Chapter 15
Biomedical Inference: A Semantic Model

Jan Doroszewski

Department of Biophysics and Biomathematics, Medical Center of Postgraduate Education, Marymoncka 99, 01-813 Warszawa, Poland
jandoro@cmkp.edu.pl

Summary. This chapter presents a model of the structure of a typical fragment of pathophysiological and diagnostic knowledge and of simple steps for biomedical problem solving. The main method used in this study consists of a structural and semantic analysis of the biomedical statements that appear in original clinical textbooks and documents. Various transformations of these statements reveal their different aspects as causal structure and meaning (causal paraphrase) and a simplified logical skeleton (propositional paraphrase), as well as temporal and uncertainty factors. The possibilities of performing some inference-making procedures based on the results of this analysis are shown. Biomedical types of understanding and use of the notions of causal relationship and probability are discussed from the point of view of their roles in biology and medicine. It is suggested that in a biomedical context, an especially important role is played by the logical type of probability, i.e., estimation of the degree of certainty of statements based on their contextual justification. The process of biomedical hypotheses verification is based, to a considerable degree, on the estimation of changes in their probabilities (confirmation) in the light of increasing evidence. This approach, called a semantic model of biomedical cognition, is exemplified by the analysis of numerous original statements and a study of a handbook text fragment, as well as by a discussion of an authentic clinical case.

1 Introduction

This chapter describes the results of a model approach to the study of some aspects of biomedical knowledge, namely, those that are especially important from the point of view of solving diagnostic problems. It is a continuation of previous research in which biomedical knowledge was studied in the systems' and semantic framework [15–17]. The approach on which this study is based consists mainly in analyzing natural-language medical texts as reflecting the fragments of natural medical knowledge. It is supposed that the results of such a study may contribute to better understanding of biomedical reasoning, perhaps also to the development of a more formal representation of biomedical knowledge in the framework of artificial intelligence.

Although the literature on various aspects of medical knowledge and problem solving, theoretical as well as practical, methodological, psychological, and others, is vast and manifold [3, 10, 12, 20, 22, 31, 43, 44], the works directly related to semantic biomedical problems are not so numerous [2, 9, 14, 25, 37, 40, 46]. This study was partly inspired by research on qualitative reasoning [26, 27] and some aspects of AI knowledge representation [36, 45]. The analytical model character of this approach should be firmly stressed since neither the methods applied nor the results obtained are directly connected with research on the psychology of thinking, in which considerable progress constantly occurs (e.g., [21, 23, 28, 39, 41]).

Biomedical knowledge is considered here as a system of statements describing features of the human organism and its parts. Due to the extreme complexity, as well as the dynamic and variable character of biological and pathological objects, the statements describing them are, on the one hand, interconnected in a multiple way and, on the other hand, denote, in an overt or a hidden way, the temporal aspects of phenomena and the degree of certainty of judgments.

One of the most specific features of practical biomedical knowledge and reasoning is its close connection with natural language and relatively limited use of formal methods of inference making in everyday problem solving, despite the rapid development of applications of mathematical methods. However, to better understand the meanings of biomedical statements, they should be extricated, as it were, from their natural linguistic surroundings.

In the search for the basic elements of biomedical statements, the method of simplifying but sense-preserving graphs and paraphrases representing linguistic expressions is useful. It resembles the logical approach to the analysis of natural language (e.g., [30, 32, 34]), but its purpose is limited here only to medical language. Theoretical structures revealed in this way may be used as a base for schemes of reasoning and applied to simplified inference making.

Scientific domains, such as medicine as a whole, specialties, for example, cardiology, problems or fragments of knowledge (e.g., description of a disease), and the like are systems of laws and other general statements that describe objects as characterized by their features. Such features are first-order statements [17], e.g., anatomical and relations connecting the features of objects and their values, which are second-order statements [17], e.g., physiological. In this work, the latter category of scientific statements is studied, especially those that describe relations linking particular values of features.

1.1 General Description of the Disease

This study is centered on the medical cognition of the patient, based on a general model description of the disease, which is composed of statements representing pathological and other phenomena and their interrelations. The reasoning of the doctor aims at making a mental reconstruction of unobserved facts (signs) on the grounds of those that have been observed. A typical set of pathological phenomena (composed of a model of a disease, or nosological unit) that are the correlates of the biomedical statements may be represented as in Fig. 1.

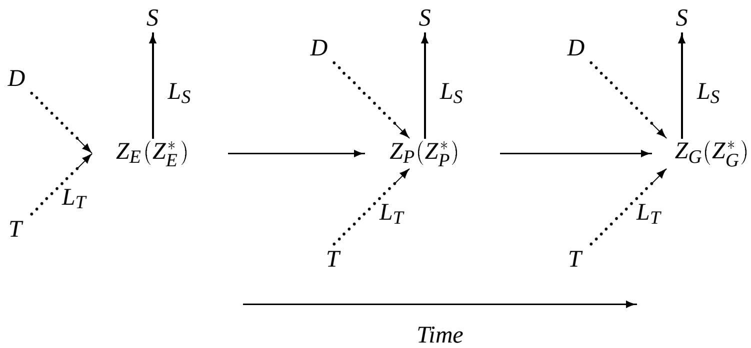

Fig. 1. A scheme of disease in general terms (a fragment of biomedical knowledge). The symbols may be interpreted as denoting the phenomena or corresponding names. Z_E: etiologic factors; Z_P: main (specific) pathological phenomena (direct and indirect effects of etiologic factors); Z_G: effects of specific pathological phenomena (Z_P) concerning the organism as a whole; *S*: observable phenomena (signs). Arrows: relations linking the phenomena (asymmetrical, mostly causal). *D*: diagnostic actions causing the appearance of signs; *T*: therapeutic actions potentially transforming undesired pathological phenomena into desired ones (Z_E^*, Z_P^*, Z_G^*).

1.2 A Biomedical Statement

A specific feature of pathological (clinical and others) statements is that they describe the relations between pathological values of the human organism and its particular parts. It should be remembered, however, that their explicit meaning is only a small part of overall, mostly normal physiological and anatomical knowledge.

The basic type of statements that are analyzed in this work may be regarded either as sentences composed of the grammatical object and a complex predicate or as statements containing two or more clauses linked by a relation; one of the clauses denotes the main object, and the other(s) describe(s) secondary ones; still other(s) represent the relation(s). The statements of this kind describe the relations between two or more values of the features exhibited by the objects; their extremely simplified general structure (neglecting the modal auxiliaries) may be summarized as follows: (1) *Functional form*: "for every object *x* (some object *x*, *f*% objects *x*) in conditions W, the values of feature E_1 of the object E are linked with the values of the feature E_2 of this object by the relation R." (2) *Particular-value form*: "for every object x (some object x, f% objects x) in conditions W, the feature E^1 of the object x is related to the feature E^2 of this object by the relation R_{12}." As examples, the statements analyzed below may be cited: (3) "*the oxygen supply* (feature E_1) *in the tissue (object x_1) depends (functional relation R) on the blood flow* (feature E_2) *in the artery* (x_2); " (4) "*the decrease in the blood flow* (E_2^1) *in the artery* (x_2) *is the cause* (R_{12}) *of the decrease in the oxygen supply* (E_1^1) *in the tissue* (x_1)." The objects x_1 and x_2 are linked by a certain spatial relation.

The professional-language statements, according to widely accepted usage, contain the phrases signifying the time and uncertainty factors in special situations rather than as a general rule. The reason probably is that these notions are so deeply embedded in medical science and the way of reasoning that they frequently seem self-evident and are automatically supplemented, as it were, by the participants in the communication process. To consider these factors in an explicit way, the above statement describing the relation between particular values (form 2), should be reformulated as follows: "for every object x (some object x, f% objects x) in conditions W, the value E_i^1 of the feature E^1 of the object E occurring in time t is connected by the relation R_{12} with the value E_j^2 of the feature E^2 of this object occurring in time t' with probability π." Besides the artificiality and complexity of such a statement, its precision and clarity are rather doubtful. Perhaps the difficulty of formulating too complex a linguistic expression is the reason that various levels of their sense are separated in practice: the basic, "core" meaning, on the one hand, and modifying factors, on the other hand. Both of them are, to a certain extent, though in a different manner, contained in the significance of a special kind of relation that may link the phenomena as cause and effect. The time sequence is one of the constitutive conditions of the biomedical causal relations (see Sect. 2) and therefore, may not be treated separately. On the other hand, the probability of the occurrence of the phenomena, even if connected with the notion of causality, is a factor that should be analyzed apart (see Sect. 2).

In comparison with many natural professional-language expressions, the above schemes of a sentence are, of course, an obvious simplification, possibly even an over-simplification. Take, e.g., an authentic sentence: "*Experimental studies suggest that the endothelial barrier, broken by mechanical or chemical injury, is associated with a tissue response that includes local platelet adhesion and aggregation* [5]." It would be impossible to render the sense of this or a similar sentence in the above scheme. A sequence, however, of several sentences (paraphrases) could convey its meaning in a relatively exact way. The above approach seems to be especially suitable for analyzing the description of pathological phenomena contained, e.g., in clinical textbooks, which present fragments of biomedical knowledge in a manner, to some extent, prepared for practical use (see Sect. 4). The remarks suggest that to make a biomedical statement clear and unambiguous (formulated, as normally, in natural linguistic shape), i.e., to express it in an overt form, it is necessary to enrich it with certain elements or to modify its wording, or both. It would be, however, an impossible task to attempt to express the whole meaning of a statement, since it would be equivalent to setting forth a fragment of knowledge of considerable size. As concerns the specific purposes of this analysis, it is sufficient, however, to show more clearly, i.e., to extract from a deeper context, only certain aspects of the sense of the statement, e.g., the time factor involved, the degree of certainty of the judgment, etc. or to modify its structure so that it could be more suitable for use in a certain type of inference. This purpose may be achieved by making various paraphrases of the analyzed statement the main types of which either express the kind of relation or simplify it. Inasmuch as one of the especially common and important types of relations described by biomedical statements is the cause-effect connection, the former transformation may be called a *causal paraphrase*; on the other hand, to show a certain type of model inference, a change based on the propositional (sentential) scheme, *propositional paraphrase*, may be useful. Besides the linguistic transformation, the use of a graphic representation is useful for semantic analysis.

1.3 A Fragment of Biomedical Knowledge

Below we present an analysis of quasi-natural statements in the frame of semantic model (a study of an authentic text is presented in Sect. 4). To be closer to real biomedical knowledge, the statements comprise more than two arguments.

Let us consider the following statements (uppercase and lowercase letters are assigned to the components of the sentences for further use).

1. *Quasi-natural statements in functional form* (a description of the relation linking the features): Oxygen content in a tissue (Q_1) depends on the blood flow in the supplying artery (P_1), arterial blood oxygen saturation (P_2), oxygen consumption in the tissue (P_3), and other factors. The supply of glucose (Q_2), the action of hormones on this tissue (Q_3), etc. depend also on arterial blood flow (P_1).

2. *Quasi-natural statements in a causal form* (a description of the relation linking particular values of features: A decrease in the oxygen supply in a tissue (q_1) is caused by a decrease in the blood flow in the supplying artery (p_1) and/or a decrease in the arterial oxygen content (p_2) and/or a decrease in oxygen consumption in this tissue (p_3) and/or by other factors. The decrease in blood flow in an artery (p_1) causes a decrease in the oxygen supply in a tissue supplying this tissue (q_1) and/or a decrease in the glucose supply (q_2) and/or a decrease in the action of hormones on this tissue (q_3), etc. 3. *Propositional form* (a description of the relation linking particular values of features by conditional sentences): If the blood flow in an artery (p_1) and/or arterial oxygen content (p_2) and/or oxygen consumption in this tissue

(p_3), etc. are decreased, then the oxygen supply in the tissue supplied by this artery (q_1) is decreased. If the blood flow in an artery (p_1) is decreased, then in the tissue supplied by this artery, the oxygen supply (q_1) and/or glucose supply (q_2) and the hormone action (q_3), etc. are decreased. In an abbreviated form:

1. If p_1 and/or p_2 and/or p_3 and/or ..., then q_1.
2. If p_1 then q_1 and/or q_2 and /or q_3 and/or ...,

where the functor and/or may be interpreted as alternative, disjunction, or conjunction.

The above sets of statements (a knowledge fragment in various linguistic shapes) may be represented as directed graphs (see Fig. 2; see also Figs. 1 and 3).

These formulations do not contain expressions indicating the time in which the phenomena occur and their degree of certainty. The time factor is implicitly assumed in the causal relation (see Sect. 2); in some cases, however, time requires an apparent expression. The same may regard the probability of the statement (see Sect. 3); more frequently, however, this factor is represented by a modal part of the sentence, e.g., "the phenomenon p_1 is a *probable cause* of the phenomenon q_1" and "the phenomenon q_1 is a *probable effect* of the phenomenon p_1." This problem will be discussed later.

The propositional representation of the statements may be useful for revealing the basic scheme (a skeleton, as it were) of various inferences; it may also show more clearly the temporal and probabilistic aspects of sentences. The statements presented as propositional paraphrases may constitute grounds for the following inference: (1) p_1 occurs; therefore q_1 and/or q_2 and/or q_3 and/or ... occur, (2) p_1 and/or p_2 and/or p_3 and/or ... occur; therefore q_1 occurs. In practice, the usefulness and effectiveness, even feasibility, of such complex reasoning is limited, and it is seldom performed (in any case, in an overt form). To arrive at a valuable conclusion, various assumptions should be adopted not only to introduce a simplification but also to make clearer some aspects of the statements, especially the meaning of the functor and/or in a given context.

An extreme simplification would consist in assuming that the influence of most of elements may be disregarded because either their insignificance is known or because they are unknown, etc. This assumption reduces the whole fragment of knowledge to the sentence: "p_i is the only cause of q_j" or "q_j is the only effect of p_i." In a propositional paraphrase, it would run: "p_i if and only if q_j" that would make possible a two-sided inference, namely, "p_i is accepted; therefore q_j," and "q_j is accepted; therefore p_i." In most situations, however, such a strong assumption would be as unjustified as taking into account all possibly relevant factors is unrealistic.

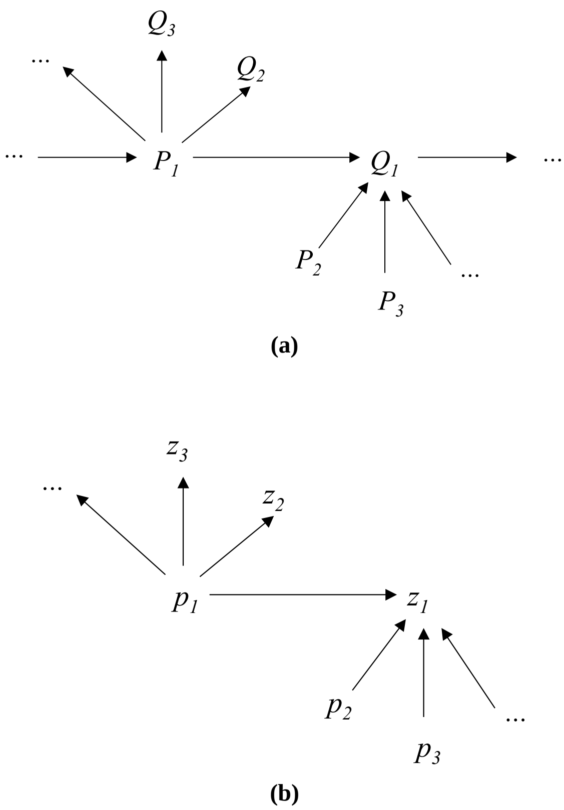

Fig. 2. (a) Graphic representation of the functional form of a knowledge fragment. P_i and Q_j denote the names of features; arrows denote a functional (multivalued) relation. **(b)** Graphic representation of the causal form of a knowledge fragment; p_i and q_j denote the names of the values of features P_i and Q_j, respectively; the arrows denote the causal relationship between these values.

The solution seems to be, in practice, to perform the reasoning without deciding in an explicit manner which elements are considered and which are disregarded. It is a kind of counterfactual thinking, of pondering "as if," or of drawing a conclusion with a certain doubt, being ready to modify it if new evidence appears. This hypothesis is corroborated by the common use of statements that are not clear in their semantic range. For example, the sentence "the phenomenon *p* is the cause of the phenomenon *q*" ("a decrease in the oxygen supply in the tissue is caused by a decrease in the blood flow in the artery") or "the phenomenon *q* is the effect of the phenomenon *p*" ("a decrease in the blood flow in the artery causes a decrease in the oxygen supply in the tissue") may be understood as signifying that *p* is the unique cause (effect) of *q* or that *p* is one of the causes (effects) of *q*. Similarly, in natural language, the statement equivalent to the conditional sentence "if *p*, then *q*" may mean "only if *p*, then *q*", "if *p*, then only *q*," etc. A conclusion based on such statements, "accepted p, therefore q" or otherwise, may be considered correct only with the above-mentioned reservation, i.e., on condition of its provisional character. The possibility of modifying the result of reasoning is assured by the fact that each statement constitutes a part of a broader fragment of knowledge that may be a source of new pieces of evidence. It seems that reasoning surrounded by a "penumbra of doubt," understood as probability or otherwise, is a normal procedure in medical problem solving.

Among the many assumptions that may constitute more or less firm grounds for accepting a conclusion in complex reasoning, one is especially important and frequently used in practice. It consists in accepting tacitly a supposition concerning the values of other features, which are components of a set of premises appearing in a given inference. For example, the conclusion that the oxygen supply in a tissue is decreased when the arterial blood flow is decreased is correct, only if it is assumed that either the causes are independent or that other features that (jointly or alternatively) are the causes of this condition display certain values (are normal, are not increased). Such an assumption would constitute the condition of the validity of the reverse inference, i.e., that the arterial blood flow is decreased based on the fact that the oxygen supply in this tissue is decreased.

In natural discourse, the modal auxiliaries are expressed in an almost infinite manner (see Sect. 3). In many cases, they are taken as self-evident and are not specified at all. On the other hand, the notions of temporal sequence and uncertainty or probability are, to a certain extent, implied in the meaning of the concept of causal relationship (see Sect. 2).

The full meaning of a causal statement is rich and complex. For the time being, it may be assumed that the statement "the phenomenon p is the cause of the phenomenon q" means that p and q occur in certain times t and t' and that the occurrence of q as the effect of p is not necessary. The paraphrase may be formulated as follows: *if in time t the phenomenon p occurs, then in time t' the phenomenon q occurs with probability* π (probably occurs), where $t = t'$ or $t < t'$ or $t > t'$; π is a qualitative estimate or quantitative measure of probability. Another formulation expresses the notion of the probability of statements rather than that of events: *if in time t, the phenomenon p occurs, then in time t', the phenomenon q occurs is probable (justified) in grade* π (see Sect. 3). According to the context, the expressions "*probable*" or "*in grade* π" may have a quantitative or a qualitative interpretation, i.e., may be equivalent to a certain number, to an adjective in a basic or comparative form, or may indicate the revision of an estimate in the light of new evidence. The inference based on the sentence *if $p(t)$, then $q(t')$ with probability* π, $t < t'$ runs: $p(t)$ is accepted, therefore q occurs (simultaneously as or later, or earlier than q) with probability π, where the phrase "probability π" may be understood in each of the above manners.

Propositional paraphrases, whose meanings are enriched by the time and probability aspects, i.e., in fully developed form, are the following:

1. If $p_1(\mathrm{t}_1)$, and/or $p_2(\mathrm{t}_2)$, and/or $p_3(\mathrm{t}_3)$, and/or ... , then $q_1(t')$ with π, where the relations of times depend on the interpretation of the functor and/or.
2. If $p_1(\mathrm{t})$, then $q_1(\mathrm{t}')$ with π_1 and/or $q_2(\mathrm{t}_2)$ with π_2 and/or $q_3(\mathrm{t}_3)$ with π_3 ... where the relations of times are as before.

A fully developed paraphrase, as well as the paraphrases that don't contain modal auxiliaries (see above) are sometimes reduced to quasi-deterministic conditional sentences: "if $p_i(\mathrm{t})$, then $q_j(\mathrm{t}')$ with π_1" and "if $q_j(\mathrm{t}')$, then $p_i(\mathrm{t})$ with π_2"; as before, a "provisional" conclusion may be drawn, namely, $q_j(\mathrm{t}')$ with π_1, and $q_i(\mathrm{t}')$ with π_2.

1.4 Inference Model

The above presented analysis may be summarized as follows:

An example of a rather typical, though not very precise, formulation of a biomedical (pathophysiological) statement may be put in words in the following manner: *a decrease in the blood flow in the artery supplying a tissue* (p_1) *is accompanied by a decrease in the oxygen supply in this tissue* (q_1). It could be used as the general premise in two kinds of inferences: (a) the blood flow in the artery is decreased (p_1), therefore probably the oxygen content in the tissue is decreased (q_1); (b) the oxygen content in the tissue is decreased (q_1), therefore probably the blood flow in the artery is decreased (p_1). In natural reasoning, both inferences would probably be accepted. The correctness of the above conclusions depends, however, on other connected statements and may be verified in various ways. Let us transform the initial statement into causal and propositional forms (paraphrases).

Effect-cause paraphrase
The decrease in the oxygen supply in a tissue (q_1) *is caused by a decrease in the blood flow in the supplying artery* (p_1).

Paraphrase of the above form:
If p_1 *occurs, then* q_1 *occurs with probability* π.

Inference:
p_1 occurs; therefore q_1 occurs with probability π; in words, *the blood flow in the*

artery is decreased (p_1)*; therefore probably (with probability* π*) the oxygen content in the tissue is decreased* (q_1).

This conclusion is arrived at without taking into consideration other than p_1 possible causes (see Fig. 2). Whether it is true or false depends on other phenomena, namely, the decrease in the oxygen saturation of the blood (p_2) and the increased oxygen consumption in the tissue (p_3). If the relation linking all these phenomena is interpreted as a weak alternative, i.e., if p_1 or p_2 or p_3 or ..., then q_1; probability π of q_1 increases when p_1 and p_2, when p_1 and p_2 and p_3 ... then the conclusion (a) is correct; taking into consideration the possible or actual appearance of p_2 and/or p_3 would only modify (enhance) the probability of the conclusion (see later).

Cause-effect paraphrase:
The decrease in the blood flow in the supplying artery (p_1) *is the effect of the decrease in the oxygen supply in a tissue* (q_1).

Paraphrase of the above form:
If q_1 *occurs, then* p_1 *occurs with probability* π'.

Inference:
q_1 occurs; therefore p_1 occurs with probability π'; in words, *the oxygen content in the tissue is decreased* (q_1)*; therefore, probably (with probability* π'*) the blood flow in the artery is decreased* (p_1).

Whether this conclusion is true or false depends on the interpretation of the functor and/or. For the general assumption, as before, if p_1 and *not*(p_2) and *not*(p_3), therefore p_1; the probability of p_1 is modified (decreases) when not only p_1 appears but also p_2, and p_1, together with p_2 and p_3.

2 Biomedical Aspects of the Notion of Causality

Causality constitutes one of the most important philosophical and logical problems of the general methodology of science (see, e.g., [5,6,29]), including the theory of computer science and artificial intelligence [33]. The analysis presented in this work concerns only limited and rather special aspects of the problem of a causal relation, namely, the manner in which the statements formulated in a causal convention are used and understood in the medical context, i.e., in connection with natural, oral, and written communication between physicians, biomedical scientists, etc. Although such an analysis cannot reveal the philosophical essence of causality, it may contribute to a better understanding of the meaning of statements falling under the scheme *phenomenon p is the cause of phenomenon q* and *phenomenon p is the effect of phenomenon q* and the role they play in biomedical discourse.

2.1 Meaning of Biomedical Causal Statements

The majority of biomedical causal statements exhibit the following features (a similar list of various aspects of the causal relation from a general viewpoint is presented by Bunge [6]):

1. Causal statements describe relations between particular values of features of objects (in contradistinction to functional laws denoting the relation linking the features, i.e., the sets of values).
2. A relation described by a causal statement is asymmetrical, i.e., if *p*, then *q* and not otherwise, where *p* is the cause and *q* is the effect.
3. Grammatical objects (object clauses) of causal statements denote dynamic phenomena, i.e., changes of values or states remaining after the occurrence of the change; in other words, in principle, the cause-effect relation links changes of the state of affairs and not the states themselves; when a statement describes the causal relation between two states, it is tacitly understood that a given change did occur earlier than the change constituting its effect.
4. The event that is the cause occurs earlier than the event that is the effect and, inversely, the effect (resulting change) occurs later than the cause (initial change); the difference in time between the occurrence of the initial change and the resulting change may be insignificant but is always present. This time sequence is related to (or is the essence of) the asymmetry of the relation.
5. The relation described by the causal statement may be either deterministic or uncertain, i.e., probabilistic or possibilistic (see Sect. 3). To underline the probabilistic character, the causal statements are formulated, e.g., *p is a probable (possible) cause of q* or similar.
6. The connection between phenomena described by causal statements is not a single relation but embraces numerous other links connecting many phenomena. In other words, every statement considered causal may be proved by and/or derived from other accepted general statements that is a condition sine qua non of belonging to this category.
7. From the above conditions follows the most important aspect of the meaning of the causal relation in a biomedical context, namely, the belief that some changes artificially provoked in a certain phenomenon (the cause) may provoke the desired changes in another phenomenon. This conviction constitutes the foundation of the majority of actions whose goal is to produce a purposeful modification of a certain state of affairs. The famous saying *felix qui potuit rerum cognoscere causas* not only reflects the intellectual satisfaction that gives the cognition of the causes of things, i.e., of the relations linking the phenomena, but also is connected with the practical value of this type of knowledge. In medical practice, acquaintance with the cause-effect relations is the basis of "causal therapy" whose essence consists in provoking the appearance of phenomena that bring about the desired effects.

As an example, let us take the statement, *impaired myocardial perfusion is caused by a decrease in coronary blood flow* [11].

1. The phrases *impaired myocardial perfusion* and *decrease in coronary blood flow* signify the values of features *myocardial perfusion* and *coronary blood flow* displayed by the object *myocardium (heart muscle)* and the complex object *blood flowing in the coronary artery*.
2. It is assumed that if coronary blood flow is decreased, then myocardial perfusion is impaired and not otherwise around.
3. The phrase *decrease in coronary blood flow* should be understood as indicating that blood flow has decreased at a certain time, whereas the phrase *impairment of myocardial perfusion* means that this state has once begun and afterward remains.
4. The beginning of the decrease in coronary blood flow preceded the impairment of myocardial perfusion.
5. The supposition that impaired myocardial perfusion is connected with the decrease in coronary blood flow is very probable, although not certain.
6. The statement under consideration may be explained on the grounds of anatomical conditions (connection of the coronary artery with the myocardial capillaries), continuity of the flow of blood in the artery and capillaries and the physical relations from which follow the direction of change of coronary blood flow and myocardial perfusion.
7. It is known that when coronary blood flow increases, myocardial perfusion will augment; therefore, if the decrease in blood flow in the coronary artery would be artificially changed to a normal value, myocardial perfusion would probably become normal; this expectation is the basis of the therapeutic management of ischemic heart disease.

To sum up, the superficial as well as the deeper meaning of the sentence formulated according to the scheme, *p is the cause of q*, may be presented as follows:

1. *p* and *q* are values of features *P* and *Q* of an object (see below).
2. The relation between *p* and *q* is asymmetrical.
3. *p* and *q* are dynamic phenomena.
4. *p* occurs or begins earlier than *q*.
5. The occurrence of *q* on the condition that *p* occurs (did occur) is usually not certain.
6. *p* and *q* are connected by other (mostly numerous) phenomena and relations.
7. An influence exerted on *p* may be followed by a modification of *q*.

Mutatis mutandis, the meaning of the statement "*q* is the effect of *p*" may be explicated in a similar way.

2.2 Causal Inference

The major premise of a causal inference is either the statement, "*p* is the cause of *q*" (identification of the cause), or the statement, "*q* is the effect of *p*" (identification of the effect), the minor premises “*q* occurs” or “*p* occurs,” and the conclusions, neglecting the time factor, “*p* occurs” or “*q* occurs,” respectively. The formal bases of these inferences are the following schemes:

1. If *p*, then *q* (propositional paraphrase of the cause-indicating statement); *p* occurs, therefore, *q* occurs.
2. If *q*, then *p* (paraphrase of the effect-indicating statement); *q* occurs, therefore, *p* occurs.

In comparison with the whole, very rich meaning of causal statements (see above), the propositional paraphrases, as well as the conclusions they may generate, are extremely poor. To represent a “semantic minimum,” they should be enriched at least by introducing an indication of the time sequence:

1. If $p(t)$, then $q(t')$, $t' < t$ (if *p*, then earlier *q*).
2. If $q(t)$, then $p(t')$, $t < t'$ (if *q*, then later *p*).

It should be stressed that the inferences based on cause-indicating and effect-indicating statements are not interchangeable (or are such only on certain assumptions), since q may have more causes than only p, and p may have more effects then only q (see below).

Another step enriching a propositional paraphrase is to include not only time, but also an uncertainty factor, as expressed, e.g., by an estimate of the probability:

1. If $p(t)$, then $q(t')$ with probability π, $t' < t$ (if *p*, then probably earlier *q*).
2. If $q(t)$, then $p(t')$ with probability π', $t < t'$ (if *q*, then probably later *p*).

The probabilities of the causes and effects may be different due to the same reasons mentioned before.

The above aspects of the meaning of causal statements, namely, time and probability, constitute only a small part of their whole sense, namely, only conditions 4 and 5 from the above list. A statement that could explicitly take into account the whole meaning of the sentence, “*p* is the cause of *q*”, and of the sentence, “*q* occurs as the effect of *p*,” besides formulating the above conditions, should embrace many other factors.

2.3 Uncertainty of Causal Statements

In most cases, inferences based on causal statements are subjectively self-evident. Usually, however, the formulation of a detailed proof of the correctness of conclusions based on premises expressed as causal propositions could be a difficult task since it would involve a more or less rich and complex fragment of biomedical knowledge. For example, a conclusion based on the general statement *impaired myocardial perfusion is caused by a decrease in coronary blood flow* (see above) contains a tacit assumption that the impairment of myocardial perfusion occurs as the cause of decrease in coronary blood flow, and such a statement is an abbreviation for a considerable number of interrelated statements. In an “unfolded” form, such a knowledge fragment would describe various possible causes, a temporal relation, and an estimate of the degree of certainty, etc. Thus, the full meaning of the causal inference is composed of explicit and implicit parts:

1. *p is the cause of q, p occurs, therefore, q occurs (will occur) as the effect of p* (where *q* is connected with *p* by the relation of being the effect).
2. *q is the effect of p, q occurs, therefore p occurs (has occurred) as the cause of q* (where *p* is connected with *q* by the relation of being the cause).

Causal statements are usually formulated either in a pseudodeterministic wording or in a manner that reveals their temporal and/or certainty factors. A typical pseudodeterministic wording of a causal statement is exemplified by the above cited statement: *impaired myocardial perfusion is caused by a decrease in coronary blood flow*. The precise characteristic of the time factor depends on numerous conditions and may be expressed accordingly in various ways. In principle, however, it is contained (as an enthymematic component) in the causal relation itself (see above). A few remarks presented below concern the degree of certainty of causal statements; a more detailed analysis of the problems connected with the notion of probability in a biomedical context is presented in Sect. 3.

The necessary condition of the validity of a pseudodeterministic form of a causal statement is the tacit assumption of isolation of the phenomena described by it. When the variable influences of other phenomena are to be taken into consideration in an overt form, the statement should be supplemented with a modal qualifier, indicating that the effect or the cause of an event probably occurs (is probable) or may occur (is possible).

The degree of certainty of causal propositions is expressed in genuine texts by a lot of phrases. For example, various causes of ischemic heart disease are described in [11] as follows (with minor modifications, partly enumerated in a box): *Although atherosclerotic obstructive coronary artery disease is the most common cause of chronic myocardial ischemia, it can result also from coronary artery spasm, congenital coronary artery anomalies, coronary artery embolism. ...Systemic collagen vascular disease, extrinsic compression of the coronary arteries by tumors can infrequently be a cause of myocardial ischemia*. The same author describes possible effects in the form of manifestations and complications in the following manner: *Acute myocardial infarction may be the first evidence of ischemic heart disease.*

Infrequently, arrhythmias and congestive heart failure can be the predominant consequence of ischemic heart disease. However, angina pectoris is by far the most common clinical manifestation. Some authors mention a quantitative estimation, e.g., *atherosclerosis is the most frequent cause of coronary insufficiency; it appears in 90% of patients, in the remaining 10% it may be the consequence of abnormal blood flow in coronary arteries, coronary embolism or other non-atherosclerotic changes in coronary arteries* [24].

Various modal expressions appear in the above phrases. They denote the probability of the cause or effect (in a qualitative or an approximative quantitative manner) or possibility, e.g., p is the most common (frequent) cause of q, p is the cause of q in 90% of cases, p can infrequently be a cause of q, q can infrequently be the consequence of q, q is by far the most common manifestation (effect) of p, q can result from p.

The set of above statements signifying causal relations partly enriched by estimation of probability (by means of linguistic or numerical variables) may be considered a fragment of knowledge describing the connections of ischemic heart disease with other pathological phenomena. Its graphic representation is given in Fig. 3.

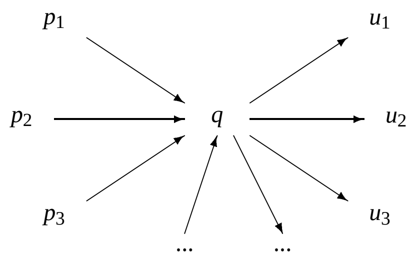

Fig. 3. Plurality of causes and effects (see text). q: ischemic heart disease; p_1: atherosclerotic obstructive coronary artery disease; p_2: coronary spasm; p_3: congenital coronary artery anomalies; u_1: angina pectoris (retrosternal pain); u_2: arrhythmias; u_3: myocardial infarction

The relations represented in Fig. 3 may be described as follows:

Causal form:

q is caused by p_1 and/or p_2 … and/or p_m.
s_1 and/or s_2 … and/or s_n are the effects of q.

Causal form enriched by probability estimation:

q is caused by p with probability π_i.
u_j is the effect of q with probability π_j.

Propositional form:

If p_1 and/or $p_2 \ldots$ and/or p_m, then q.
If q, then earlier p_1 and/or $p_2 \ldots$ and/or p_m.
If q, then u_1 and/or $u_2 \ldots$ and/or u_n.
If u_1 and/or $u_2 \ldots$ and/or u_n, then earlier q.

Propositional form enriched by probability estimation:

If p, then q with probability π_i.
If q, then u_j with probability π_j.

The degree of certainty (π) is estimated by means of an adjective (e.g., high, low), in a comparative manner (e.g., most probable), by a numerical value (e.g., 90%), etc. or by mentioning it as a possibility (without probabilistic qualification). The time factor is represented above by the adverb "earlier," and it may also be expressed more precisely.

The range of possible elementary problem solving processes based on the above (or similar) set of statements is great; the kind of particular inferences, however, depends in the first place on the type of links symbolized by the "and/or" functor that may be equivalent to a weak or strong alternative or conjunction.

In the above example, most of the elements of the sets of causes and effects are linked by weak alternatives which, however, for practical purposes are sometimes interpreted disjunctively.

3 Biomedical Aspects of the Notion of Probability

From a general point of view, the notion of probability is an even more complex and multifaceted problem in comparison with that of causality and since antiquity has been treated by philosophers, mathematicians, logicians, and thinkers from other fields. The problem of probability compared with that of causality, has greater direct impact on the development of the majority of sciences, from physics to sociology, and is more important from the viewpoint of practical applications. In biomedical sciences, various branches of mathematics related to probability are the cornerstones of research and clinical activity, and their role is rapidly increasing with the progress of applications of computers in medicine and the development of artificial intelligence (for a review, see [4]).

3.1 Logical Aspects of Probability in a Biomedical Context

A considerable part of biomedical knowledge is based on the results of statistical research and quantitative measurements. They constitute the very foundation of modern biological and medical sciences. For practical medical problem solving, however, probability based on estimation of relative frequencies, even if available, is not expressed mostly in numbers, but by means of linguistic variables expressed as modal qualifiers. Such is the usual manner in which the numerical estimates are introduced into the processes of natural thinking. It is in this way that the statements such as *if coronary arterial resistance is increased, then in 90% of patients myocardial perfusion is impaired*, are transformed into their qualitative counterparts, *if coronary arterial resistance is increased, then very probably myocardial perfusion is impaired*.

The remarks preceding the discussion of the notion of causality apply also to this analysis of the concept of probability in biomedical sciences. The following analysis is centered on those aspects of probability that are typical for biological and medical natural knowledge and reasoning (perhaps, however, some of these views might have also a more general bearing). In other words, this study concerns the problems of uncertainty as expressed in the professional medical language of everyday communication. It is based mainly on the analysis of biomedical texts that are easily accessible and are a reliable source of information, enriched by the results of behavioral observation. It is evident that this study, whose aim and methods are restricted by such research material, could contribute to a psychological or formal theory of medical problem solving only in an indirect way.

The majority of biological and biomedical laws describe events that are time-related and whose appearance is not certain or not quite determined, although in many scientific statements, time and uncertainty factors are not indicated in an explicit way. Another characteristic of typical descriptions of biological phenomena and biomedical problems is the fact that they are composed, to a considerable extent, of statements that link, in a general way, concrete values of features instead of describing the relations between sets of values, as, e.g., in physical sciences. Among various reasons for such an approach is the fact that in biomedical sciences an important role is played by qualitative rather than numerical scales. Hence, concrete statements (concerning various objects) formulated as conditional sentences, such as *if the value p of the feature P of the object O occurs, then the value q of the feature Q occurs* appear more frequently than the general laws stating that *the values of the feature Q of the object O depend in such and such a manner on the value of the feature P of this object*. The concrete character of biomedical knowledge is one of the principal causes for the necessity of using the concept of probability (frequently linked with the temporal aspects of phenomena) as a basic component of biomedical laws as well as the statements describing particular facts.

Similarly, as in other sciences and in everyday life, in medicine the notion of probability denotes, apart from a strictly formal way of understanding, the degree of anticipation of the appearance (presence) of an event (phenomenon) founded on past experience (frequency) and/or rational justification. The sentence, *if p occurs, then q occurs (or will probably occur),* means (in an extreme simplification) that on the grounds of the occurrence of p, we expect the occurrence of q rather than *not*(q). In the context of this work, the expressions denoted by p and q mean either purely qualitative states (nominal values) or intervals of magnitudes distinguished on a numerical scale, as expressed in the form of linguistic variables. They represent either a qualitative description of basically quantitative values or an attempt to make basically qualitative ones more exact. In both cases, in every probabilistic statement (in the above meaning,) appears an exact "core" surrounded by a "shadow" of uncertainty (see below).

Sometimes, we have at our disposal relevant statistical evidence on whose grounds it is possible to assess probability quantitatively. For, however, a biomedical context, such a possibility is far from common not only due to the lack of appropriate data, but also because the relevant conditions of applicability frequently are not fulfilled. Also, in most cases, the quantitative probability does not represent the essence of biomedical problems and only partly meets the needs of reasoning processes.

The second qualitative aspect of the notion of probability is related to the logical grounds on which it is estimated, i.e., on the evidence furnished by a certain set of general, already accepted statements or a fragment of general biomedical knowledge as a whole. In medicine, as well as in biology and perhaps in other sciences, such as psychology and sociology, the notion of the probability of a statement (hypothesis) is related to the totality of evidence that speaks in favor or against it. Thus, the notion of probability connected with biomedical sciences seems to be closer to the logical than to purely numerical interpretation. According to this approach, *probability measures the degree of confidence that would be rationally justified by the available evidence* (Salmon [38]). The definition given by Ajdukiewicz [1] runs, *the logical probability of a statement A relative to a statement B is the highest degree of certainty of acceptance of the statement A to which we are entitled by a fully certain and valid acceptance of the statement B* and stresses that *the condition of conclusiveness of an inference is relative to the premises and the body of knowledge*. It should be underlined that in this interpretation, probability is an attribute of the statements describing the events and not of the event described by these statements. Although Carnap distinguished various kinds of probability [7], according to Ajdukiewicz [1] and Reichenbach [35], it is possible that logical and mathematical probability have a common basis; this problem, however, exceeds by far the question of probability as used in biomedicine. The psychological aspects of probability are outside the scope of this chapter; in this context, however, note that the qualitative assessment of probability was studied in connection with probabilistic networks [42].

3.2 Justification of Uncertain Biomedical Statements

The probability of a statement describing biological or pathological phenomena, understood as the degree of its general justification, depends on various factors. The most important are (1) the quantitative characteristic of the description of the phenomena, (2) the semantic aspects of statements, (3) the interconnections of various phenomena with other ones, and (4) the place of a given statement in a certain fragment of the system of science.

1. The quantitative characteristics of phenomena are related to the distributions of continuous or discrete values that represent the direct or indirect results of measurements and observations performed in the frame of various experiments. Due to the variability inseparably connected with biological and pathological phenomena, the classes of values and objects are unsharp and overlap more or less considerably. Therefore, it is commonly impossible to ascribe a particular case (value or object) to a given class with absolute certainty, but only with a certain probability, which may be expressed as a numerical magnitude or as a linguistic variable. The latter approach is useful because natural reasoning (in which it is normally involved) is not performed on numbers. Moreover, the statistical data describe external data rather than the "inside" of phenomena and represent only one of many facets of the reality, frequently not the most significant. The numerical probability, when available and reliable, is an important component of the estimation of the degree of certainty of judgments only when it is integrated into the general bulk of knowledge, i.e., regarded as one of the pieces of evidence.
2. Typical general statements that are considered in this work are conditional sentences composed of objects (in a grammatical sense) and predicates containing relational and adjectival parts. The range of denotation of the majority of these expressions is frequently broad, i.e., they describe classes composed of more or less numerous kinds of objects, relations, and features. Moreover, their meaning is usually ambiguous and unsharp, and the sentences are mostly formulated so that an important part of their meaning is not expressed in an explicit way. It is evident that the statements exhibiting such a semantic characteristic cannot contain precise expressions denoting the uncertainty factor.
3. Biomedical phenomena appear as elements of other phenomena and, accordingly, the statements that describe them constitute the components of complex systems of laws. Their probability, therefore, depends to a considerable extent on their connections with various parts of a greater whole. Although a "core probability" of a statement, i.e., its probability considered in abstraction from the broader environment, is relatively stable, it is surrounded by a shadow of uncertainty due to the fact that in different situations, the phenomena are subject to various influences. Obviously, the variability of phenomena, e.g., of the values of features of the organism or of its different parts, is more conspicuous in the state of disease than in normal conditions. Therefore, reasoning, whose aim is estimation of the probability of pathological phenomena, i.e., their identification, including the degree of certainty, is especially typical for the process of diagnosis.
4. From the statistical point of view, the justification of a statement consists in performing mathematical operations on numerical magnitudes. On the other hand, in a biomedical context, the majority of premises are expressed not by numerical values but by linguistic variables, and the way of arriving at the conclusions

is not a calculus but operations on the judgments of another kind. On the other hand, a strictly formal approach to qualitative (linguistic) expressions, which are the basic units of biomedical knowledge and reasoning, may be used only with considerable limitations. The logical structure of biomedical statements, general as well as particular, constitutes a "deterministic skeleton" of their meaning that is surrounded by an informal "flesh." The answer to the question whether, or with what certainty, a given statement should be accepted, may be arrived at by proving or justifying it by using various methods of inference making, i.e., making a logical analysis on various levels of knowledge. By "logical approach," I mean here an analysis as detailed, ordered, and precise as possible and useful, rather than the application of a formal system.

3.3 Typical Expressions Denoting Probability and Their Interpretation

In natural, professional medical language, various levels of probability are designated by a great variety of expressions, e.g., "in most (some) cases," "sometimes," "frequently (seldom, exceptionally)," "probably," "very likely," "something may (can) occur," "something is a probable cause (effect)," "something is supposed to be," etc. The meaning of these and similar expressions is ambiguous and unsharp to the same extent as those that are used in everyday language. However, experience with the verbal usage of doctors and the study of professional texts and other sources suggest that the following main types of medical subjective probability may be distinguished: formal certainty, quasi-certainty, high probability, intermediate probability, low probability, and possibility.

The above types may be illustrated by the following sentences; it should be remembered, however, that their sense overlaps to a considerable extent.

Formal certainty: blood flow is either normal or decreased (still, these ranges are unsharp).

Quasi-certainty: if the artery is constricted (and blood pressure is not increased and so on), then almost certainly blood flow in this artery is decreased.

High probability: if the patient complains of retrosternal pain, then very probably this patient suffers from ischemic heart disease.

Intermediate probability: if the patient suffers from atherosclerosis, this patient probably displays an elevated level of a certain fraction of cholesterol.

Low probability: the probability that the patient suffering from mild myocardial ischemia will develop severe heart failure is small.

Possibility: the patient suffering from myocardial ischemia may display a normal electrocardiographic recording.

Although the majority of doctors will probably agree with the above examples, nobody would be inclined to consider them exact. Sometimes, however, similar statements, or rather their meanings, appear as important elements in the decision making process.

The above classification embraces qualitative degrees of probability, i.e., is related to a discrete scale. Although it may be considered a “strong” type of probability estimation (probabilistic verification), not less important is a “weaker” type of the assessment of probability, namely, (1) comparison of the probability of various statements and (2) comparison of the probability of the same statement in the light of different pieces of evidence, i.e., its confirmation.

1. The former type, i.e., *verification of hypotheses*, consists of indicating which of the statements describing the two values of the same feature is more probable. For example, the judgment stating that the patient suffering from atherosclerosis probably displays an elevated blood level of a certain fraction of cholesterol is more probable than the statement suggesting that this patient will display a normal or diminished level of this compound.
2. The notion of *confirmation* of statements is based on the assumption that if new (supplementary), previously unknown or neglected, evidence is taken into account, then the probability of a given statement increases or decreases in comparison with that estimated before this information has been obtained or considered. For example, the probability of the statement, *if the patient suffers from atherosclerosis and retrosternal pain, then this patient suffers from myocardial ischemia,* is higher in comparison with the statement, *if the patient suffers from atherosclerosis, then this patient suffers from myocardial ischemia*. The problems related to the confirmation of hypotheses play especially important roles in the approach presented in this work and are analyzed in more detail in the next paragraph.

The above considerations may be summarized as follows:

The notion of probability has several meanings that depend on the context in which it is used by the physicians and/or appears in biomedical texts. On the grounds of general usage, several qualitative degrees of uncertainty (probability) may be distinguished. In practice, they are well understood and discriminated by the doctors such a categorization, however, may serve mainly to introduce a certain order (of a rather soft kind) rather than using them as a basis for a precise scheme of inference. The interpretation of the way of understanding the notion of the probability of a statement (of a phenomenon) may be either strong or weak. The strong meaning corresponds to the qualitative interval scale or a nominal scale and is expressed by adjectival expressions such as high, medium, low, almost equal to certainty (positive or negative), etc. The weak interpretation is based on a comparative scale; this category embraces the confirmation of hypotheses.

The probability of a statement depends on its justification on the grounds of certain knowledge (general and particular), i.e., on the kind, strength, and amount of evidence. In some cases the evidence exists that may constitute sufficient basis for a strong kind of probability assessment (including practical certainty), in other situations, only a weak assessment is justified on the grounds of existing evidence. In the latter case, a certain number of "weak" pieces of evidence may be sufficient for a strong assessment. This is important point since probability is one of the factors on which the process of decision making is based, and for a decision concerning an action, frequently, a strong type of its assessment, instead of a weak one, is necessary. Commonly, in the absence of a single strong piece of evidence, the decision making subject is compelled to use a set of weak pieces of evidence to arrive at a sufficient estimate of probability.

The appearance of a phenomenon or the acceptance of a statement may have a different influence on the probability of a given phenomenon (statement); in other words, the revision of probability may be greater or smaller. In clinical practice we say that the symptoms and signs are deciding, pathognomic, strong, weak etc. and a similar "strength" of every evidence may be distinguished. Thus, when a series of observations is performed, the probability of a given phenomenon (statement) undergoes a sequence of changes depending on the number of pieces of evidence as well as on their confirmatory strength.

3.4 Confirmation of Hypotheses: A Theoretical Model

The assignment of a certain value of probability to a hypothesis may be called *probabilistic verification*, whereas the estimation of the direction and approximative magnitude of the change (revision) of probability of a hypothesis is called confirmation [8]. As the latter type of probability evaluation seems to be especially important for a biomedical problem solving theory, especially in a model approach, it deserves a more detailed presentation.

In research on the fundamental problems of the methodology of sciences, the authors [1,8,38] frequently use a notation that shows the connection of the probability of statements with general knowledge. It is assumed that an analogy exists between the estimate of numerical and logical probability that justifies the similarity of these approaches. According to this usage, the probability of the hypothesis h in the light of general knowledge k may be designated by symbol $p(h/k)$, and its probability in the light of general knowledge k and evidence e by the symbol $p(h/k \text{ and } e)$. Usually, the symbol p may be interpreted as denoting either quantitative or qualitative probability. In the present analysis, I use the symbol π for probability based on various kinds of justification (logical as well as quantitative), and a piece of evidence here is designated by the letter s (from sign), h and k denoting the hypothesis and general knowledge, respectively.

According to my previous work [13] and in a way similar to the usage of this term by above mentioned authors, the *confirmation* or *degree of confirmation* $c(h,k,s)$ of the hypothesis h is equal to its probability, as estimated in the light of general knowledge k and observation (acceptance) of novel evidence (e.g., sign s) divided by the probability of this hypothesis h estimated without considering this evidence. Hence,

$$c(h,k,s) = \frac{\pi(h/k \text{ and } s)}{\pi(h/k)}.$$

For two alternative and mutually exclusive hypotheses h_1 and h_2 [i.e., h_2 is $not(h_1)$], it may be shown (on the grounds of the Bayes formula) that if we accept the evidence s, whose probability is higher on the grounds of hypothesis h_1 than on the grounds of hypothesis h_2 (the evidence s speaks in favor of hypothesis h_1 and against hypothesis h_2 or, using statistical terminology, the likelihood of h_1 in the light of s is greater than the likelihood of h_2 in the light of s, or), then the probability of hypothesis h_1 increases, and that of the hypothesis h_2 decreases.

The following conditions are equivalent:

1. $\pi(s/k \text{ and } h_1) > \pi(s/k \text{ and } h_2)$.
2. $c(h_1,k,s) > c(h_2,k,s)$.
3. $\pi(h_1/k \text{ and } s) > \pi(h_1/k)$.
4. $c(h_1,k,s) > 1$.
5. $\pi(h_2/k \text{ and } s) < \pi(h_2/k)$.
6. $c(h_2,k,s) < 1$.

Hence, in particular, the following implications hold:

• If $\pi(s/k \text{ and } h_1) > \pi(s/k \text{ and } h_2)$ then $c(h_1,k,s) > c(h_2,k,s)$.

• If $\pi(s/k \text{ and } h_1) > \pi(s/k \text{ and } h_2)$ then $\pi(h_1/k \text{ and } s) > \pi(h_1/k)$.

• If $\pi(s/k \text{ and } h_1) > \pi(s/k \text{ and } h_2)$ then $c(h_1,k,s) > 1$. In other words, in the situation described by any of conditions (1)–(6), confirmation of hypothesis h_1 is greater than one, and the confirmation of the hypothesis h_2 is less than one (evidence s strengthens hypothesis h_1 and weakens hypothesis h_2). The above conclusion is valid irrespective of the initial probability of hypotheses h_1 and h_2. In other words, to estimate the direction of the change (revision) in the probability of a hypothesis, we do not need to know its initial probability. The magnitude, however, of the change in probabilities in the light of evidence s depends on the initial probabilities of the hypotheses (for hypothesis h_i it is the greater, the smaller the initial probability of this hypothesis), as well as on the magnitude of likelihood values.

Analogously, one can obtain:

$$\text{if } \pi(s/k \text{ and } h_1) < \pi(s/k \text{ and } h_2), \text{ then } c(h_2,k,s) < 1.$$

Every empirical examination, scientific as well as practical, consists of gathering successively the pieces of evidence (e.g., signs s_1, s_2, and so on) that are relevant, novel, and independent (or at least are assumed to be independent). The greater the number of observed pieces of evidence (signs) that speak in favor of (suggest) a given hypothesis, the greater the degree of its confirmation. Theoretical aspects of this problem are discussed by Ajdukiewicz [1] and Caws [8].

Assuming, s_1, s_2 are conditionally independent (with respect to conditions described by k and h) and $\pi(h/k \text{ and } s_1) > \pi(h/k)$ as well as $\pi(h/k \text{ and } s_2) > \pi(h/k)$, one can show that the probability of hypothesis h in the light of general knowledge k and the pieces of evidence s_1 and s_2 satisfies the following inequality:

$$\pi[h/(k \text{ and } s_1 \text{ and } s_2)] > max[\pi(h/k \text{ and } s_1), \pi(h/k \text{ and } s_2)].$$

The above concept of probability revision in the light of growing evidence constitutes the basis of the semantic model of hypothesis verification presented in this work.

3.5 An Example of the Semantic Model (a General Presentation)

The subject (e.g., a doctor) verifies (in a broad sense) the hypothesis h_1 versus an alternative hypothesis h_2 [or $not(h)$] in the light of a set of accessible pieces of evidence or signs s_1, s_2, and so on. On the grounds of relevant general knowledge, the hypotheses are a priori neither excluded nor certain. Some of the signs confirm hypothesis h_1 (and disconfirm h_2), others confirm hypothesis h_2 (and disconfirm h_1). The signs have various confirmational strength, i.e., the degree of confirmation of each hypothesis in the light of a given sign may be weak or strong. In other words, every sign confirms hypothesis h_1 or h_2 weakly or strongly. It is assumed that the weak and strong confirmation mean that the change in probability is equal to one or two confirmational steps, respectively. The subject performs the appropriate examination, i.e., successively observes (or accepts on theoretical grounds) the occurrence of signs and intends to assess the direction of revision of the probability of hypotheses to decide which hypothesis is more probable than its counterpart (to make a weak verification), or to assign to the hypotheses certain values of probability (to verify them in a strong sense). If all observed signs or their majority confirm the same hypothesis and the number of signs as well as their strength are considerable, then the probability of this hypothesis unequivocally increases, and one of them may be verified or even considered practically certain in the light of a given set of signs. If, however, the observed signs don't confirm (weakly or strongly) the same hypothesis, the examination permits only (in the best case) choosing the hypothesis whose probability is higher in the light of the set of signs (weak verification).

Let us consider two possible situations: at first, two signs s_1 and s_2 having equal strength, secondly, signs of various strengths.

1. *Signs of equal strength:*
   (a) Both signs confirm hypothesis h_1 and disconfirm h_2, i.e.,
   $\pi(s_1/k \text{ and } h_1) > \pi(s_1/k \text{ and } h_2)$,
   $\pi(s_2/k \text{ and } h_1) > \pi(s_2/k \text{ and } h_2)$,
   where k is the relevant general knowledge.
   Therefore,
   $\pi(h_1/k \text{ and } s_1) > \pi(h_1/k)$,
   $\pi(h_1/k \text{ and } s_2) > \pi(h_1/k)$,
   $\pi(h_1/k \text{ and } s_1 \text{ and } s_2) > max[\pi(h_1/k \text{ and } s1), \pi(h_1/k \text{ and } s1)]$.
   (b) s_1 confirms h_1, s_2 confirms h_2, i.e.,
   $\pi(s_1/k \text{ and } h_1) > \pi(s_1/k \text{ and } h_2), \pi(s_2/k \text{ and } h_1) < \pi(s_2/k \text{ and } h_2)$.
   Therefore,
   $\pi(h_1/k \text{ and } s_1) > \pi(h_1/k)$,
   $\pi(h_1/k \text{ and } s_2) < \pi(h_1/k)$,
   in view of various directions of confirmation of hypotheses in the light of $s_1 and s_2$, the overall confirmation cannot be determined.
2. *Signs of various strengths:*
   We assume that the confirmation of h_1 in the light of s_1 is weak and the confirmation of h_1 in the light of s_2 is strong:
   $\pi(s_1/k \text{ and } h_1) > \pi(s_1/k \text{ and } h_2), \pi(s_2/k \text{ and } h_1) >> \pi(s_2/k \text{ and } h_2)$.
   Therefore,
   $\pi(h_1/k \text{ and } s_1) > \pi(h_1/k)$,
   $\pi(h_1/k \text{ and } s_2) >> \pi(h_2/k)$.
   $\pi(h_1/k \text{ and } s_1 \text{ and } s_2) >> \pi(h1/k \text{ and } s_1 \text{ or } s_2)$.

Similarly, the observation of a greater number of signs possibly exhibiting various degrees of strength may be interpreted from the point of view of the confirmation and verification of hypotheses. For some purposes (see Sect. 5) an alternative notation of the direction and strength of confirmation may be useful, namely, $\uparrow$ or $\uparrow\uparrow$, etc., for a weak and strong positive confirmation and $\downarrow$ or $\downarrow\downarrow$, etc. for a weak and strong negative confirmation, respectively.

The problem of the qualitative influence of signs in the frame of probabilistic networks was discussed by Wellman [42] and Druzdzel [18, 19].

The application of the above described model to the analysis of an authentic clinical case is presented in Sect. 5.

4 Analysis of an Original Text (Ischemic Heart Disease) Using the Semantic Model

The application of the above described approach to the analysis of a clinical text is presented here using as an example a fragment of the chapter entitled “Ischemic heart disease” (K. Chatterjee) from the textbook edited by J.H. Stein [11]. The statements form partly a continuous text, are partly cited from different places, and are only slightly abbreviated. Here and there the wording is insignificantly simplified (numbering is inserted for the use in the discussion).

1. *Obstructive coronary artery disease caused by atherosclerosis is the most common cause of chronic ischemic heart disease.*
2. *Ischemic heart disease may be totally silent or may manifest in angina, arrhythmias, and heart failure.*
3. *Myocardial ischemia stems from the imbalance between myocardial oxygen requirements and oxygen supply which can occur from a primary decrease in coronary blood flow or from an increase in myocardial oxygen requirements or their combination.*
4. *Impaired myocardial perfusion, caused by a decrease in coronary blood flow resulting from an increase in the coronary arterial resistance and/or abnormalities of the coronary vascular autoregulatory mechanisms, appears to be the principal cause for myocardial ischemia in other clinical syndromes.*
5. *Angina pectoris in by far the most common clinical manifestation of ischemic heart disease.*
6. *Downsloping electrocardiographic S-T segments are highly specific for coronary artery disease.*
7. *The definitive diagnosis of coronary artery disease can be made only by coronary angiography.*

4.1 Causal Paraphrases of the Original Statements

In the sentences below, the names of phenomena (complexes object, feature, value) are distinguished by italics and marked by uppercase letters for use in Fig. 4a.

1. The most probable cause of *chronic ischemic heart disease* (E) is *atherosclerotic obstructive coronary artery disease* (L).
2. *Angina* (H), *arrhythmias* (J), and *heart failure* (K) may be the effects of *ischemic heart disease* (E).
3. *Myocardial ischemia* (E) may be caused by a *decrease in coronary blood flow* (B) and/or from an *increase in myocardial oxygen requirements* (F).
4. *Myocardial ischemia* (E) is caused by *impaired myocardial perfusion* (A).
5. *Impaired myocardial perfusion* (A) is caused by a *decrease in coronary blood flow* (B).

6. *Decrease in coronary blood flow* (B) is caused by an *increase in coronary arterial resistance* (C) and/or *abnormalities of coronary vascular autoregulatory mechanisms* (D).
7. *Angina pectoris* (M) is the most probable effect of *ischemic heart disease* (E).
8. Appearance of *downsloping electrocardiographic S–T segments* (N) is a very probable effect of *coronary artery disease* (E).
9. *Coronary artery disease* (E) causes the *appearance of a specific image in coronary angiography* (T).

The fragment of knowledge composed of the above causal statements may be represented as a directed graph, as in Fig. 4a.

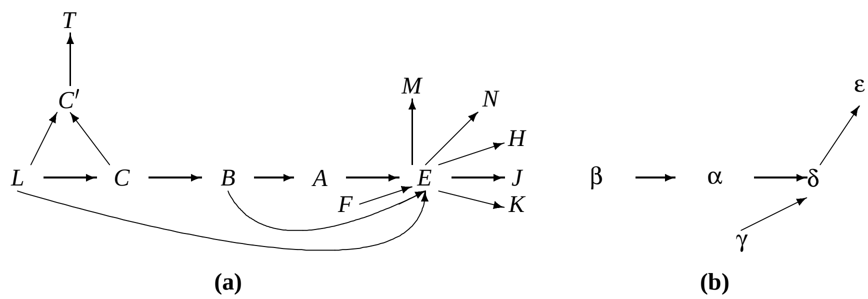

Fig. 4. (a) Graphic representation of a causally paraphrased fragment of knowledge (see text). Letters correspond to phrases designating the value of a feature of an enthymematic object; e.g., E : "myocardial ischemia," i.e., the phrase meaning a decreased value of the blood supply (feature) to the myocardium (heart muscle, object). Arrows denote causal relations described by the phrase "is the cause (effect) of" in which time and probability factors are enthymematically comprised. Long arrows represent shortcut relations mentioned in the original text. The element C' means "atheromatic constriction of the coronary arteries" absent from the original text and introduced here to ensure appropriate links among other elements.
(b) Arrows denote functional relations (phrase "depends on", see text).

4.2 Paraphrases of Functional Laws

According to the natural understanding of the sentences contained in the original text, they constitute the concretization of more comprehensive (general, functional) laws describing the relation linking whole features instead of their particular values. The following statements represent the general laws that embraced some of the concrete statements analyzed above (Greek letters are used in Fig. 4b).

1. Myocardial perfusion (α) depends on coronary blood flow (β).

2. Myocardial oxygen supply (δ) depends on myocardial perfusion (α) and/or myocardial oxygen requirements (γ).
3. Retrosternal pain sensations (ε) depend on myocardial oxygen supply (ε).

4.3 Propositional Paraphrases (Conditional Sentences)

To present the propositional form, three statements are represented below in the form of conditional sentences; in this case, the time and probability factors are added or preserved (in the original texts, they are commonly omitted).

1. If in time t coronary blood flow is decreased (B), then in time $t' \geqslant t$, myocardial perfusion is probably impaired (A).
2. If in time t, myocardial perfusion is impaired (A) and/or in time t, myocardial oxygen requirements are increased (F), then in time $t' \geqslant t$, myocardial ischemia (E) probably occurs.
3. If in time t, myocardial blood supply is decreased (i.e., myocardial ischemia occurs, E), then in time $t = t'$, retrosternal pain occurs (H).

The regressive form based on the assumption of isolation of the above statements is the following (the time factor is rendered by typical linguistic expressions):

1. If myocardial perfusion is impaired (A), coronary blood flow has probably been decreased (B).
2. If myocardial ischemia (E) occurs, then myocardial perfusion has probably been impaired (A), and/or myocardial oxygen requirements have probably been increased (F).
3. If retrosternal pain occurs (H), then myocardial blood supplies probably are decreased (i.e., myocardial ischemia (E) probably occurs).

The following sentences are complementary for the above statements (1)–(3):

1. If coronary blood flow is normal (B′), then myocardial perfusion is probably normal (A′).
2. If myocardial perfusion is normal (A′) and/or myocardial oxygen requirements are increased (F′), then myocardial blood supply is normal (i.e., myocardial ischemia does not occur, E′).
3. If myocardial blood supply is normal (i.e., myocardial ischemia does not occur, E′), retrosternal pain does not occur (H′).

Propositional paraphrases may be used as a basis for a pathophysiological inference.

Progressive inference:

1. In time t, coronary blood flow is decreased (B); therefore, in time $t' \geqslant t$, myocardial perfusion is probably impaired (A).

2. In time t, myocardial perfusion is impaired (A); therefore, in time $t' \geqslant t$, myocardial ischemia (E) probably occurs.
3. In time t, myocardial oxygen requirements are increased (F); therefore, in time $t' \geqslant t$, myocardial ischemia (E) probably occurs.

Regressive inference (with the assumption of isolation):

1. Myocardial perfusion is impaired (A); therefore, coronary blood flow has probably decreased (B).
2. Myocardial ischemia (E) occurs; therefore, myocardial perfusion has probably been impaired (A) and/or myocardial oxygen requirements have probably increased (F).

The diagnostic inferences based on propositional paraphrases may be exemplified as follows.

1. *Progressive form:*
   In time t, myocardial blood supply is decreased (i.e., myocardial ischemia occurs, E); then in time $t = t'$, retrosternal pain (will) occurs (H).
2. *Regressive:*
   Retrosternal pain occurs (H); therefore, myocardial blood supply probably is decreased (i.e., myocardial ischemia probably occurs, E).

5 Analysis of an Authentic Clinical Case Using the Semantic Model

The present clinical example is based on the following text extracted from a document concerning the treatment of a patient during his stay in the clinical department (epicrisis).

"*66-year-old patient was admitted to the Coronary Care Unit with a severe retrosternal pain of three hour duration. Acute ischemic heart disease with a possibility of myocardial infarction was diagnosed. The chest pain resolved after morphine and the treatment with subcutaneous low molecular heparine, aspirine and intravenous nitrates was administrated. The diagnosis of myocardial infarction was ruled out based on a creatine kinase level estimation and subsequent electrocardiographic recordings. Urgent coronary angiogram was performed which showed total occlusion of the left anterior descending artery and two lesions of the proximal left circumflex artery (stenosis of 60 and 90%). Percutaneous transluminar coronary angioplasty was performed.*"

The above document contains the description of signs exhibited by the patient, diagnostic hypotheses (statements), and therapeutic procedures. The last component is presented only to underline the inseparable connection between medical cognition and decision making. The signs (partly completed on the grounds of the context) are the following: retrosternal pain, normal blood level of creatine kinase, ECG recordings not characteristic for myocardial infarction, and a coronarographic image characteristic of coronary stenosis. The diagnostic hypotheses embraced acute ischemic heart disease and myocardial infarct. In other words, the epicrisis contains (apart from therapeutic information) the premises and conclusions of the doctor's hypothetical reasoning.

According to the model presented in this work, in the process of solving a biomedical problem, e.g., a diagnostic one, the final steps are based on relatively simple propositions (working statements) derived from statements belonging to general knowledge that have various forms and exhibit different levels of generality and degree of precision. In a model approach, the connection of the general (theoretical) statements with the working propositions may be represented as a sequence of paraphrases of original, e.g., textbook, sentences in which they are transformed into typical causal, propositional, and working statements (see Sect. 1). For example, an original textbook formulation: "angina pectoris (retrosternal pain) is by far the most common manifestation of ischemic heart disease" [11], may be expressed as a causal proposition: "the most probable cause of retrosternal pain is ischemic heart disease," which in its turn is approximately equivalent to the propositional paraphrase "if myocardial ischemia occurs, then probably retrosternal pain occurs." Finally as a working proposition, "if retrosternal pain occurs, then probably myocardial ischemia occurs."

A general working proposition in conjunction with the particular statement "retrosternal pain occurs," is the basis for the conclusion "therefore, probably myocardial ischemia occurs."

Similarly other working statements may be derived and applied in the inference-making process.

If myocardial ischemia occurs, then, myocardial infarction is possible (the same as a propositional paraphrase).

If blood CK is normal, then, very probably myocardial infarction does not occur.

If the ECG curve does not exhibit typical features, then, myocardial infarction very probable does not appear.

If myocardial ischemia occurs, then, very probably the coronary artery is constricted.

If a coronary angiogram exhibit typical features, then, the coronary artery is almost certainly constricted.

On the grounds of the above working statements and the statements describing the signs presented by the patient, the following conclusions are drawn:

1. Severe retrosternal pain occurs; therefore, very probably the patient suffers from ischemic heart disease, possibly with myocardial infarction.
2. Assuming that ischemic heart disease occurs, myocardial infarction is possible.
3. Creatine kinase (CK) level is normal; therefore, probably myocardial infarction does not occur.
4. Electrocardiographic (ECG) recordings are normal; therefore, probably myocardial infarction does not occur.
5. Assuming that ischemic heart disease is present, obstruction of the coronary artery is very probable.
6. Occlusion and stenosis of coronary arteries appear in the angiogram; therefore, obstruction of the coronary artery is practically certain.
7. Considering that the obstruction of the coronary artery is almost certain, therefore, ischemic heart disease is very probable.

The final solution of the problem may be formulated as follows: the hypothesis of ischemic heart disease is accepted (ischemic heart disease is certain), and the hypothesis of myocardial infarction is rejected (myocardial infarction is improbable).

It should be stressed that each of both diagnostic hypotheses, i.e., that of ischemic heart disease, as well as that of the absence of myocardial infarction, is supported by two signs: "retrosternal pain" and "occlusion of coronary arteries in the angiogram" as concerns the former, and "normal CK level" and "normal ECG recordings," as concerns the latter. Therefore, it seems justified to estimate the resultant probability of these statements as "almost certain," despite the fact that in the light of the signs taken separately, the probability of these hypotheses would rather be "very high."

The clinical document, the extract of which is being analyzed here, presents the results of the patient's examination without mentioning the obvious fact that the appropriate tests were previously designed. It seems worthwhile, however, to complete the present study by indicating the logical grounds for the choice of tests.

The reasoning connected with the planning of the diagnostic examination is approximately the following. For example, the doctor suspects myocardial infarction and wishes to verify this hypothesis. He or she knows that if the patient suffers from this condition, the CK blood level would very probably be elevated. If myocardial infarction does not appear, the CK blood level would very probably be normal. In the former case, he or she would accept (with appropriate probability) the hypothesis of myocardial infarction, in the latter case he or she would reject it. Thus, the doctor takes into consideration two statements: "if myocardial infarction occurs, then the blood level of CK is very probably elevated" and "if myocardial does not occur, then probably the blood CK level is normal," which play the role of working hypotheses. The conclusion is based on hypothetical reasoning: assuming that myocardial infarction is present, it should be supposed that the blood CK level is elevated, on the contrary, assuming that myocardial infarction is absent, it should be supposed that the blood CK level is normal. It is easy to see that in the process of planning examinations, the working statements are equivalent to the propositional paraphrase; in other words, as concerns the planning reasoning, the statements' transformation may be simpler in comparison with the interpretation of already observed signs. On the other hand, to decide which test is necessary for proving a given hypothesis, at least two statements, instead of only one, should be considered.

Below is a list of general statements (and their interrelations) on which the process of solving the problem under consideration could be based, together with some examples of inferences. General statements, as well as particular premises and conclusions, are artificially reconstructed on the grounds of the semantic model described in this work. The list embraces functional laws and statements formulated in causal and propositional form; the model inference is presented according to the confirmational approach (see Sects. 3.4 and 3.5).

Functional Laws (see Fig. 5)

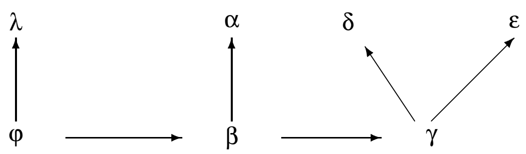

Fig. 5. Scheme of functional laws (see text)

1. Retrosternal pain sensations (α) depend on myocardial oxygen supply (β).
2. The morphological state of the heart muscle cells (γ) depends on the myocardial oxygen supply (β).
3. The level of blood CK (δ) depends on the morphological state of the heart muscle cells (γ).
4. Some features of the ECG curve (ε) depend on the state of the heart muscle cells (γ).
5. The myocardial oxygen supply (β) depends on the morphological features of the coronary arteries (φ).
6. The features of the coronary angiogram (λ) depend on the morphological features of the coronary arteries (φ).

Causal Statements

1. Myocardial ischemia (E) is the (a probable) cause of retrosternal pain (M).
2. Myocardial infarction (P) is the effect (a probable, possible effect) of myocardial ischemia (E).
3. Myocardial infarction (P) is the cause of the elevated blood CK level (U).
4. Myocardial infarction (P) is the cause of the appearance on the ECG curve of features typical of this condition (V).
5. Constriction of coronary arteries (W) is the (a probable) cause of myocardial ischemia (E).
6. Constriction of coronary arteries (W) is the cause of characteristic features in the coronary angiogram (Z).

The fragment of knowledge (general functional laws) on which the above problem and its solution are based may be represented in graphic form (Fig. 5).

A list of propositional probabilistic paraphrases follows:

1. If myocardial ischemia occurs (E), then, probably retrosternal pain occurs (M).
2. If retrosternal pain (M) occurs, then, probably myocardial ischemia occurs (E).
3. If myocardial ischemia occurs (E), then, myocardial infarction is possible (P).
4. If myocardial infarction occurs (P), then, myocardial ischemia certainly occurs (E).
5. If myocardial infarction occurs (P), then, probably the blood CK level is elevated (U).
6. If the blood CK level is elevated (U), then, probably myocardial infarction occurs (P).
7. If myocardial infarction does not occur (P′), then, probably the blood CK level is normal (U).
8. If blood CK level is normal (U′), then, probably myocardial infarction does not occur (P′)
9. If myocardial infarction occurs (P), then, almost certainly the ECG curve exhibits features (V) typical of myocardial infarction.
10. If the ECG curve exhibits typical features (V), then, myocardial infarction very probably occurs (P′).
11. If myocardial infarction does not occur (P′), then almost certainly the ECG curve typical (of myocardial infarction) does not appear (V′).
12. If the ECG curve does not exhibit typical features (V′), then, myocardial infarction very probably does not occur (P′).
13. If constriction of coronary arteries occurs (W), then, probably myocardial ischemia occurs (E).
14. If myocardial ischemia occurs (E), then, constriction of coronary arteries probably occurs (W).
15. If constriction of coronary arteries occurs (W), then, almost certainly characteristic features appear in the coronary angiogram (Z).

16. If characteristic features appear in the coronary angiogram (Z), then, almost certainly constriction of coronary arteries occurs (W).

Diagnostic conclusions (interpretation of signs) based on the above list of propositional probabilistic paraphrases:

17. Retrosternal pain (M) occurs; therefore, probably myocardial ischemia occurs (E).
18. Myocardial ischemia occurs (E); therefore, myocardial infarction is possible (P).
19. The blood CK level is normal (U′); therefore, probably myocardial infarction does not occur (P′).
20. The ECG curve does not exhibit typical features (V′); therefore, myocardial infarction very probably does not occur (P′).
21. Characteristic features appear in the coronary angiogram (Z); therefore, almost certainly constriction of coronary arteries occurs (W).

Pathophysiological inference based on the above list of propositional probabilistic paraphrases 13, 14:

22. Regressive conclusion: assuming that myocardial ischemia occurs (E); therefore, probably constriction of coronary arteries occurs (W).
23. Progressive conclusion: assuming that constriction of coronary arteries occurs (W); therefore, myocardial ischemia occurs (E).

Confirmational model based on the above list of propositional probabilistic paraphrases.

24. Retrosternal pain (M) strongly suggests myocardial ischemia (E)
    If M, then, $\uparrow\uparrow E$ or $\downarrow\downarrow E'$.
25. Myocardial ischemia (E) suggests the possibility of myocardial infarction (P).
    If E, then, $\uparrow$P or $\downarrow$P′.
26. The normal blood level of CK (U′) strongly suggests the absence of myocardial infarction (P′).
    If U′, then $\uparrow\uparrow$P′ or $\downarrow\downarrow$P.
27. The absence of features typical of the myocardial infarction in the ECG curve (V′) strongly suggests absence of myocardial infarction (P′).
    If V′, then, $\uparrow$P′ or $\downarrow$P.
28. Constriction of coronary arteries (W) strongly suggests myocardial ischemia (E).
    If W, then, $\uparrow\uparrow$E or $\downarrow\downarrow$E′.
29. Myocardial ischemia (E) strongly suggests constriction of coronary arteries (W).
    If E, then, $\uparrow\uparrow$W or $\downarrow\downarrow$W′.

30. The appearance of characteristic features in the coronary angiogram (Z) proves the constriction of coronary arteries (W).
If Z, then W and W′.

Conclusions

31. (E and U′ and V′); therefore, P.
32. If Z, then, W; if W, then ↑↑E or ↓↓E′; Z; therefore, ↓↓↓E or ↓↓↓E′.

Acknowledgment

I deeply appreciate the kind help and criticism of Professor Piotr Kułakowski of the Department of Cardiology of the Medical Center of Postgraduate Education in Warsaw. Needless to say, deficiencies in this work are my sole responsibility. The research has been partially supported by the State Committee for Scientific Research of the Republic of Poland (KBN) research grant 8 T11C 025 19.

References

1. K. Ajdukiewicz. *Pragmatic Logic*. PWN, Warsaw, 1974.
2. R. Baud. Present and future trends with nlp (natural language processing). *International Journal of Medical Informatics*, 52: 133–139, 1998.
3. M.S. Blois. *Information and Medicine. The Nature of Medical Descriptions*. University of California Press, Berkeley, CA, 1984.
4. G.W. Bradley. *Disease, Diagnosis and Decisions*. Wiley, Chichester, 1993.
5. M. Bunge. *Causality – the Place of the Causal Principle in Modern Science*. Harward University Press, Cambridge, MA, 1959.
6. M. Bunge. Conjonction, succession, determination, causalité. In *Les Theories de Causalité*, 112–132, University of Paris Press, Paris, 1971.
7. R. Carnap. *Logical Foundations of Probability*, (3rd ed.). University of Chicago Press, Chicago, 1950.
8. P. Caws. *The Philosophy of Science. A Systematic Account*. Van Nostrand, Princeton, NJ, 1965.
9. W. Ceuster, P. Spyns, G. de Moor. From syntactic–semantic tagging to knowledge discovery in medical texts. *International Journal of. Medical Informatics*, 52: 149–157, 1998.
10. T. Chałubiński. *The Method of Arriving at Medical Indications*. Gebethner and Wolff, Warsaw, 1874 (in Polish).
11. K. Chatterjee. Ischemic heart disease. In J.H. Stein, editor, *Internal Medicine,* (4th ed.), 149–169, Mosby, St. Louis, 1994.
12. P. Cutler. *Problem Solving in Clinical Medicine*. Williams and Wilkins, Baltimore, MD, 1979.
13. J. Doroszewski. Evaluation of probability in the process of verification of diagnostic hypotheses. *Studia Filozoficzne*, 10: 121–142, 1972 (in Polish).

14. J. Doroszewski. Polish medical language. In W. Pisarek, editor, *Polszczyzna 2000*, Jagiellonian University Press, Kraków, 1999.
15. J. Doroszewski. An analysis of medical knowledge and reasoning. In A. Tymieniecka, Z. Zalewski, editors, *Analecta Husserliana*, 57–66, Kluwer, Dordrecht, 2000.
16. J. Doroszewski. Solving pathophysiological problems. *Med. Sci. Mon.*, 6: 1–7, 2000.
17. J. Doroszewski. A model of medical knowledge based on systems' and semiotic approach. *Studia Semiotyczne*, 23, 2001 (in Polish).
18. M.J. Druzdzel, M. Henrion. Intercausal reasoning with uninstantiated ancestor nodes. In *Proceedings of the 9th Annual Conference on Uncertainty in AI*, 317–325, Washington, DC, 1993.
19. M.J. Druzdzel. Qualitative verbal explanations in Bayesian belief networks. *AISB Q.*, 94: 43–54, 1996.
20. B. Eiseman, R. Wotkyns. *Surgical Decision Making*. Saunders, Philadelphia, 1978.
21. H.T. Engelhardt, S.F. Spicker, B. Towers. *Clinical Judgment: A Critical Appraisal*. Riedel, Dordrecht, 1977.
22. A.R. Feinstein. *Clinical Judgment*. Williams and Wilkins, Baltimore, MD, 1957.
23. M. Henle. On the relation between logic and thinking. *Psych. Rev.*, 69: 366–578, 1962.
24. M. Hoffman. Diseases of heart and blood vessels. In R. Brzozowski, editor, *Vademecum of Diagnosis and Theraphy*, 97–184, Wyd. Lek. PZWL, Warszawa, 1993 (in Polish).
25. G. Hripcsak. Writing arden syntax medical logic modules. *Comp. Biol. Med.*, 24: 331–363, 1994.
26. K.D. Forbus Qualitative process theory. In D.G. Bobrow, editor, *Qualitative Reasoning about Physical Systems*, MIT Press, Cambridge, MA, 1995.
27. B. Kuipers. Qualitative simulation. *Artificial Intelligence*, 29: 289–338, 1986.
28. H. Llewelyn, A. Hopkins, editors. *Analysing How We Reach Clinical Decisions*. Royal College of Physicians, London, 1993.
29. J.L. Mackie. *The Cement of the Universe. A Study of Causation*. Oxford University Press, Oxford, 1974.
30. W. Marciszewski. *Methods of Analysis of a Scientific Text*. PWN, Warsaw, 1977 (in Polish).
31. E.A. Murphy. *The Logic of Medicine*, (2nd ed.) Johns Hopkins University Press, Baltimore, MD, 1978.
32. J. Packard, J.E. Faulconer. *Introduction to Logic*. Van Nostrand, New York, 1980.
33. J. Pearl. *Causality — Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2000.
34. H. Reichenbach. *Elements of Symbolic Logic*. Macmillan, New York, 1948.
35. H. Reichenbach. *The Theory of Probability*. University of California Press, Berkeley, CA, 1949.
36. C. Rieger. An organization of knowledge representation. *Artificial Intelligence*, 7: 89–127, 1976.
37. N. Sager, C. Friedman, M.S. Lyman. *Medical Language Processing. Computer Management of Narrative Data*. Addison-Wesley, Reading, MA, 1987.
38. W.C. Salmon. *The Foundations of Scientific Inference*. University of Pittsburgh Press, Pittsburgh, PA, 1967.
39. S. Schwartz, T.M. Griffin. *Medical Thinking. The Psychology of Medical Judgment and Decision Making*. Springer, New York, 1986.
40. J. Stausberg, M. Person. A process model of diagnostic reasoning in medicine. *Journal of Medical Informatics*, 54: 9–23, 1999.
41. P. Szolowitz, S.G. Pauker. An organization of knowledge representation. *Artificial Intelligence*, 11: 115–144, 1978.

42. M.P.Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44(3): 257–303, 1990.
43. H.J. Wright, D.B. Macadam. *Clinical Thinking and Practice*. Churchill Livingstone, Edinburgh, 1979.
44. H.R. Wulff, S.A. Pedersen, R. Rosenberg. *Philosophy of Medicine. An Introduction*. Blackwell, Oxford, 1990.
45. L.A. Zadeh. A new direction in AI: Toward a computational theory of perceptions. *AI Magazine*, 22(1): 73–84, 2001.
46. D. Zigmond, L.A. Lenert. Monitoring free-text data using medical language processing. *Computers and Biomedical Research*, 26: 467–481, 1993.