


Peirce on Statistical Explanation



Ilkka Niiniluoto

The Commens Encyclopedia
The Digital Encyclopedia of Peirce Studies
New Edition

Edited by Mats Bergman and João Queiroz

URL <http://www.commens.org/encyclopedia/article/niiniluoto-ilkka-peirce-statistical-explanation>
Retrieved 21.04.2020
ISSN 2342-4257
License Creative Commons Attribution-NonCommercial-ShareAlike

Abstract:

Carl G. Hempel's classical papers in 1942 and 1948 formulated in precise terms the deductive-nomological (D-N) model of scientific explanation. After brief hints in these articles, Hempel started the serious study of statistical explanation in the end of the 1950s, and published his inductive-probabilistic (I-P) model in 1962. Wesley Salmon's extensive historical summary of the "four decades of scientific explanation" (1989) asserts that Hempel's paper in 1962 is "the first attempt by any philosopher to give a systematic characterization of probabilistic or statistical explanation". However, this claim is unfair to Charles S. Peirce who was concerned with scientific explanation already in the 1860s and gave a detailed model for statistical explanation in his 1883 article on "probable" and "statistical deduction". In fact, Peirce's account covers explanations of singular events and statistical facts, and thus is richer than most later theories of probabilistic explanation.

Keywords: Explanation, Probable Inference, Propensity, Reference Class, Statistical Deduction, Truth-frequency

I. The Subsumption Theory of Explanation

Carl G. Hempel's seminal papers "The Function of General Laws in History" (1942) and "Studies in the Logic of Explanation" (1948, with Paul Oppenheim) started a new research area in the philosophy of science. One of the main items in the new philosophical agenda of the logical empiricists was Hempel's proposal to make precise the notion of scientific explanation (see Hempel, 1965). According to Hempel's *deductive-nomological (D-N) model*, scientific explanations are arguments which answer why-questions about particular events or general regularities by "subsuming" them under general laws and particular antecedent conditions. The *explanandum* is a statement E about a fact, an event, or a regularity which is already known (or believed) to be true, and this statement E is required to be a deductive consequence of the *explanans*. This "subsumption theory" of explanation has also been called the "covering-law model", since it requires that the explanans contains at least one law. Variants of this view were defended by Richard Braithwaite (1953) and Ernest Nagel (1961). However, the task of distinguishing genuine "nomic" laws from merely accidentally true generalizations turned out to be a hard problem (see Fetzer, 1981; Psillos, 2002). Also the general task of giving sufficient and necessary conditions for adequate deductive scientific explanations proved to be surprisingly difficult (see the survey in Tuomela,

1977). But if the logical form of lawlike statements is expressed by (or at least entails) universal generalizations, simple paradigmatic examples of D-N explanations are represented by

$$(1) \quad \begin{array}{l} (x)(Fx \rightarrow Gx) \\ Fa \\ \hline Ga \end{array}$$

$$(2) \quad \begin{array}{l} (x)(Fx \rightarrow Gx) \\ (x)(Hx \rightarrow Fx) \\ \hline (x)(Hx \rightarrow Gx). \end{array}$$

To understand and appreciate Hempel's achievement, it is important to consider also his precursors. This is a topic which is largely neglected in Wesley Salmon's (1989) in many other ways illuminating historical survey of the "four decades of scientific explanation". According to Salmon, "the 1948 Hempel-Oppenheim article marks the division between the prehistory and the history of modern discussions of scientific explanation" (p. 10). His historical account jumps directly from Aristotle to Hempel, mentioning only John Stuart Mill and Karl Popper in a footnote (p. 187).

When Morton White attributed the deductive-nomological pattern of historical explanation to Hempel's 1942 paper, Karl Popper immediately complained that Hempel had only reproduced his theory of causal explanation, originally presented in *Logik der Forschung* (1935) (see Popper 1945, Ch. 25, note 7, and Popper 1957a, p. 144). With his charming politeness, Hempel (1948) pointed out that his account of D-N explanation is "by no means novel" but "merely summarizes and states explicitly some fundamental points which have been recognized by many scientists and methodologists". Hempel went on to quote definitions of explanation as subsumption under laws from the 1858 edition of John Stuart Mill's *A System of Logic* (1st ed. 1843), Stanley Jevons' *Principles of Science* (1st ed. 1873), and from the books of Ducasse (in 1925), Cohen and Nagel (in 1934), and Popper (1935) (see Hempel 1965, p. 251). Later he added N.R. Campbell (in 1920) in this list (Hempel, 1965, p. 337). In the same spirit, with sarcasm directed at Popper, G. H. von Wright has remarked that "in point of fact the 'Popper-Hempel' theory of explanation had been something of a philosophic commonplace ever since the days of Mill and Jevons" (von Wright, 1971, p. 175).

In an often quoted passage, Mill says that "an individual fact is said to be explained by pointing out its cause, that is, by stating the law or laws of causation of which its production is an instance ... and in a similar manner, a law of uniformity in nature is said to be explained when another law or laws are pointed out, of which that law itself is but a case, and from which it could be deduced." (Mill, 1906, p. 305.) This is an elaboration of Auguste Comte's assertion in 1830 that science endeavors to discover the "actual laws of phenomena", or "their invariable relations of succession and likeness", so that "the explanation of facts ... consists henceforth only in the connection established between different particular phenomena and some general facts." (Comte, 1970, p. 2.)

Comte and Mill defended the subsumption theory of explanation in a form where laws express verifiable general connections between phenomena. Many later positivists and instrumentalists excluded explanations from science, since they thought that science should drop *why*-questions in favor of descriptive *how*-questions. Pierre Duhem in 1907 explicitly expressed the fear that the aim of explanation would "subordinate" science to metaphysics (Duhem, 1954, p. 10).

The subsumption theory of explanation is a natural ally of the hypothetico-deductive method. René Descartes, Robert Boyle and Isaac Newton in the 17th century, and John Herschel and William Whewell in the 19th century demanded that a good hypothesis or theory should "explicate", "explain", "prove", "demonstrate", or "account for" known facts. In this view, a theoretical hypothesis, even if it is not directly verifiable and refers to theoretical entities and processes beyond observation, receives inductive support or confirmation from those observed facts that it successfully explains.

The origin of the theory of deductive explanation goes even further back in history - to the Aristotelian ideal of demonstrative science. In distinguishing the four types of "causes" or "explanatory factors" (Greek *aitia*), Aristotle argued that inquiry proceeds from knowing *that* to knowing *why*: "We cannot claim to know a subject matter until we have grasped the 'why' of it, that is, its fundamental explanation." (Aristotle 1961, p. 28.) First we know by observation that there is a fact; then the answer to a why-question is provided by a *scientific syllogism* which demonstrates the fact as an effect of its cause. This explanatory stage of science was called *compositio* by medieval and Renaissance Aristotelians (see Niiniluoto, 2018). The paradigmatic examples (1) and (2) of Hempelian D-N explanations are arguments which can be formulated in the mode of Aristotle's *Barbara* syllogism. Aristotle's *Posterior Analytics* can thus be viewed as the first systematic attempt to disclose the deductive structure of scientific explanation.

II. Inductive-probabilistic Explanation

Already in 1942, Hempel hinted that some explanatory arguments may replace the universal or "deterministic" laws of the D-N model with "probability hypotheses" which (together with the antecedent conditions) make the explanandum event "highly probable" (Hempel, 1965, p. 237). His example is Tommy's coming down with the measles two weeks after his brother; here the law asserts that contagion occurs "only with high probability". In 1948, Hempel mentioned examples from economics (supply and demand) and linguistics (phonologic decay), and added that the involved laws may be statistical (ibid., pp. 252-253). He remarked that the subsumption under "statistical laws" has "a peculiar logical structure" which "involves difficult special problems" (ibid., p. 251).

One of the "difficulties" in this context is the problem of giving an adequate formulation of *statistical laws*. The simplest statistical counterparts of universal generalizations are obviously statements about the *relative frequency* $rf(G/F)$ of an attribute G in a class F; this relative frequency is one if all Fs are G. As a student of the famous frequentist Hans Reichenbach in Berlin, Hempel had himself, already in the early 1930s, attempted to defend a "finitistic" version of the frequency interpretation where probability statements are applied only to finite reference classes, but Reichenbach immediately rejected that account in his comment in *Erkenntnis*. Later Hempel defined probability statements as limits of relative frequencies of properties in infinite sequences. In Hempel (1962), statistical laws were characterized as probability statements of the form $P(G/F) = r$, where 'r' takes values between zero and one and the reference class F includes all the potential instances of F (Hempel, 1962, p. 123). One way of making sense of this requirement is to treat statistical probability as a disposition, as in Popper's (1957) propensity interpretation. The dispositional interpretation was mentioned again in Hempel (1965), p. 378. In Hempel (1968) probability is again a long-run frequency, but the relevant predicates in a lawlike probability statement are required to be "nomic".

It may seem surprising that, after the suggestive remarks in 1948, the formulation of Hempel's model of statistical explanation was delayed for 14 years - and no one else picked up the topic for study in the meantime (but see Nagel, 1961, and Rescher, 1962). Finally in 1962 Hempel published his article "Deductive-Nomological vs. Statistical Explanation". The reason for this delay was Hempel discovery of the problem of the "ambiguity of statistical syllogisms" (Hempel 1962, p. 138). This problem was not noted by D.C. Williams (1947) in his extensive treatment of predictive statistical syllogisms,

but it was discussed in detail by Stephen Barker (1957). The problem arises, because the statistical probability $P(G/F)$ of outcome G in the class F depends on the reference class F . Then two arguments of the form

- (4) $P(G/F)$ is nearly 1
 Fa
 So, it is almost certain that a is G
- (5) $P(\sim G/H)$ is nearly 1
 Ha
 So, it is almost certain that a is not G

may both have true premises, even if their conclusions are inconsistent with each other. In "Inductive Inconsistencies" (1958), Hempel offered the solution that, in a statistical syllogism, probability should not be understood as a *modal qualifier* of the conclusion but rather as a *relation* between the premises and the conclusion. This relation involves logical or inductive probability, or a degree of confirmation in the sense of J. M. Keynes (1921) and Rudolf Carnap (1950). Thus, instead of (4), we should say that ' Ga ' is highly probable relative to the statements ' $P(G/F) \approx 1$ ' and ' Fa ' (Hempel 1965, p. 60). Schema (4) can now be written in the form

- (6) $P(G/F)$ is nearly 1
 Fa
 ===== [makes almost certain]
 Ga

where the double line indicates that (6) is an inductive rather than a deductive argument.

With this reasoning, Hempel was finally ready to formulate his model of *inductive-statistical (I-S) explanations* in 1962. This model can be expressed by the schema

- (7) $P(G/F) = r$
 Fa
 ===== [r]
 Ga

where 'r' in the lawlike premise 'P(G/F)=r' is a *statistical probability*, and 'r' in the brackets indicates the inductive probability of the explanandum 'Ga' given the explanans. Further, Hempel required that r should be close to one.

Hempel's I-S model still allows that there may be inductive arguments of the form (7) with true premises but with incompatible conclusions 'Ga' and '~Ga'. For the purpose of prediction this is clearly a problem, as it is not yet known whether 'Ga' is true or not. But for explanation the choice of the correct explanation is already determined by the assumption that the explanandum sentence is known to be true (see Coffa, 1974; Salmon, 1989, p. 69). Hempel knew this, but he insisted that it is unnatural to admit that, in giving an I-S explanation of a fact, we could have "just as readily" explained its opposite from true premises (see Hempel, 1968, p. 119).

To handle this problem of *ambiguity*, Hempel (1962), p. 146, demanded that the rule (7) should be "based on the statistical probability of G within the narrowest class, if there is one, for which the total evidence available provides the requisite statistical probability". This is a counterpart for explanation of Hans Reichenbach's (1938) advice in the context of prediction: use the statistical probability statement with the narrowest available reference class for determining the "weights" of single cases. Similarly, von Wright (1945), pp. 35-36, advocated the principle that scientific prediction should be based on a minimal epistemically homogenous reference class.

A precise formulation of the *Requirement of Maximal Specificity* was given by Hempel (1965), pp. 397-400. The total evidence is now represented by the set K of "all statements accepted at the given time". K is assumed to be deductively closed and to contain the axioms of probability theory. If K contained the explanandum 'Ga', then trivially the logical probability of this statement relative to K would be equal to one. Therefore, K is assumed not to contain the explanandum (but see Hempel, 1968). In a situation, where the premises of the argument (7) are known, the relevant knowledge situation K should satisfy

(RMS) For any class F_1 for which K implies that F_1 is a subclass of F and that F_1a , K also contains a law to the effect that $P(G/F_1) = r_1$, where $r_1 = r$, unless that law is a theorem of probability theory.

The unless-clause excludes the use of classes like $F \cap G$ and $F \cap \neg G$ for the choice of F_1 . RMS thus attempts to specify what information in our knowledge situation K is of potential explanatory relevance to the explanandum. When RMS is satisfied, the

probability r expresses the nomic expectability of G_a on the basis of the explanans. Hempel (1968) reformulated RMS so that it applies to predicates rather than classes, and is less demanding on the existence of known statistical laws:

(RMS*) If K contains statements ' F_1a ' and ' $(x)(F_1x \rightarrow Fx)$ ' and the lawlike statement ' $P(G/F_1) = r_1$ ', then $r_1 = r$, unless r_1 is one or zero.

Both RMS and RMS* are relativized to K . Hempel claimed that this is unavoidable, so that the notion of potential I-S explanation (unlike D-N explanation) makes sense only if relativized to a knowledge situation. This is Hempel's thesis of the *epistemic relativity* of statistical explanation (Hempel, 1965, p. 402).

Later in his 'Nachwort 1976' to the German edition of his *Aspects of Scientific Explanation*, Hempel (1977) reformulated RMS by dropping the condition ' F_1a ' from its antecedent. In this form, it requires that we do not know of any subclass F_1 of F such that the probability of G in F_1 differs from the probability of G in F . In other words, the reference class F should be *epistemically homogeneous* for G in the sense of Salmon (1971, 1984). At the same time, Hempel dropped the requirement that the probability r is high. Further, he stated that it would be "very desirable" to find an objective, not epistemically relativized formulation of maximal specificity, but left it open whether such a definition can be found (Hempel, 1977, p. 123). These modifications were responses to lively debates about Hempel's I-S model.

In the debates following the appearance of Hempel's I-S model in the 1960s, some philosophers still denied altogether the idea that probabilistic arguments could serve as explanations. In particular, G. H. von Wright (1971) and Wolfgang Stegmüller (1973) argued that Hempel's I-S model does not answer explanatory why necessary -questions but is valid only for inductive predictions and other reason-giving arguments.

One of the early objections to Hempel's I-S model is that high probability is not necessary for statistical explanation. This question was raised by Rescher (1962) and Salmon (1965). A forceful statement against the high probability requirement - and, more generally, against the view that statistical explanation are arguments - was presented by Richard Jeffrey (1969) (see also Salmon, 1970). Some philosophers still demanded that the probability of the explanandum has to be at least $1/2$; this was called the "Leibniz condition" by Stegmüller (1973). Hempel's (1977) response was that eventually he dropped the high probability requirement.

Another issue is that the idea of inductive support or confirmation could be explicated,

instead of high probability, by the *positive relevance* criterion (see Carnap, 1962; Hempel, 1965, p. 50). Thus, we might require that the explanans increases the probability of the explanandum. This proposal was made by Wesley Salmon (1965), who developed his *statistical relevance (S-R) model* of explanation as a rival to Hempel's I-S model (Salmon 1970, 1984). With Jeffrey (1969), Salmon rejects Hempel's high probability requirement. He also replaces the epistemic principle RMS with a requirement that the reference class F should be *objectively homogeneous* for attribute G, i.e., no property H (independent of G) divides F in a "statistically relevant" way to a subclass $F \cap H$ such that $P(G/F) \neq P(G/F \cap H)$.

One difference between the S-R and I-S models is Salmon's requirement that the homogeneous reference class should be maximal. This relevance condition leads to unintuitive results already in the deductive case, since in some cases we may be unwilling to combine two separate causes into one class (e.g., to explain why a piece of white substance melted in water by the fact that it was salt *or* sugar) – and the same holds for probabilistic causes that produce the same effect with the same statistical probability (see Hempel, 1977; Niiniluoto, 1982). By the same token, one should not generally demand the choice of a minimal maximally specific reference class (see Hempel, 1968).

In Salmon's S-R model, in contrast to his original 1965 proposal, it may happen that the posterior probability of the explanandum is higher, smaller, or equal to the initial probability (positive relevance, negative relevance, irrelevance, respectively). It is an interesting question whether this move could be justified by a suitable "global" conception of explanatory power (see Salmon et al., 1971; Niiniluoto and Tuomela, 1973; Niiniluoto, 1981, 1982). Stegmüller (1973) argues that Salmon's S-R model is not an explication of statistical explanation but instead of "statistical depth analysis". Salmon (1984) defends his model as exemplifying the *ontic* conception of explanation: the explanandum event should be fitted into the nexus of causal or lawlike regularities, and the epistemic idea of expectation is irrelevant to the aim of explanation (see also Salmon, 1989, pp. 117-122).

The crucial question about the ontic conception is the possibility of giving a reasonable explication of the notion of "objectively homogeneous" reference class. Hempel's doubts about objective or non-epistemic formulations of RMS led him to announce the epistemic relativity of I-S explanation, but – as he acknowledged later (Hempel, 1977) – this argument was not conclusive. On the other hand, J. Alberto Coffa's (1974) and Salmon's claim that the epistemic relativity thesis commits Hempel to determinism was

not conclusive, either (see Niiniluoto, 1976).

Salmon's strategy in developing the ontic conception has been to rely on a long run frequency interpretation of statistical probability, together with a theory causal processes. Another approach is to base the analysis of statistical laws on the single-case propensity interpretation of probability, defended in the early 1970s by Ron Giere and James Fetzer. Instead of viewing physical probabilities as sure-fire dispositions to produce long-run frequencies, the single-case interpretation takes propensities to be degrees of possibility that are displayed by a chance set-up in each single trial of a certain kind. According to this view, a lawlike probability statement $(x)[Hx \rightarrow Px(G/F) = r]$ asserts then that every chance set-up x of type H has a dispositional tendency of strength r to produce a result of type G on each single trial of kind F . In the special case $r = 1$ this analysis reduces to a non-Humean intensional analysis equating lawlikeness either with "physical necessity" (i.e., truth in all physically possible worlds) or counterfactual conditionality (see Fetzer, 1981). Probability in physical laws $P(G/F) = r$ (where $r < 1$) is then a modal operator which is weaker than physical necessity. If $0 < r < 1$, such a law implies that set-ups of type H are indeterministic systems.

A clear formulation of a propensity model of probabilistic explanation was given by James H. Fetzer (1974). A typical explanation with such single-case propensity laws is as follows:

$$(8) \quad (x)[Hx \rightarrow Px(G/F) = r]$$

$$Ha \ \& \ Fa$$

$$===== [r]$$

$$Ga$$

where r in brackets is again the degree of nomic expectability of outcome G on the relevant trial with chance set-up a . Alternatively, r is the degree of "nomic responsibility" of the causally relevant conditions to produce Ga in the given situation (cf. Fetzer, 1992, p. 258). In this case, a separate RMS condition is unnecessary, since already the law in (8) presupposes that F is objectively homogeneous for G .

Most philosophers have concentrated their efforts in analyzing explanations of singular facts and events. However, as Hempel noted, the D-N model can be applied also to the explanation of deterministic laws. Similarly, the explanandum could be taken to be a probabilistic law (Hempel, 1962, p. 147). Thus, while schema (6) is a statistical counterpart to the singular D-N inference (1), Hempel's concept of *deductive-statistical*

(D-S) explanation corresponds to the universal syllogism (2). A D-S explanation is an argument where a statistical probability statement is derived from other such statements by means of the theory of probability (Hempel, 1965, pp. 380-381).

D-S explanations have received very little attention in the philosophical literature. Nagel (1961), pp. 509-520, suggests that the formal structure of explanations of statistical generalizations in the social sciences is always deductive. As he has in mind probabilistic laws, his examples may be taken to be instances of Hempel's D-S explanation. Hempel himself stated that "ultimately, however, statistical laws are meant to be applied to particular occurrences and to establish explanatory and predictive connections among them" (ibid., p. 381).

Hempel (1962), p. 166, concluded his major essay on statistical explanation by stressing the need of a "statistical-probabilistic concept of 'because' ". His example of such probabilistic causality quotes Richard von Mises: "It is because the die was loaded that the 'six' shows more frequently". Earlier in the paper, he gave examples from Mendelian genetics, where the argument explains "the approximate percentages of red- and white-flowered plants in the sample" (ibid., p. 142), and theory of radioactivity, where the statistical law about radon's half-life explains the behavior of a large sample of such atoms (ibid., p. 142). In these examples, it is clear that the explanandum is not a statistical or probabilistic law, but rather a *statistical generalization* or *fact* about a particular *finite* class. The explanation of such statistical generalizations does not follow the structure of I-S and D-S models (see Niiniluoto, 1976, pp. 357-358); yet, it may be the most typical of the applications of statistical ideas in science.

Three different models for the explanation of statistical facts are distinguished in Niiniluoto (1981), pp. 440-442. First, a universal law may be combined with a statistical fact to give a deductive explanation of another statistical fact. For example, assume that a disease G is deterministically caused by a gene F, and that the relative frequency of genes F in a given population H is r. Then the relative frequency of disease G in H is at least r; if F is also a necessary condition for G, the latter relative frequency equals r. This inference follows the pattern

$$\begin{array}{l}
 (9) \quad (x)(Fx \rightarrow Gx) \\
 \quad \quad rf(F/H) = r \\
 \quad \quad \text{-----} \\
 \quad \quad rf(G/H) \geq r
 \end{array}$$

Secondly, statistical facts can also be inductively explained by probabilistic laws and universal generalizations. For example, if a radon atom decays within 3.82 days with probability $1/2$, and the decays are probabilistically independent, then by Bernoulli's Theorem it is probable that in a large sample of radon atoms the number of decays within 3.82 days is approximately $1/2$. This inference, which is a straightforward generalization of Hempel's I-S model for singular explanation, has the form

$$(10) \quad \begin{array}{l} P(G/F) = r \\ H_n \text{ is a large finite sample of } F\text{s} \\ \text{=====} \\ \text{rf}(G/H_n) \approx r \end{array}$$

Thirdly, as a generalization of (9) and (10), a probabilistic law may be combined with a statistical fact to give an inductive explanation of another statistical fact:

$$(11) \quad \begin{array}{l} P(G/F) = r \\ \text{rf}(F/H) = s \\ \text{=====} \\ \text{rf}(G/H) \approx rs \end{array}$$

For example, if gene F produces disease G with probability r , then it is highly probable that the relative frequency of G in a subclass $F \cap H$ of a given finite population H is approximately r . If now the relative frequency of gene F in population H is s , then the relative frequency of G in H is at least rs .

III. Peirce on Statistical Explanation

The surveys of the subsumption theory (Section I) and recent discussions of statistical explanation (Section II) give us a useful background for assessing Peirce's contributions to scientific explanation (see Niiniluoto, 1993, 2000).

Hempel's *Aspects of Scientific Explanation* (1965) does not mention at all the name of Charles S. Peirce. Also others who have listed 19th century advocates of D-N explanation have failed to refer to Peirce's contribution. Even though Peirce's reputation grew only slowly in the twentieth century, and some of the important evidence comes from his early *Writings* published in the 1980s, this silence is quite surprising in view of the fact that Peirce's paper "Deduction, Induction, and Hypothesis"

(1878), which gives examples of deductive explanations, was available in the collection *Chance, Love, and Logic* in 1923. Further, Peirce's *Collected Papers* (1931 - 35) contain a reprint of the article "A Theory of Probable Inference" (1883), which formulates a theory of probabilistic explanation, with deductive explanation as a special case.

Peirce's writings on probability and induction were of course well-known to many philosophers before 1950: they were discussed, among others, by Keynes, Ramsey, Braithwaite, Nagel, von Wright, Carnap, and Williams. While it was recognized that Peirce was interested both in inductive inference from a sample to a population and in "probable deduction" from a population to a sample or to single cases, it was almost always thought that the latter type of inference - variously called "the use of *a priori* probabilities for the prediction of statistical frequency" (Keynes, 1921), "the problem of the single case" (Reichenbach, 1938), "statistical syllogism" (Williams, 1947), or "direct inference" (Carnap, 1950) - was concerned with *prediction* rather than *explanation*. Peirce himself never made such a restriction, however.

In Niiniluoto (1982), p. 160, I pointed out that "strangely enough, it seems that the modern literature on statistical explanation does not contain even a single reference to Peirce's theory of explanatory statistical syllogisms". In another article, I ventured to suggest that "Peirce should be regarded as the true founder of the theory of inductive-probabilistic explanation" (Niiniluoto, 1981, p. 444).

Salmon, who took 1962 to be "the year in which the philosophical theory of scientific [statistical] explanation first entered the twentieth century" (Salmon, 1983, p. 179), expressed disagreement with my judgment, since "one isolated and unelaborated statement" about explanatory statistical syllogisms "can hardly be considered even the beginnings of any genuine theory" (Salmon, 1984, p. 24). In his survey article some years later (Salmon, 1989), Peirce is not mentioned nor listed even in the "prehistorical" part of the bibliography.

However, it can be argued against Salmon that Peirce had a serious and systematic concern for scientific explanation ever since 1865, and that his 1883 account of "probable" and "statistical deduction" gives a rich and detailed model for the structure of statistical explanation. Indeed, it seems to me that the relation of Peirce's work to the I-S model parallels the relation of Aristotle to the D-N model (see Niiniluoto, 1993).

Peirce's interest in the structure of scientific explanation arose from his early studies in Aristotle's logic. In his Harvard Lectures during the spring of 1865, the young Peirce observed that there is a type of reasoning that is neither deductive nor inductive (W

1:180). This reasoning, which Peirce called Hypothesis (and later abduction), can be represented as the inference of the minor premise of a syllogism, or inference of a cause from its effect. The classification of inferences into Deduction, Induction, and Hypothesis was elaborated in Peirce's Lowell Lectures in the fall of 1866, and published in the next year. Ten years later, this distinction was presented in the article "Deduction, Induction, and Hypothesis".

Already in the Harvard Lectures 1865, Peirce made it perfectly clear that Hypothesis is an inference *to an explanation* (W 1:267). In the Lowell Lectures 1866, Peirce said that hypothesis - which alone "enables us to see the *why* of things" - is the inversion of the corresponding *explaining syllogism*:

"Ether waves are polarizable.

Light is ether waves.

□ Light is polarizable."

In general, "to explain a fact is to bring forward another from which it follows syllogistically", i.e., "we say that a fact is *explained* when a proposition - possibly true - is brought forward, from which that fact follows syllogistically." (W 1:428, 425, 440, 452.) This is a clear anticipation of the D-N-model of explanation.

In his early remarks, Peirce shows that he was aware of problems of relevance, later discussed by Salmon and others. He points out that if "D is C" can be deductively explained by the middle terms I and J, then it can be explained also by the more extensive predicate "I or J" (W 1:293).

Another important influence came from Peirce's fascination with probability theory. In his 1867 review of John Venn's *The Logic of Chance* (1866), the first systematic treatment of the frequency interpretation of probability, Peirce discussed inferences of the form

- (12) A is taken at random from among the B's
 2/3 of the B's are C
 □ A is C.

Peirce's justification for schema (12) is in terms of *truth-frequencies*: in the long run an argument of form (12) would yield a true conclusion from true premises two thirds of the time (CP 8.2).

Peirce formulated induction, hypothesis, and analogy as probable arguments in 1867,

where probability is measured by the proportion of cases in which an argument "carries truth with it". In 1878, he formulated probabilistic versions of the *Barbara* syllogism and its inversions by replacing the universal law with a statistical generalization of the form "Most of the beans in this bag are white" (CP 2.508 - 516, 2.627).

The article "A Theory of Probable Inference" (1883) gives several models of probable "deduction" from a statistical premise about a population. *Simple Probable Deduction* is a statistical version of singular syllogism in Barbara (cf. (1)):

- (13) The proportion r of the F's are G's;
 a is an F;
 It follows, with probability r , that a is a G.

As Peirce noted, the conclusion here is 'a is a G', and probability indicates "the modality with which this conclusion is drawn and held to be true". Here Peirce anticipates Hempel's discussion of "inductive inconsistencies", i.e., the patterns (3) - (5). Further, it is required that a "should be an instance drawn *at random* from among the F's".

"The volition of the reasoner (using what machinery it may) has to choose a so that it shall be an F; but he ought to restrain himself from all further preference, and not allow his will to act in any way that might tend to settle what particular F is taken, but should leave that to the operation of chance."

"... the act of choice should be such that if it were repeated many enough times with the same intention, the result would be that among the totality of selections the different sorts of F's would occur with the same relative frequencies as in experiences in which volition does not intermeddle at all. In cases in which it is found difficult thus to restrain the will by a direct effort, the apparatus of games of chance - a lottery-wheel, a roulette, cards, or dice - may be called to our aid." (CP 2.696)

This condition guarantees that the result G is obtained with the long-run frequency within the unlimited population of possible drawings from the class of F's (CP 2.731). Hence, the inference schema (13), i.e.,

- (14) $rf(G/F) = r$
 a is a random member of F
 □ a is a G

with a relative frequency statement as a premise, can be formulated as a probabilistic argument with a lawlike statistical premise

$$\begin{array}{l}
 (15) \quad P(G/F) = r \\
 \quad \quad Fa \\
 \quad \quad ===== [r] \\
 \quad \quad Ga
 \end{array}$$

where r in the brackets is a long-run truth-frequency.

The schema (15) corresponds to Hempel's model (6) of I-S explanation - with the difference that $[r]$ indicates an objective probability or truth-frequency rather than an epistemic or inductive probability. But even this difference with Hempel is not very great. Even though Peirce was critical of Bayesian epistemic probabilities (CP 2.780), he added - by appealing to Fechner's law - that there is a relation between objective probabilities and degrees of belief: as a matter of fact we do have "a stronger feeling of confidence about a sort of inference which will oftener lead us to the truth than about an inference that will less often prove right" (CP 2.697).

If probability is understood as a long-run propensity, as Peirce suggested in his later work in 1910 (CP 2.664), pattern (15) comes close to the propensity model (7) of probabilistic explanation, even though Peirce failed to directly associate probabilities with single-case propensities (see Fetzer, 2014).

Besides Simple Probable Deduction (13), Peirce (1883) formulated a schema for *Statistical Deduction* which proceeds from a population to a sample:

- (16) The proportion r of the F 's are G 's,
 a', a'', a''' , etc. are a *numerous* set, taken at random from among the F 's;
 Hence, *probably* and *approximately* the proportion r of the a 's are G 's.

This inference can be formalized in the following way:

$$\begin{array}{l}
 (17) \quad rf(G/F) = r \\
 \quad \quad \{a', a'', a''', \dots\} \text{ is a random sample of } F\text{'s} \\
 \quad \quad \square rf(G/\{a', a'', a''', \dots\}) \approx r.
 \end{array}$$

Again the condition for randomness allows us to reformulate (17) by

$$\begin{array}{l}
 (18) \quad P(G/F) = r \\
 \quad \quad Fa' \ \& \ Fa'' \ \& \ Fa''' \ \& \ \dots \\
 \quad \quad \text{-----} \\
 \quad \quad \text{-----} \\
 \quad \quad \text{fr}(G/\{a', a'', a''', \dots\}) \approx r
 \end{array}$$

where, as Peirce showed by the Binomial Formula, r is the most probable value of the relative frequency in the conclusion. Further, by Bernoulli's Theorem, the probability of the conclusion given the premises approaches one when the size of the sample a', a'', a''', \dots increases without limit (CP 2.698-700). This schema is the same as the often neglected pattern (10) for the explanation of statistical facts.

It is interesting to note that Peirce's discussion of probable reasoning was anticipated by Mill and Venn. Mill's *System of Logic* contained a brief discussion of the application of an approximate generalization (like 'Most A are B' or 'Nine out of ten A are B') to its individual instances. Mill required that we should know nothing about such instances "except that they fall within the class A" (Mill, 1906, p. 391; Niiniluoto, 1981, p. 444). In a rough form, this guarantees that Hempel's RMS is satisfied - without yet implying the stronger condition that the class A itself is objectively or epistemically homogeneous.

Similarly, Venn's *Logic of Chance* formulated the rule that statistical inferences about an individual case should refer it to the narrowest series or class which still secures "the requisite degree of stability and uniformity" (Venn, 1888, p. 220; cf. Reichenbach, 1938).

However, Mill and Venn never said that such a statistical inference could be applicable for the purpose of *explanation*. It is clear that for Mill approximate generalizations were important primarily "in the practice of life", but in science they are valuable as "steps towards universal truths". Moreover, Venn explicitly restricted his attention to attempts "to make real inferences about things as yet unknown", i.e., to *prediction* (Venn, 1888, p. 213).

Peirce, on the other hand, in so many words reasserted his earlier view that "Inductions and Hypotheses are inferences from the conclusion and one premiss of a statistical syllogism to the other premiss. In the case of hypothesis, this syllogism is called the *explanation*." Indeed, Peirce repeated, "we commonly say that the hypothesis is adopted *for the sake* of the explanation". A statistical syllogism "may be conveniently termed the explanatory syllogism" (CP 2.716 - 717).

Peirce was also aware that in explanation it is usually impossible to choose the individual under consideration by a random selection: one might argue that the schema (15) is not applicable to explanation, since in explaining why a is a G we already know the individual a, and therefore we cannot draw it randomly from the class F. Here is Peirce's reply:

"Usually, however, in making a simple probable deduction, we take that instance in which we happen at the time to be interested. In such a case, it is our interest that fulfills the function of an apparatus for random selection; and no better need be desired, so long as we have reason to deem the premiss 'the proportion r of the F's are G's' to be equally true in regard to that part of the F's which are alone likely ever to excite our interest." (CP 2.696)

The intuition seems to be that the "interesting" subclasses of F should preserve the probability of G, that is, the schema (15) can be used for explanation as long as the reference class F is epistemically homogeneous.

Peirce did not return to the logic of probabilistic inference after his 1883 paper. But it is unfair to say with Salmon that his treatment was "unelaborated" and "isolated", since it was systematically connected to the most central tenets of his work on probability, scientific method, and metaphysics. Peirce's views of the nature of probability and of the role of chance in nature went through changes which deepened his insight of the indispensability of statistical explanation. Evidence from science suggested that some of the best theories have a statistical character: already in "The Fixation of Belief" (1877), Peirce referred to the 1851 theory of gases by Clausius and Maxwell and the 1859 theory of evolution by Darwin (CP 5.364, 1.104, 6.47, 6.613). In his 1892 papers for *The Monist*, Peirce formulated his evolutionary metaphysics, with its principles of absolute chance (*tychism*) and continuity (*synechism*) in nature (CP 6.13). In the first years of the 20th century, Peirce radicalized his criticism of Hume's "Ockhamist" views of the laws of nature, adopted a realist view of dispositional and modal conceptions (see Fisch, 1967), and proposed a propensity interpretation of probability as a "would-be" of a physical chance set-up (cf. Niiniluoto, 1988, Fetzer, 2014). Peirce's tychism thus led him to a philosophical position, where the world is governed by evolving probabilistic laws – and all explanation of natural phenomena is probabilistic. This version of indeterminism has affinities with Salmon's ontic account of statistical explanation with objectively homogeneous reference classes, even though Salmon accepted propensities only as causes of frequencies (cf. Niiniluoto, 1988).

After reading my paper for the 1989 Peirce Congress (see Niiniluoto, 1993), Hempel told me "with a strong sense of embarrassment" that he had been "unaware of Peirce's

writings on probabilistic theorizing and statistical explanation".¹ The orientation of philosophical inquiry in the Berlin group and the Vienna Circle, he told, was "basically a-historical". Hempel acknowledged that Peirce's ideas on statistical explanation "constitute important pioneering contributions to the field", even though "they are not as precisely formulated and as theoretically integrated and comprehensive as are more recent accounts, of which Salmon's theory is a fine example".

After his conclusion in the letter, Hempel still added one remark: "I have never thought of myself as the founder of the theory of statistical explanation, but only as the proponent of one explicatory approach to that important problem complex". This is a remarkably modest statement. If Peirce was the true founder of the theory of statistical explanation, Hempel was the first philosopher who was able to give definitions and arguments that convinced his contemporaries of the existence of probabilistic explanations.²

References

- Aristotle (1961). *Physics*. Lincoln: University of Nebraska Press.
- Barker, S. (1957). *Science and Hypothesis*. Ithaca, New York: Cornell University Press.
- Braithwaite, R.B. (1953). *Scientific Explanation*. Cambridge: Cambridge University Press.
- Carnap, R. (1950/1962). *The Logical Foundations of Probability*. Chicago: The University of Chicago Press.
- Coffa, J. A. (1974). Hempel's Ambiguity, *Synthese* 28, 141-163.
- Comte, A. (1970). *Introduction to Positive Philosophy*. Indianapolis: Bobbs-Merrill.
- Duhem, P. (1954). *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.
- Fetzer, J. H. (1974). A Single Case Propensity Theory of Explanation. *Synthese* 28, 171-198.
- Fetzer, J. H. (1981). *Scientific Knowledge*. Dordrecht: D. Reidel.
- Fetzer, J. H. (1992). What's Wrong with Salmon's History: The Third Decade. *Philosophy of Science* 59, 246-262.
- Fetzer, J. H. (2014). Peirce and Propensities. *The Commens Encyclopedia of Peirce Studies*.
- Hempel, C. G. (1942). The Function of General Laws in History. *Journal of Philosophy* 39, 35-48. Reprinted in Hempel (1965), 231-243.
- Hempel, C. G. (1960). Inductive Inconsistencies. *Synthese* 12, 439-469. Reprinted in Hempel (1965), 53-79.
- Hempel, C. G. (1962). Deductive-Nomological vs. Statistical Explanation. In H. Feigl and G. Maxwell (eds.),

- Minnesota Studies in the Philosophy of Science*, Vol. 3 (pp. 98-169). Minneapolis: University of Minnesota Press,
- Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.
- Hempel, C. G. (1968). Maximal Specificity and Lawlikeness in Probabilistic Explanation. *Philosophy of Science* 35, 116-133.
- Hempel, C. G. (1977). Nachwort 1976. In *Aspekte wissenschaftlicher Erklärung*, Berlin: Walter de Gruyter.
- Hempel, C. G. and Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science* 15, 135-175. Reprinted in Hempel (1965), 245-295.
- Jeffrey, R. C. (1969). Statistical Explanation vs. Statistical Inference. In N. Rescher (ed.), *Essays in Honor of Carl G. Hempel* (pp. 104-113). Dordrecht: Reidel. Reprinted in Salmon et al. (1971), 19-28.
- Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.
- Mill, J. S. (1843/1906). *A System of Logic*. London: Longmans, Green, and Co.
- Nagel, E. (1961). *The Structure of Science*. London: Routledge and Kegan Paul.
- Niiniluoto, I. (1976). Inductive Explanation, Propensity, and Action. In J. Manninen ja R. Tuomela (eds.), *Essays on Explanation and Understanding* (pp. 335-368). Dordrecht: D. Reidel.
- Niiniluoto, I. (1981). Statistical Explanation Reconsidered. *Synthese* 48, 437-472.
- Niiniluoto, I. (1982). Statistical Explanation. In G. Floistad (ed.), *Contemporary Philosophy: A New Survey*, vol. 2 (pp. 157-187). The Hague: M. Nijhoff.
- Niiniluoto, I. (1988). Probability, Possibility, and Plenitude. In J. Fetzer (ed.), *Probability and Causality: Essays in Honor of Wesley C. Salmon* (pp. 91-108). Dordrecht: D. Reidel.
- Niiniluoto, I. (1993). Peirce's Theory of Statistical Explanation. In E. C. Moore (ed.), *Charles S. Peirce and the Philosophy of Science* (pp. 186-207). Tuscaloosa and London: The University of Alabama Press.
- Niiniluoto, I. (2000). Hempel's Theory of Statistical Explanation. In J. Fetzer (ed.), *Science, Explanation, and Rationality: The Philosophy of Carl G. Hempel* (pp. 138-163). Oxford: Oxford University Press.
- Niiniluoto, I. and Tuomela, R. (1973). *Theoretical Concepts and Hypothetico-Inductive Inference*. Dordrecht: D. Reidel.
- Popper, K. R. (1945). *The Open Society and Its Enemies*. London: Routledge and Kegan Paul.
- Popper, K. R. (1957a). *The Poverty of Historicism*. London: Routledge and Kegan Paul.
- Popper, K. R. (1957b). The Propensity Interpretation of the Calculus of Probability, and the Quantum Mechanics. In S. Körner and M. H. L. Pryce (eds.), *Observation and Interpretation* (pp. 65-70). London: Butterworths.

- Psillos, S. (2002). *Causation and Explanation*. Chesham: Acumen.
- Reichenbach, H. (1938). *Experience and Prediction*. Chicago: University of Chicago Press.
- Rescher, N. (1962). The Stochastic Revolution and the Nature of Scientific Explanation. *Synthese* 14, 200-215.
- Salmon, W. C. (1965). The Status of Prior Probabilities in Statistical Explanation. *Philosophy of Science* 32, 137-146.
- Salmon, W. C. (1970). Statistical Explanation. In R. G. Colodny (ed.), *Nature and Function of Scientific Theories* (pp. 173-231). Pittsburgh: University of Pittsburgh Press. Reprinted in Salmon et al. (1971), 29-87.
- Salmon, W. C. et al. (1971). *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.
- Salmon, W. C. (1983). Probabilistic Explanation: Introduction. In P. D. Asquith and T. Nickles (eds.), *PSA 1982*, vol. 2 (pp. 179-180). East Lansing: Philosophy of Science Association.
- Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Salmon, W. C. (1989). Four Decades of Scientific Explanation. In P. Kitcher and W. C. Salmon (eds.), *Scientific Explanation* (pp. 253-282). Minneapolis: University of Minnesota Press. (Published also as a monograph by the University of Minnesota Press, 1990.)
- Stegmüller, W. (1973). *Personelle und statistische Wahrscheinlichkeit*. Berlin: Springer-Verlag.
- Tuomela, R. (1977). *Human Action and Its Explanation*. Dordrecht: D. Reidel.
- Venn, J. (1888). *The Logic of Chance*, 3rd ed. London: Macmillan.
- von Wright, G. H. (1941/1957). *The Logical Problem of Induction*. Helsingfors: Acta Philosophica Fennica 3. (2nd ed. Oxford: Blackwell, 1957).
- von Wright, G. H. (1945). *Über Wahrscheinlichkeit*. Helsingfors: Acta Societatis Scientiarum Fennicae A.III.11.
- von Wright, G. H. (1971). *Explanation and Understanding*. Ithaca, New York: Cornell University Press.
- Williams, D. C. (1947). *The Ground of Induction*. Cambridge, Mass.: Harvard University Press.

Notes

1. A private letter dated December 20, 1989. Quoted with permission by Professor Hempel.
2. This paper is mainly based on the articles Niiniluoto (1993) and Niiniluoto (2000).