STUDIA LOGICA
Tom XII – 1961

JERZY GIEDYMIN

AUTHORSHIP HYPOTHESES AND RELIABILITY OF INFORMANTS

I

In the methodology of historical research the concept "historical source" plays an essential role: the historian reconstructs the past from the material remains that survived the passage of time and these are termed "sources". They are either objects produced with the intention to communicate information, e. g. chronicles, or other results of human activities, e. g. tools, or other remains such as skeletons of men and of animals. E. BERNHEIM's classification of historical sources into tradition and remains draws our attention to that type of sources which were produced with the intention to communicate information[1].

In order that sources may be used as evidence they have to be subjected to interpretation and criticism. In case of written documents it is customary to distinguish the so called external and internal criticism. The criticism of documents is guided by a sequence of questions the ultimate aim of which is to establish the significance of the text as intended by its author (the culmination of the external criticism) and to evaluate the testimonies involved in the text as evidence for or against certain historical hypotheses (the culmination of the internal criticism). The rules of utilizing documents as evidence are in terms of reliability of informants and of independence of their testimonies. One of these rules allows to consider a fact as established if it is asserted by at least two reliable and independent informants. Obviously the historian has to test his hypotheses against sources that are reliable to some extent only and then he has to do all his best to estimate their bias or prejudice. Even completely unreliable sources are forced by historians to yield some information. As a rule, however, the historian has to base his research on partially conteminated and defective reports and his aim is to maximize trustworthy information that may be obtained from these reports; this is why there is an analogy between the methodology of the criticism of documents and the theory of information.

[1] E. BERNHEIM: *Lehrbuch der geschichtlichen Methode*, Leipzig 1930, s. 231–233.

In the present article I am going to discuss the procedure of testing historical hypotheses against interpretative statements of the form: "N asserts (in the given document) that p".

The first section deals with analogies between statistical inferences (testing statistical hypotheses and estimating parameters) and solving the problems of the authorship of documents which is often necessary in order to arrive at the interpretative statements mentioned above.

The second section shows the methodological role of the concept of reliability of informants and of the concept of independent testimonies. Some of the methodological rules accepted by historians are shown to be rational on the definition of reliabity. Relations between the reliability, confirmation and explanation functions are briefly studied.

The treatment of reliability given in this article forms a part of a theory in which testimonies (assertions) are considered as a type of rational behaviour.

II. THE AUTHORSHIP PROBLEM. SOME ANALOGIES WITH STATISTICAL INFERENCE

The historian, *qua* historian, is interested in objects referred to as sources, only because they enable him to solve at least some of historical problems. Before, however, attempts at solving problems concerned with historical facts are undertaken, usually it is necessary to answer several questions concerning the sources themselves. In the so called external criticism of the sources among such questions are — the problem who was the author of the document, when and where the document was issued whether or not the copy in our possession is genuine etc. In the present section I am going to discuss a typical procedure of discovering the author of an anonymous document and some analogies between this procedure and statistical inference.

There exist a general concept of historical source and a relativized one. In the first sense, sources are objets which "... owing to their destination, their existence, origin or other circumstances are capable of imparting information about historical facts"[2]. With this general concept in mind the methodologist of history proposes several classifications of sources and it is just in terms of this concept that he explicates the phrase saying that historical knowledge is indirect. The relativized concept of sources appears in the criticism of written documents when the question is asked whether or not the given document is a source of information with respect to a given problem. This relativized sense of "sources" is closely related with the concept of reliability. In what follows we shall be interested exclusively in the relativized concept of historical sources.

From the logical point of view another relativization is essential, *scil.* rela-

[2] E. BERNHEIM: op. cit. p. 227.

tivization to methodological (inferential) rules. The following explication is, therefore, suggested:

Explication 1: An object x is a source of information with respect to a question q if and only if there exist: a body of knowledge K, basic statements $O_1 \ldots O_n$ referring to the observable properties of x, a hypothesis h offered as a solution of q and an accepted rule R such that the relation of confirmation or disconfirmation holds between $O_1 \ldots O_n$ and the hypothesis h.

The above explication emphasizes that objects are never historical sources "by their nature", that the set of sources at our disposal is determined by the available knowledge of natural and social phenomena and by our logic and methodology. Some of the relevant methodological rules will be discussed in the subsequent sections of this article. Our attention will be focused on linguistic sources and among these especially on cases of concrete linguistic behaviour, *scil.* on utterances.

Symbolic notation will be helpful in further analysis[3]. Let the formula

$$w_{tm}(x, z)$$

designate the phrase: "x uttered z in the spatio-temporal region tm"
Let

$$as_{tm}(x, p)$$

stand for: "x asserts p in the spatio-temporal region tm", where only statements in the logical sense may be substituted for the variable p (p may be a conjunction of sentences forming a text T).

Both formulae given above are schemes of interpretations or, synonymously, of interpretative hypotheses which result from the criticism and interpretation of written documents. If necessary, other schemes of interpretations may be introduced, e. g.

$$b_{tm}(x, p)$$

to be read as: "in tm x believes that p".

In complete interpretations the variables t, m, x, p are replaced by constants: tm by spatio-temporal coordinates, x by the individual name of the author of the text, p by the uttered statement. If the value of at least one of the variables t, m, x remains unknown, the interpretation is incomplete. Incomplete interpretations may also be used in testing hypotheses, with the possible exception of the case when p itself is unknown which makes the utterance incomprehensible.

Let us now apply the relativized concept of sources to interpretations:

Explication 2: An interpretation of the form $w_{tm}(x, z)$ is a source with respect to the question q if there exists a hypothesis h offered as a solution of q and a rule R such that the interpretative hypothesis $w_{tm}(x, z)$ either confirms or disconfirms the hypothesis h.

[3] Similar notation was used by H. S. LEONARD in the article: *Authorship and Purpose*, „Philosophy of Science", vol. 26, No. 4.

In case we are interested in the utterance of several expressions forming a text T written by an anonymous informant and want to discover who was the author, the methodologist of history gives us the following heuristic rule:

"... if the task is to discover the author... one should begin by determining the character of the author as revealed by his text in order to be able afterwards... to determine the person and the name of the author... The language of the work is of utmost importance... then follow the construction of the text, treatment of the subject and the individual approach (sympathetic, negative, objective etc.). The second stage of the procedure consists in comparing the anonymous text with other contemporary documents whose authors are known to us... "[4].

The quoted advice refers to situations in which we approach the authorship problem without any hypothesis to test. If we had one, our problem could be formulated in the form of the following decision question:

"Was N the author of the anonymous text T?"

Having no such hypothesis ready we formulate our problem in the form of a complementation question. As in many other problem situations we make some assumptions of the problem which must not be violated by any satisfactory solution. In conformity with the first part of the above heuristic rule the assumptions of the authorship problem may result from the analysis of the anonymous text and the problem may therefore be formulated as follows:

"Who qualifies as the author of the anonymous text T given the results of the analysis of the text?".

Let us distinguish two formulations of the authorship problem according to whether maximal or minimal assumptions are made:

(a) the problem with maximal assumptions:

"Given the results of the analysis of the text T:

$$f_1(y) \dots f_n(y)$$

where $f_1 \dots f_n$ are properties of the unknown author of T inferred from the text T,

find the value of x satisfying the condition:

$$x = (\imath y)\,[A(y, T) \cdot f_1(y) \dots f_n(y)]$$

where $A(x, T)$ stands for "y is the author of T" ".

The properties $f_1 \dots f_n$ may be e. g. properties referred to in the following sentences: "y was a well-educated man at his time", "y knew the works of Aristotle", "y used a language full of provincialisms" etc.

---

[4] M. HANDELSMAN: *Historyka*. Warszawa 1928, p. 141.

(b) the problem with minimal assumptions:
"Given the results of the analysis of T:

$$f_1(x) \dots f_n(x)$$

where $f_1 \dots f_n$ are properties of the unknown author of T,
indicate the elements of the set

$$(\hat{x}) f_1(x) \dots f_n(x)".$$

The above two formulations presuppose the solution of the following problem:
"By analyzing the text T determine the properties

$$f_1 \dots f_n$$

such that if x is the author of T, then $f_1(x) \dots f_n(x)$".

By $f_1(x) \dots f_n(x)$ the set of all admissible hypotheses generated by the function "x wrote T" is defined. By an admissible hypothesis is meant here any hypothesis obtained from the function "x wrote (is the author of) T" by substituting any value from the set $(\hat{x}) f_1(x) \dots f_n(x)$ for the variable x. The set $(\hat{x}) f_1(x) \dots f_n(x)$ is the set of all values of x such that we would not reject a hypothesis of the form: "x wrote T" if we tested it assuming as our criteria of rejection:

$$\sim[f_1(x) \dots f_n(x)]$$

The set: $(\hat{x}) f_1(x) \dots f_n(x)$ is, therefore, an analogue of the confidence limits in the procedure of estimating statistical parameters.

The historian tries to do all his best to derive the maximum amount of information about the unknown author from the analysis of the text, this information, moreover, being such that the set $(\hat{x}) f_1(x) \dots f_n(x)$ contains as few elements as possible. This may be achieved not only by maximizing n but also by a careful selection of the properties (*scil.* rare properties). If the information derived from the analysis of the text T is poor and therefore the set $(\hat{x}) f_1(x) \dots f_n(x)$ too "generously" determined, the number of all admissible hypotheses to be tested may become enormous and the chances of solving the problem small.

By defining the conditions necessary for the acceptance of any hypothesis of the form: "x was the author of T" and the rejection criteria on the assumption of the results of the analysis of the given text T, we simultaneously find a partial solution of the problem, what is the content of the function: "x is the author of T" (given the results of the analysis of T), and therefore the content of any admissible hypothesis. Information about the author inferable from the analysis of the text is incomplete and this is why the solution of the content problem may only be partial; it would be solved by producing an effective method of deciding of any relevant statement whether or not it is incompatible with "x is the author of T".

The second part of the quoted heuristic rule suggests that — having completed the analysis of the anonymous text, and so having solved the authorship problem on the minimal assumptions — we should compare the anonymous text with

other documents whose authors are known to us, were contemporary with the anonymous author and satisfy the condition: $f_1(x) \ldots f_n(x)$. This second stage of the procedure may be shown to be analogous in some respects with the testing of statistical hypotheses.

Let us consider an example analogous with the testing of a statistical hypothesis stating that the difference between two means is significant, i. e. when we want to decide whether or not given the results of experiments we are to reject the hypothesis that two sample means came from the same universe.

When comparing the given anonymous text with others whose authors are known to us we test in each case a hypothesis of the form:

(H_a): "T_1 and T_2 have been produced by the same author".

Let us assume that when compared both texts T_1 and T_2 exhibited a number of similarities, e. g. both contained quotations from one and the same classical author (with whose works very few were acquainted at the time), there were striking coincidences in the vocabulary and the style of both etc.

Let us, further, assume that the historian arrives at the conclusion that the similarities being as they are (numerous and with respect to rare properties) they cannot be satisfactorily explained without accepting H_a: it is namely improbable that such striking similarities should occur in the texts of two different authors.

To trace the analogy between the procedure outlined above and the testing of a statistical hypothesis that the difference between two sample means is significant, we shall consider the following example[5]:

Assume now that we want to find out whether a certain drug is effective in shortening the convalescence period of a given illness. We form two comparable groups of patients consisting, say, of 100 patients each; patients in the first group (the experimental group) are given the drug while we withhold it from patients of the second group (the control group). A record of the convalescence period of patients of both groups is kept.

Suppose that the mean calculated for the convalescence period of the control group was 19.7 days and of the experimental group 17.1 days. We ask the following question: is the observed difference of 2.6 days due to the drug or is it due to chance. We formulate the following alternative hypotheses:

H_0: The observed difference of 2.6 days is due to chance (we can expect the future difference to be sometimes positive and sometimes negative with the most likely value of about zero).
Or in other words: The two sample means came from the same universe.

[5] This is a summary of the example discussed by V. GOEDECKE in *Introduction to the theory of statistics*, 1953, p. 214.

H_1: The observed difference of 2.6 days is due to the drug.
Or in other words: There is a significant difference between the universe of treated patients and the universe of untreated patients.

The above hypotheses may be tested either by attempting to show that H_1 is to be accepted (as possibly true) or that H_0 is to be rejected (as possibly false). The latter approach has some advantages over the former. If we decide to choose it, we ask the following question: what is the probability, given H_0, of observing the difference of 2.6 days is an experiment in which two samples are selected at random. Furthermore we adopt the following rule:

R_1: If the probability, given H_0, of observing the difference of 2.6 days between two sample means chosen at random is very small, the hypothesis H_0 is to be rejected.

R_1 is a particular case of a more general rule:

R_0: A hypothesis H is to be rejected if an experimental result has been observed improbable given H[6].

What is meant by "small probability" or by "improbable" has to be made precise in each case according to the data of the problem situation. By doing so we choose a rejection region (critical region, criteria of rejection) for the given hypothesis.

It is easy to see that in our former example the historian testing two alternative hypotheses was following the rule R_0: the similarities between the texts T_1 and T_2 were in his opinion so great that the probability of their occurrence on the assumption that T_1 and T_2 came from two different authors was very small; therefore this assumption is to be rejected and its negation — H_a — accepted. The conditions necessary for the rejection of *non-*H_a vary according to the severity with which the authorship problem is approached, analogously with the choice of a larger or smaller critical region in statistical procedure.

Both texts T_1 and T_2 are interpreted as samples drawn from two universes U_1 and U_2, i. e. from two sets of texts which were written by one or by two different authors. If the similarities (or differences) between them are significant — are neither accidental nor due to imitation — we suppose that there are significant similarities (or differences) between both universes U_1 and U_2, so that they are indistinguishable and form one universe.

There are, however, differences between both examples. We have assumed that "probability" in historical terminology has the same meaning as "statistical probability", and this assumption does not seem to be too unrealistic so far as inferences of the type discussed here are concerned. But normally a historian does not assign numerical values to probabilities, merely using such phrases

[6] The inferential scheme corresponding to this rule has been termed by Z. CZERWIŃSKI „the weakened modus tollendo tollens" in the article *On the relation of statistical inference to traditional induction and deduction*, Studia Logica, vol. 7, 1958.

as "very rarely", "in most cases", "exceptional occurrence" etc. Owing to the vagueness of theses and to the vagueness of probability estimates in historical inferences, the results are highly subjective, controversial, imprecise. When, therefore, we emphasize analogies between historical and statistical inferences the difference in precision between them must not be neglected.

In the testing of a statistical hypothesis — the rejection region, once it is chosen, is unambiguously determined; in the former it is never sharply defined and one relies on intuitive appreciation, unless—of course—the authorship hypothesis is given a statistical interpretation. As a rule, the historian tests his hypotheses without unambiguously determining the region of rejection. If the criteria of rejection (and acceptance) are meant to determine the content of hypotheses (by analogy with falsification criteria), then the content of historical hypotheses is determined vaguely; there does not exist an effective method of deciding in each relevant case whether or not an experimental result is sufficient for the rejection of the given historical hypothesis.

To end this section we will note that the problems of the external criticism of sources, to which the authorship problems belong, are attacked by making some assumptions. One of the criteria of a satisfactory solution of a problem is that it does not contradict the assumptions. If it does, we decide either to reject the solution or else to change our assumptions. Sometimes a further going condition for satisfactoriness of solutions is accepted, namely that a satisfactory solution should be a satisfactory explanation of our assumptions.

III. RELIABILITY OF INFORMANTS

We are interested in the behaviour of informants using linguistic signs to communicate items of information. The fact that a person uttered a sign at a moment, is stated in an interpretative hypothesis of the form:

$$w_{tm}(x, z)$$

to be read as: "x uttered z on tm". Henceforth in our analysis the place index m shall be omitted in any interpretative hypothesis on the assumption that the time index t is sufficient to determine uniquely the event referred to in the hypothesis.

Our interest will be focused on the assertive behaviour and therefore on the interpretative hypotheses of the form:

$$as_t(x, p).$$

In particular we shall be interested in the assertive behaviour of informants in a given domain D, being a set of declarative statements on a given subject. For some purposes it will be convenient to construct the domain D of true statements exclusively (or — of statements believed to be true by historians at a given moment) and moreover of such that are selected as important or rele-

vant from some point of view; for the purposes of this discussion it will be assumed that we have a sequence of numbered statements $p_1 \ldots p_n$ where $p_i \in D$.

Let the functor "*as*" in the interpretative hypotheses satisfy the following conditions:

(1) $(x, t, p)\ [as_t(x, p) \rightarrow w_t(x, p)]$

"for every x, t, p: if x asserts p at time t, then x utters p at time t". This condition confines our attention to observable assertions exclusively; we will disregard assertions that are not communicated either in writing or in speech.

(2) $(x, t, p)\ [as_t(x, p) \dotplus as_t(x, \bar{p}) \dotplus (\overline{as}_t(x, p) \cdot \overline{as}_t(x, p))]$

"for every x, t, p: either at time t x asserts that p, or at time t x denies that p, or at time t x neither asserts that p nor denies that p".

(3) "For every x, t, p: if at time t, x's answer to the question: — is p true? — is "yes", then at time t x asserts that p".

The last condition is, strictly speaking, not "operational" when applied to historical documents: it is not usually possible to have the question "is p true?" answered by the author of a historical document. However, it may be possible to find indications that the answer to this question would have been "yes".

In the present essay we shall confine ourselves to discussing unqualified assertive behaviour, characterized by the unqualified answer "yes" to the question "is p true?". It is possible, however, to modify condition (3) as follows:

(3′): "If at time t x's answer to the question "is p true?" is "probably (possibly, maybe etc.) yes", or "people say that p is true", then at time t x makes a qualified assertion that p".

This modificatfon is superfluous if we agree to consider the qualification as a part of the sentence p that is being uttered. Alternatively if the answer to the question "is p true?" is e.g. "so people say", then we may say that what is asserted by the informant under consideration is not the statement p but the phrase "people say that p". This adds, of course, another level to the hierarchy of languages involved and likewise to the genealogy of information.

The results of the test under (3) may be unambiguous only if we know that the sentence p occurring in the question is understood by the informant in the same way we understand it. As the language in which p is formulated is not formalized we cannot appeal to its rules but rather have to appeal to the informant's behaviour. Assume, for instance, that the sentence p uttered by an informant N is the following: "*for every* x: *if* $f(x)$ *then* $g(x)$" and that N's answer to our question: "is p true?" is "yes". However, our attempts to refute P by finding an x which has the property f but does not have the property g, are considered by N to be a misunderstanding, for p is untestable for N. This shows, of course, that the result of the test (involved in condition (3) above) was ambiguous owing to the ambiguity of p.

Having arrived at interpretations of the form

$$w_t(x, z)$$

the historian faces the main problem of the internal criticism of documents. This problem may be formulated as follows:

— What relation (confirmation, disconfirmation or neither) exists between the interpretation

$$w_t(a, z)$$

and a historical hypothesis h considered as a solution of a problem q? —

We shall distinguish between two cases of the above question with reference to interpretations involving the functor "*as*":

— What relation holds between the interpretation:

$$as_t(a, p)$$

and a historical hypothesis h in which we are interested? —

In case h and p are identical with respect to their content the problem may be stated as follows:

— When is it reasonable to accept (tentatively) a hypothesis h on the ground of interpretations of the form: "x asserts at t that h"? —

To answer the last two questions the methodologist of history formulates rules in terms of reliability of informants (or, eventually, the degree of their reliability) and of independence of their assertions communicating information. The following rule is an answer to the last question:

(R_2): It is reasonable to accept (tentatively) a hypothesis h if h is asserted by at least two reliable and independent informants.

Let

$$Rel(x, DT)$$

stand for: "x is reliable in the domain D and the time period T".

Let

$$Indp\,[as_t(x, h), as_t(y, h)]$$

stand for: "x's assertion of h at t and y's assertion of h at t are mutually independent". (R_2) may then be formulated as follows:

It is reasonable to accept (tentatively) a hypothesis h if there exist x, y, t such that

$$as_t(x, h) \cdot as_t(y, h) \cdot Rel(x, DT) \cdot Rel(y, DT) \cdot Indp\ as_t(x, h), as_t(y, h).$$

$$(h \in D) \cdot (t, \in T)$$

We will now discuss the concepts of reliability and of independence which occur essentially in the above rule.

The reliability of an informant (author of an assertion) is usually characterized (a) in terms of purposeful cognitive behaviour in favourable circumstances, (b) in terms of the frequency of false assertions made by him. The reason is that purposeful cognitive behaviour in favourable circumstances and the results of this behaviour, *scil.* assertions with a frequency of false among them, seem to be interrelated; moreover, the estimation of the informant's reliability is usually hypothetical and it is rather essential in testing a reliability hypothesis to be able to choose from among a variety of consequences or — should some of them be beyond our reach — to have recourse to others. There are cases in which the historian is entirely ignorant as to the truth-value of the informant's assertions but is able, to some extent at least, to reconstruct the informant's behaviour at the time the information was collected and communicated. Quite naturally the (qualitative) appraisal of the informant's reliability is then based on this reconstruction of his behaviour and this gives the historian a corresponding estimate of the frequency of false assertions that occur in the behaviour-line. This estimate is often expressed in quasi-quantitative formulations such as: "in most cases...", "very rarely...", "more often than not..." etc. In other cases the historian knows which of the informant's assertions in the given domain have been accepted as true on the ground of other investigations and this is the basis of his estimate of the informant's reliability. So the proportion of true statements (or believed to be true) in the class of those asserted by an informant is essential for his reliability in the given domain. This possibility to make translations from the terminology of purposeful behaviour to frequencies of false assertions is welcome by the methodologist for it enables him to study the relations between the concept of reliability and other methodological concepts such as the concept of confirmation, explanation etc. defined in terms of probabilities.

Let us start now with the first approach to reliability. In this approach reliability is characterized in terms of the aims or intentions with which the assertion was made ("the author wanted to learn the truth and to communicate true information"), in terms of the means at the informant's disposal including his mental and physical capacities, his knowledge and competence in the given domain ("the author was able to learn the truth and was able to communicate true information"), in terms of the situation in which the author of the assertion acquired the given item of information and in which the assertion was made ("the author was not under any restraint or pressure when collecting the information and when communicating it").

It should be noted that there are at least two different concepts of reliability, both in terms of purposeful behaviour. The former refers to the informant's (subjective) properties exclusively, *scil.* to his aims, his knowledge, competence, sincerity etc. The latter refers both to the properties of the informant and to the (objective) situation in which the informant behaves, and which (in

the historian's opinion) is necessary for the attainment of cognitive ends. In the former case it is conceivable that a reliable informant is in a situation in which he is unable either to test the information or else to communicate it freely; then he is expected either to be silent or to make a suitable qualified assertion. If the latter conception is adopted any informant unable (for whatever reasons, objective as well as subjective) either to test his information or to communicate it freely is unreliable, no matter how honest he is and how competent in the given domain. This latter conception is adopted in the present essay, because we want to appeal to the correlation between reliability and the frequency of false assertion. According to the former conception, however, a reliable informant (i. e. honest and competent) may in extreme cases be forced by circumstances to produce exclusively false assertions.

The following explication is proposed:

Explication 3: The author of a report R_{p_n} consisting of n statements $p_1 \ldots p_n$ is reliable in the time period T (including at least the time when he was collecting the information and was uttering R_{p_n}), and in the domain D to which $p_1 \ldots p_n$ belong, if and only if in T he sincerely intended to learn whether $p_1 \ldots p_n$ are true and to communicate the true information, was physically and mentally capable to do so, was competent in D and did all his best to test $p_1 \ldots p_n$, and was not restrained by any physical or moral threat in T.

We suppose that the author of the given item of information is reliable (in DT) if we have reasons to believe that his main purpose in DT was cognitive (and he had definite hypotheses to test), that he acted purposefully and had (according to our opinion) adequate means to achieve the cognitive aim. The references to"sincere intention to learn and communicate the truth" in Explication 3 or to "cognitive aims" above are in need of elucidation. We might require either

(a) that a reliable informant be a rational agent in whose preference scale cognitive aims (i. e. the aim to test whether, say, p or p is true and to communicate the truth) are not dominated by any other aim — and we say that A_1 dominates A_2 if and only if A_2 is rejected partly or entirely whenever the realization of A_2 is incompatible with A_1; or

(b) that a reliable informant be a rational agent who in the given period (of his reliability) has no aim incompatible with the cognitive aim and dominating the latter.

According to (a) an informant is reliable only if he is truthful in critical situations when his decision to be truthful involves risk or even is sure to be cause of danger to him. It is accordingly only in such critical situations that we can test an informant's reliability. The requirement (b) is not so demanding. It labels an informant reliable if no aim incompatible with the cognitive one competes for his preference at the given period, no matter how the agent would

behave in critical situations. Criterion (a) seems attractive because of its ethical austerity. However, in order to keep as close as possible to actual usage and hesitant to impose conditions that may never be satisfied, we choose criterion (b).

As explained before, a reliability hypothesis of the form

$$x \in Rel_{DT}$$

may be tested either by attempts to show that x's behaviour was not cognitive, that x did not possess edequate means etc. or else it may be tested against the frequency of false assertions made by him. In the former case we may construct a list of conjectures, each of them incompatible with the reliability hypothesis. The following list may serve as an illustration:

$f_1(N)$: "N was interested personally in concealing or distorting the facts" (this amounts to saying that in N's preference scale non-cognitive aims were above the aim to communicate true information).

$f_2(N)$: "N was at time t a casual observer, likely to overlook or confuse details".

$f_3(N)$: "N was colourblind (and the ability to distinguish colours properly was essential in testing $p_1 \ldots p_n$ forming N's report)"

$f_4(N)$: "N did not have access to information relevant to his report"

$f_5(N)$: "As most people of his time N interpreted even simple natural phenomena as results of the intervention of supernatural forces, he expected e.g. the shapes of the clouds to be used by God as signs to communicate important messages" (this may be taken as a declaration of the inadequacy — from our point of view — of the informant's knowledge or of his lack of criticism).

$f_6(N)$: "N was aware of the fact that he would be subjected to severe penalty if he revealed the facts and he was not likely to take the risk".

The reliability hypothesis $N \in \mathrm{Rel}_{DT}$ cannot be accepted unless all conjectures on the list of its criticism are refuted. The severity or level of our criticism depends on how this list is construed. Usually it is intended to include such conjectures as would exhaust the list of factors responsible (according to our opinion) most frequently in the given domain for regularly recurring false assertions or failures to report facts, whether deliberate (lies) or not (errors). Regularly recurring errors and lies due to these most frequent factors may be termed — systematic. If the list of tests was constructed in this way and all conjectures appearing in it have been falsified, we have some ground for believing that the informant N did not posses any of the properties $f_1 \ldots f_m$, therefore that none of the frequent causes of regular errors was operative in his case, that eventual errors the possibility of which was not eliminated by our procedure could only be due to chance (rare and irregularly recurring causes) and that their frequency is small.

Presently we will outline a procedure with the help of which the errors of a reliable informant may be shown to be random.

Let C be a text written by N and describing a (temporal) sequence S of observable events, e. g. a revolution, or the performance of an experiment or a battle etc. Let $p_1 \dots p_m$ be a sequence of n true statements about S (or statements that we believe to be true and moreover to be important in some sense). By comparing C with the sequence $p_1 \dots p_n$ we are able to construe another sequence representing N's assertive behaviour expressed in C. For this purpose let 1 stand for N's assertion and 0 for his denial or failure to assert. The two sequences we are interested in may be represented as follows:

$$\text{I } p_1, p_2, p_3, p_4 \dots p_n$$
$$\text{II } 1, 1, 0, 1, \dots 1$$

We may try to predict N's errors or lies or failure to assert facts, i. e. the recurrence of zeros in sequence II, by suitable selections of the statements from the sequence I. For instance, we select all such statements that are obviously related with politics, are likely to be inconvenient for the adherents of an ideology popular at the time T when the text C was written, or else such that are easily overlooked by casual observers. If N was reliable in the time period T and in the domain D and the statements $p_1 \dots p_m$ all belong to D, then the sequence II — which is a so called alternative — is insensitive to selection according to any of the listed principles — political or ideological bias, incompetence etc. — in the sense that the frequency of zeros corresponding to p's in any subsequence of I constructed according to any of the chosen principles, is not greater than the frequency of zeros in II. The principles of selection to which the sequence representing a reliable author's assertive behaviour must be insensitive correspond to the most frequent causes of regularly recurring errors or lies in the domain D and likewise to the properties $f_1 \dots f_m$ in the list of tests of the reliability hypothesis. To the m-level of criticism there corresponds the m-freedom of the sequence representing the informant's assertive behaviour[7]. We are also entitled to say that if the informant is reliable there does not exist a predictive gambling system based on any of the principles of selection to which the alternative representing his behaviour is insensitive. If the reliability hypothesis has successfully passed all tests so devised that they eliminate all frequent causes of regularly recurring errors in the given domain D, then the errors of the informant in question are shown to be random and their frequency small, i. e. they are shown to be due to chance alone.

On the basis of what has been said so far we propose the following characterization of reliability:

[7] For the concepts of insensitiveness or freedom of random sequences see K. R. POPPER: *The Logic of Scientific Discovery*, pp. 153—163, especially p. 162.

. Assume that we are considering only a two-valued reliability function, i. e. assume that any informant is either completely reliable in the given domain and time period or unreliable altogether. This is, of course, a far going simplification of the concept of reliability as used by historians; this simplification, however, may easily be removed.

Let $h_1 \ldots h_n$, where $h_i \varepsilon D$, be a sequence of true statements, or statements believed to be true and important from some point of view, and let D be referred to as the domain of reliability of a given informant.

Assume that an informant's assertive behaviour in D may consist either in his asserting a statement belonging to D, or in denying one or in his failure to mention a statement in any form. Our sample space would consist, therefore, of three points corresponding to three possible outcomes of the informant's behaviour:

$$as_t(N, h),\ as_t(N, \bar{h}),\ \overline{as}_t(N, h) \cdot \overline{as}_t(N, \bar{h})$$

However, we decide to make a further simplification and to treat denial and failure to mention as indiscernible, *scil.* as cases of error (or lies). This enables us to use two symbols only: 1 for N's assertion and 0 for his denial of any statement belonging to D or for his failure to mention one; N's possible line of assertive behaviour in D will then be represented by a sequence 1,0 ... 1. The frequency of zeros represents the frequency of N's errors or lies; let F stand for this frequency.

Let α be a small number, the probability of errors due to chance alone (as explained before) and characteristic of a domain D.

Df 1: If a long behaviour line of N is given in DT, then N is reliable in DT if and only if

(a) $F(0) \leqslant \alpha$,
(b) zeros in the behaviour line are random.

A few comments may help to understand the above definition: Errors of a reliable informant are said in it to be random and rare and he himself is informative in the sense that he rarely fails to mention a true and relevant statement in the field of his reliability. The laws of reliable assertive behaviour, however, according to Df 1 operate in longer behaviour lines. If we were in posession of one assertion of an informant N only and knew it to be true, this would by no means indicate that N is reliable, although the frequency of his errors in this short behaviour line equals zero; a silent fool must not be mistaken for a sage or a notorious liar for a truthful person on account of one single truth pronounced.

The more the number of errors in the informant's behaviour line excedes the limit α, the more his assertive behaviour is independent of the reality he pretends to be describing. In so far as his errors are predictable in that far they

may be shown to depend on factors other than the reality referred to in the assertions. As special cases of such dependence which is to be discovered in the procedure of testing the randomness of errors described before, we may mention the following two examples:

Let $A_{MD'}$ be the behaviour line of M in D' and let $D' \neq D$. If N's errors in D are predictable on the basis of $A_{MD'}$ then N is unreliable in D.

Let A_{ND} be the behaviour line of N in D and let D' be the extension of D such that $D' \neq D$. If N's errors in D' are predictable on the basis of A_{ND}, then N is unreliable in D'.

If N's errors are not numerous (the limit $\varkappa$ is not exceded) but are systematic, it is possible to determine the sub-domain D_s in which they occur and so obtain the new reliability domain D-D_s such that N is reliable in it in the sense of Df 1.

One of the consequences of Definition 1 may seem objectionable. If it is the frequency of zeros that is essential for the reliability of an informant, then a denial of a statement belonging to D and a failure to mention one are of equal importance; similarly, errors and lies are of equal importance among themselves. This simplification is partly neutralized, we may hope, by the fact that the statements belonging to D are not only true (or believed to be true) but also important or relevant from some point of view. Besides, this simplification might be removed by suitable weighting of errors and lies and by counting some of the zeros in the behaviour line twice, three times or n-times.

Our Definition 1 refers to relations that hold between the behaviour lines of informants and the domains of reliability, conceived as sets of true statements. However, in actual inferences in which the concept of reliability appears the information on either the behaviour line or the domain D is incomplete. Sometimes the unkown part of D is extrapolated on the basis of the known behaviour lines, but it may be the other way round as well. The estimation of the parameter α, discussed before in connection with testing the reliability hypothesis, determines ambiguously the behaviour line of the informant in DT. Strictly speaking, we should use the expression $\varkappa$-reliable. The level of reliability, the level of criticism with which the reliability hypothesis was tested and the level of freedom of the informant's behaviour line in D, are all simultaneously determined. In practice numerical values are not assigned to these levels, but it is conceiveable that we might do so e. g. we might at least assign ordinal numbers to various lists of tests devised to refute the reliability hypothesis and expect the frequencies of errors to vary accordingly, so that the higher the number of the list the smaller the frequency of errors. Anyway, it should be borne in mind that the presented Definition 1 defines a simplified concept. It does justice to some at least ideas behind the actual usage of the concept of reliability. It cannot do to all, for the simple reason that some of them are inconsistent. In conjuction with Explication 3 the Definition 1 provides a the-

oretical framework relating observable assertive behaviour with hypothetical reconstruction of the informant's aims, knowledge and other means possessed by him and with the situation in which he acted, including the actual states of affairs referred to in some way in his assertions.

The following are some of the methodological rules or procedural schemes in which the concept of reliability occurs, one of them is our rule R_2 formulated at the opening of this section:

(R_3) If N is reliable in DT and h_s belongs to D, then it is reasonable to expect that N asserts h_s in T.

(R_3) allows to expect that N asserts h_s in T, provided the antecedent of the above theorem has been satisfied. The procedural scheme corresponding to this rule may be termed the *weakened modus ponens*.

(R_4) Assume that N has denied (negated) in T n statements belonging to D.

If N were reliable in DT, then the probability of his denying a (true) statement belonging to D would be small.

Therefore we may reject the conjecture that N was reliable in DT.

(R_5) Assume that a number — at least two — of informants reliable in DT communicate their information about D independently of each other; let N, M be their names and let D_m designate the statements in D asserted by M. We define:

Df 2: $Indp\ [as_{t_1}(N, h_i),\ as_{t_2}(M, h_i)] \equiv {}_{ND_m}F(0) = {}_{ND}F(0)$

Assume further that there exists a statement h_r not mentioned in any way in either of the behaviour lines of our informants and such that were it true it would belong to D (but not necessarily its negation $\bar{h}_r$).

The probability of a reliable informant's failure to mention a statement belonging to the domain of his reliability is — on the Definition 1 — very small; the probability of several independent and reliable informants failure to mention such a statement is even smaller.

Therefore it is reasonable to suppose that *non-*h_r is true.

The procedure of rejecting a hypothesis h_r according to the above scheme has been termed by historians argumentum ex silentio. It may be presented as an application of the rule (R_0) or the *weakened modus tollendo tollens*.

(R_6) Assume, as in the former scheme, that a number of informants have failed to mention a statement h_k which belongs to a domain D and which we know to be true.

Were the informants reliable in D, the probability of their failure to mention h_k would be, according to Definition 1, very small.

Therefore it is reasonable to suppose that the informants in question are unreliable in D, in the sense that the probability of their failure to mention h_k

is not small, or even that this failure is subject to a law that may enable us to predict their "silence".

The above is a variant of argumentum ex silentio: in both R_5 and R_6 the "silence" (failure to assert) of some informants improbable on some hypothesis is the ground for the rejection of that hypothesis; in R_5, however, the informants are reliable and the hypothesis is a statement that does not occur in their behaviour lines; in R_6 on the other hand the rejected statement is the reliability hypothesis.

The following inference may serve as an illustration of R_6: All inscriptions on the monuments left by the ancient Mayas were found to be dates of events; none of these inscriptions communicates any information on, say, the religion of Mayas, their political and social institutions, their food, their habits etc. It would be absurd, of course, to suppose ex silentio that all these phenomena were absent in the lives of Mayas. Rather we suppose that the Mayas never communicated such facts in writing and that therefore they were not reliable informants in any domain except, may be, in dating events.

(R_7): An informant may be unreliable in several possible ways. Therefore we cannot predict his assertive behaviour in D unless we possess a suitable transformation rule which is the law of his behaviour and which determines a correspondence between the set of statements accepted by us to be true and the elements of the set of statements asserted by the unreliable informant. Such a rule may say, for example, that N flatly denies everything that is against his interest; or that he exaggerates the achievements of his countrymen or that he avoids discussing religious or political matters etc. Some rules are, of course, more specific than those mentioned; imitation relations may serve as examples. On the assumption that *Einhard*'s description of *Charlemagne* is an imitation of the portrait of *Augustus* given by *Svetonius*, we establish a correspondence between *Svetonius*' statements characterizing *Augustus* and *Einhard*'s statements on *Charlemagne*. Given this rule we could predict *Einhard*'s assertive behaviour in the description of *Charlemagne* independently (to some extent at least) of what *Charlemagne* was like in fact. This is why *Einhard* is unreliable so far as his account of Charlemagne is concerned and why this account is not a historical source.

(R'_2): The rationality of the methodological rule (R_2) which allows to accept as a fact something that is asserted by at least two reliable and independent informants is established by Definition 1 in the way we appealed to it presenting (R_5). Thus R_2 may be shown to be an application of the more general methodological rule R_0, provided the reference class D is constructed as a sequence of true negated statements $\bar{h}_1 \ldots \bar{h}_n$. The procedural scheme is as follows:

The probability that several (two at least) informants reliable in D will independently deny a (true) statement belonging to D, is — on Definition 1 — the definition of *Indp* and the multiplication theorem, small.

If, however, they have denied a particular statement h_s which, were it true would belong to D, then it is reasonable to accept that h_s is not true and its negation is.

Sometimes it is insisted that more than two reliable authors are necessary in order that the rule R_2 should be applicable to their independently asserting a given fact, two coincidences, it is argued, are quite common. It is only accumulation of coincidences that is sufficiently improbable.

To end this discussion of some methodological rules involving the concepts of reliability and of independence, a brief comment on the latter of the two concepts. Independence of two informants is usually assumed not on the basis of calculating the frequencies of their errors according to the definition of *Indp*, although this does not seem to be impossible; the common procedure is to test an independence (or dependence) hypothesis against textual similarities or differences in the usage of vocabulary, in style etc., similarly to testing an authorship hypothesis as presented in the former section of this essay. This procedure is, therefore, analogous to the testing of a statistical hypothesis which states that the difference between two sample means is significant.

The rule R_2 is an answer to the question: When is it reasonable to accept (tentatively) a hypothesis h on the ground of accepted statements of the form: "X asserts at time t that h"? This amounts to saying that R_2 states the conditions under which a hypothesis h is confirmed by interpretative hypotheses, $as(x, h)$, referring to assertive behaviour of informants. These conditions may be formulated:

(a) in terms of the reliability of informants and of independence of their assertions,

(b) in terms of purposeful cognitive (testing) behaviour, competence of the agent and adequacy of his means,

(c) in terms of the probability of the occurrence of certain assertions involving the hypothesis, given this hypothesis.

(d) in terms of satisfactory explanation.

The last point (d) will now be briefly discussed. Following K. R. POPPER let us by an explanation in terms of situational logic mean an explanation of a person's behaviour in terms of his aims, means at his disposal and of the situation in which he acts. Given the meaning of reliability as explicated above, a reliability hypothesis refers to the logic of the situation in which an assertion was made and therefore qualifies as an explanation of the assertive behaviour in terms of the logic of the situation.

Given a case of assertive behaviour stated in the interpretative hypothesis $as_t(N, \bar{h})$, the historian has to consider four possibilities which might explain it (or its result); these are written on the right side of the following equivalence:

$$as_t(N,\bar{h}) = h \cdot \overline{Rel}(N,DT) + \bar{h} \cdot \overline{Rel}(N,DT) + h \cdot Rel(N,DT) + \bar{h} \cdot Rel\,N,DT)$$

where $t \in T$ (for the sake of brevity h is written instead of $h \in V$, i. e. instead of "h is true").

In order to decide which of the four alternatives may be a satisfactory explanation of the fact that a asserted $\overline{h}$ in t we need criteria of a satisfactory explanation. By slightly modifying K. R. Popper's criteria of explanation we will accept the following conditions[8]:

1) a hypothesis h is not a satisfactory explanation of an event e if h is not refutable (i. e. if there exist no falsification and no rejection criteria for h);

In particular, we shall refuse to accept as satisfactory an explanation involved in the statement of the form: "N acted as he did because he (sometimes) behaves irrationally"; this condition and the rejection of the view that there exist cases of behaviour inexplicable "because of the irrationality of human nature" may be termed the rationality principle.

We decide likewise not to explain events by assuming that they resulted from an accumulation of accidents;

2) if a hypothesis h has been falsified, then it may no longer be a satisfactory explanation;

3) if the probability of e given $\overline{h}$ is great (and thus the probability of e given h is small), then $\overline{h}$ may be a satisfactory explanation of e;

if the probability of e given h is small, then h is not a satisfactory explanation of e;

4) if h is a satisfactory exlanation for e, then there exists independent evidence for h;

Having ascertained the reliability of the given informant, a, the historian rejects the first two components of the disjunction of considered explanations: they do not qualify as satisfactory explanations being falsified (rejected) statements (Cf. condition 2). There remain the third and the fourth component of the disjunction. The 3d one, however, must also be rejected: for if N is reliable in DT, which was assumed, the probability of his denying a true statement in DT is small (Cf. condition 3); we have much stronger grounds for its rejection if there are several independent assertions. It is only the fourth part of the disjunction on the right side of the equivalence that may be a satisfactory explanation of the assertive behaviour, $as_t(N, \overline{h})$, on the assumption that the informant N is reliable. Condition 4 is satisfied only in case there are several independent testimonies.

We reformulate the rule R_2 as follows:

"It is reasonable to accept (tentatively) a hypothesis h on the ground of inter-

---

[8] K. R. POPPER: *Philosophy of Science: Personal Report, British Philosophy in the Mid-Century*, p. 187; *The Aim of Science*, Ratio, 1, 1957. *The Open Society and its Enemies*, 1957, vol. 2, p. 265.

pretations of the form $as_t(x, h)$, if the statement" h is true" is an element in the logic of the situation in terms of which the fact that $as_t(x, h)$ is satisfactorily explained".

To complete the list of comparisons and analogies between various methodological concepts it should be noticed that an informant may be regarded as an operator, the fact which he observed or wanted to learn in some way — an operand and the assertion we find in a historical document — a transform, in the sense these terms are used in cybernetics. (I have used the concept of transformation before, when discussing the bahaviour of an unreliable author). The proposed Definition 1 and the rules based on it may be interpreted as defining a behavioural relation between two points — the informant and the historian or whoever interprets the given text — and so creating a channel of communication which more often than not is affected by "noise" distorting the information. An important part of the normative methodology of historical criticism is a theory of maximizing information from unreliable sources. If an informant is unreliable and we have not discovered a transformation rule relating his assertive behaviour with the state of affairs in which his behaviour originated, so that to any statement h_i asserted by N there might correspond in reality either h_i or $\bar{h}_i$, then N's assertive behaviour does not communicate any information about that reality.

In an article entitled *A Generalization of the Refutability Postulate*[1] I have proposed a rule restricting the confirmability of hypotheses by facts compatible with them and have introduced the concepts "reliable and unreliable indicators". Evidence given by unreliable informants is never a reliable indicator of the truth of a hypothesis. The concept of reliability serves in the methodology of historical criticism to impose certain conditions on the confirmability of historical hypotheses by the evidence given by informants.

The concept of reliability of informants is, of course, applicable outside the criticism of historical sources. The scientific value of the results of the testing of scientific hypotheses depends on the reliability of those who test the hypotheses. Not only because the results of single experiments performed in the course of testing an universal hypothesis have to be observed and reported by somebody (*scil.* an informant), but also because "confirmation of the hypothesis by results of experiments" or "the weight of evidence" seem to depend on the pragmatical factor of competence, ingenuity, sincerity and honesty of the scientist who devised the whole testing procedure with the help of which the evidence was acquired.

Allatum est die 9 Decembris 1960

[1] J. GIEDYMIN: *A Generalization of the Refutability Postulate*, Studia Logica, vol. 10, 1960..