International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Joint Action of Binary Factors in the Sufficient Causes Theory and Its Classification

Julia V. Nagrebetskaya, Vladimir G. Panov

Abstract*: *We consider a problem of formal definition of joint action in the binary sufficient causes framework based on the theory of Boolean algebras. This theory is one of the general causality concepts in epidemiology, environmental sciences, medicine and biology. Its correct mathematical form allows us to regard the binary version of this theory as a specific application of Boolean functions theory. Within the formalism of Boolean functions, a strict definition of the joint action is given and various criteria for the presence of joint action of factors in a Boolean function are obtained. The methods previously developed for analyzing joint action in binary sufficient causes framework allows us to split all the Boolean functions into disjoint equivalence classes. The relationships among these classes however remain uncertain. In the present paper, an integer invariant is introduced which allows one to order joint action types in a certain way. We consider examples of two- and three-factor theories of sufficient causes with the ordinary epidemiological symmetry group. Estimation of the time complexity of determining the type of joint action are considered as well.

Keywords: Boolean algebra, Boolean function, joint action, group action on a set, integer-valued invariant, sufficient causes theory, time complexity

I. INTRODUCTION

An important theoretical and practical problem of the environmental regulatory authorities is to assess a type of joint impact of a multicomponent burden on the environment and population of industrial cities [1-4]. This problem is closely related to the problem of assessing the type of joint action of toxic substances or physical factors in toxicology and pharmacology [5–8]. However, toxicological studies as a rule consider a small number of acting factors due to the difficulty of understanding multifactorial effects and controversial interpretation of the results of application complex multivariate models [9–11]. In addition, in many cases acting factors are the causes of a disease which prevention often requires not only eliminating the actual exposure of the harmful factor that is not always possible but, rather, an accurate assessment of which a joint effect of these factors is. For instance, it is known that the cyanides in industrial wastes are quite poisonous to aquatic life. However, in presence of nickel, a nickel-cyanide complex is formed whose toxicity is comparatively low [12].

Chelating agents, such as calcium disodium defeated, fall into the category of antagonists and operate to minimize the lethal effects of heavy metals such as mercury or lead [13].

Thus, the presence of certain agents in a multicomponent mixture affecting the environment or living systems can significantly weaken or eliminate the total effect of the mixture. The knowledge of such effects can change the assessment of severity of joint effects of pollutants. For example, if the joint effect is known to be weaker than the sum of one-factor effects, this saves money or resources on the necessary measures to reduce the environmental burden of pollutants.

One of the causal analysis model in epidemiology and environmental sciences is the so-called *sufficient causes framework* [14–21]. This approach addresses, inter alia, representations of various causation mechanisms of acting factors in a response, as well as the problem of identification possible acting agents' synergism [15,18–21].

These issues were discussed in detail in [15,17,19–21] for the binary theory. As it turned out, it is possible to build a formalized model of sufficient causes theory which, however, does not adequately represent both the initial ideas of that theory and means of research.

A more rigorous formal presentation of the binary sufficient causes theory is possible on the basis of finite Boolean algebras theory. That was shown in [22,23] for two variables and in [24] for general case of n variables. An adequate mathematical apparatus is presented therein for studying a classification of factors joint action types in the binary theory of sufficient causes.

It is important to note that the problem of classification joint action types from the viewpoint of risk assessment is not posed explicitly and remains unclear in all formal models of the theory of sufficient causes.

Briefly, the structure of the binary theory of sufficient causes can be described as follows [22–24]. The state space of a binary experiment forms a finite Boolean algebra consisting of the set of all binary vectors of length n. The response (outcome) is considered as a Boolean function on this space and the set of all responses forms the Boolean algebra of all Boolean functions defined on the state space. Empirical symmetries [15,17,21] play an important role in the structure of a binary experiment.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Julia V. Nagrebetskaya, Department of Mathematics, Mechanics and Computer Science, Institute of Natural Sciences and Mathematics, Ural Federal University, Ekaterinburg, Russia. Email: I.V.Nagrebetckaia@urfu.ru
Vladimir G. Panov*, Laboratory of Mathematical Modelling in Ecology and Medicine, Institute of Industrial Ecology Ural Branch of RAS, Ekaterinburg, Russia. Email: vpanov@ecko.uran.ru

In the proposed formalization [22–24] these symmetries can be written as automorphisms on the algebra of Boolean functions. These automorphisms generate a group of experiment's symmetries that action on the Boolean algebra of responses determines classes of equivalent responses representing different types of joint action.

In the usual epidemiological situation, these symmetries form a group of all symmetries of the *n*-dimensional cube [22–24]. As a result, it is impossible to distinguish the types of joint action which could be considered antagonistic (in a sense, weakening the total effect of agents) from the synergistic types (that are stronger than the sum of one-factor effects). This feature of the sufficient causes theory with such a group of symmetries requires a more careful study.

Today it can be argued that, depending on the symmetry group under consideration, one can obtain a complete list of response functions that forms a certain type of joint action, i.e. response functions that are in one such a class represent the same type of joint action, and functions from different classes have different types of joint action. However, the relationship between these classes remains uncertain. For example, it cannot be said *a priori* that one type of joint action is stronger or weaker than another.

In the present paper, we consider the problem of ordering the types of joint action with the help of an integer invariant on the space of all Boolean functions defined on a finite Boolean algebra of experiment states. This invariant allows us to order the types of joint action in such a way that it can be interpreted as some characteristic of the joint action strength.

In order the mathematical formalism to be effective in application to the binary sufficient causes theory basic concepts of that theory should be translated to the Boolean algebra language. This formalization allows us to consider analysis of joint action of binary factors as a specific application of Boolean algebras theory. That is why the main statements that have direct applications in epidemiology and medicine are given as mathematical theorems, although in most cases without a proof. However, we should remember that "Eine gute Theorie ist das Praktischste was es gibt", or "A good theory is the most practical thing" (Kant, Kirchhoff and many others). Note, however, that the definitions of notions used below as well as omitted proofs can be easily recovered using the sources from the reference list.

II. PROBLEM FORMULATION AND NOTATIONS

A.Mathematical framework

A complete description of the formalism of binary sufficient causes theory based on Boolean algebras theory is presented in [22–24]. The paper [24] also describes the relationship between the theory of sufficient causes and the Neyman-Holland-Rubin causality model [25]. The general formalism of Boolean algebras and Boolean functions see in [26–28]. We here describe briefly the basic concepts of the Boolean formalization and notations used in that follows.

In biomedical and environmental studies acting factors $X_1, X_2,\dots, X_n$ may be the presence of some harmful agent (toxin, chemical agent, pollutant, some physical factor etc.) at a certain dose, concentration or exposure. A response Y represents a certain effect which has often two levels, for example, indicating that value of a physiological index is less than or greater than some threshold value, or just the absence or appearance of some effect.

In the binary theory, it is assumed that the levels of acting factors and response take on two possible values that can be encoded with the numbers 0 and 1.

As shown in [22–24], the formalization of the binary sufficient causes theory can be expressed by Boolean functions notions, for example, response Y is represented by a Boolean function f of Boolean variables $x_1,\dots, x_n$, which encode the acting factors $X_1, X_2,\dots, X_n$.

More precisely, it can be said that in the formal model of the sufficient causes theory there are two components representing different aspects of the experiment in which the variables X_1, X_2, ..., X_n and the response Y arise. First is the space of experiment states, i.e. set of possible values of the independent variables $x_1,\dots,x_n$. Obviously, the set of experiment states forms a Boolean cube $\mathbb{B}^n$. Second is the space of all responses defined on the set of experiment states. This space forms a free Boolean algebra $\mathbb{B}(x_1,\text{K} ,x_n)$ of all Boolean functions of $x_1,\dots, x_n$.

According to this binary sufficient causes theory representation the problem of classifying the types of joint action of factors X_1, X_2, ..., X_n should be posed as the problem of calculating orbits of a certain group of automorphisms action on the algebra $\mathbb{B}(x_1,\text{K} ,x_n)$ [22–24].

We give examples of such a formalization in the case of two acting factors represented by the Boolean variables x_1, x_2. In this case the space of experiment states is the Boolean cube $\mathbb{B}^2$ consisting of $2^2 = 4$ elements. The responses' space is the Boolean algebra $\mathbb{B}(x_1,x_2)$, consisting of $2^{2^2}=16$ functions. Each such a function can be represented in various forms one of which is the canonical disjunctive normal form (canonical DNF) [27]. Thus, any function on $\mathbb{B}(x_1,x_2)$ can be written in the form

$$f(x_1,x_2)= f_{00}\overline{x}_1\overline{x}_2 \vee f_{01}\overline{x}_1 x_2 \vee f_{10}x_1\overline{x}_2 \vee f_{11}x_1x_2 = \bigvee_{\alpha,\beta\in\mathbb{B}} f_{\alpha\beta}x_1^{\alpha}x_2^{\beta},$$

where

$$f_{\alpha\beta} = f(\alpha,\beta),\ \alpha,\beta\in\mathbb{B},\ x_i^{\gamma} = \begin{cases} x_i, \text{if } \gamma=1 \\ \overline{x}_i, \text{if } \gamma=0 \end{cases},\ \mathbb{B}=\{0,1\}.$$

As shown in [17, 18, 21–24], a juxtaposition of different responses to determine whether they are identical in the nature of their joint action makes significant use of symmetry considerations. In [17, 18, 21] this is presented in an informal descriptive form. A more formal representation in the Boolean formalism leads to the fact that, along with the constructions introduced above, it is necessary to consider a certain group of automorphisms on the response algebra. This group is usually generated by some empirical symmetries, which can be formalized as automorphisms on the space of Boolean functions [22–24].

For instance, for the two-factor case described above, natural symmetries are determined by the following transformations [22–24]

(A1) The character of the joint action produced by the factors X_1 and X_2 is the same as for the factors X_2 and X_1.

(A2) The character of the joint action produced by the factors X_1 and X_2 is the same as for the factors $\overline{X}_1$ and X_2.

As shown in [22–24], these conditions can be written as transformations on the set of literals as follows (transformations T_1 and T_2 correspond to symmetries (A1) and (A2) respectively)

$$T_1(x_1) = x_2, \quad T_1(x_2) = x_1$$
$$T_2(x_1) = \overline{x}_1, \quad T_2(x_2) = x_2,$$

and then continue to automorphisms of the Boolean cube $\mathbb{B}^2$ by the equality

$$T_i\left(x_1^\alpha x_2^\beta\right) = T_i(x_1)^\alpha\, T_i(x_2)^\beta,\ \alpha,\beta \in \mathbb{B},\ i = 1,2$$

Automorphisms T_1 and T_2 generate a group G of all automorphisms of the Boolean cube $\mathbb{B}^2$ which is being considered here as a graph [29]. Geometrically, this group is a group of all symmetries of a square [22,23]. Acton of the group G on the Boolean cube $\mathbb{B}^2$ continues to the action of that group on the algebra $\mathbb{B}(x_1, x_2)$ of all Boolean functions of two variables x_1, x_2 in such a way

$$T_i\left(f(x_1,x_2)\right) = f_{00}T_i(\overline{x}_1\overline{x}_2) \vee f_{01}T_i(\overline{x}_1x_2) \vee$$
$$f_{10}T_i(x_1\overline{x}_2) \vee f_{11}T_i(x_1x_2),\ \text{where } f_{\alpha\beta} = f(\alpha,\beta)$$

The action of this group on the response's space forms various classes of equivalent responses. For the considered two-factor case, we obtain the following classes written by one of its representatives in angle brackets [22–24]

$$\langle 0\rangle = \{0\}, \quad \langle 1\rangle = \{1\}, \quad \langle x_1\rangle = \{x_1, \overline{x}_1, x_2, \overline{x}_2\}$$
$$\langle x_1 \vee x_2\rangle = \{x_1 \vee x_2, \overline{x}_1 \vee x_2, x_1 \vee \overline{x}_2, \overline{x}_1 \vee \overline{x}_2\}$$
$$\langle x_1x_2\rangle = \{x_1x_2, \overline{x}_1x_2, x_1\overline{x}_2, \overline{x}_1\overline{x}_2\}$$
$$\langle x_1x_2 \vee \overline{x}_1\overline{x}_2\rangle = \{x_1x_2 \vee \overline{x}_1\overline{x}_2, \overline{x}_1x_2 \vee x_1\overline{x}_2\}$$

Thus, the presented formalism allows us to obtain a complete list of Boolean functions that have the same type of joint action of factors, though relationships among these classes remains unclear in most cases. Moreover, an increase in the number of acting factors leads to a steep increase in the number of joint action types and complication of classes' structure.

For example, for three-factor sufficient causes theory with symmetries similar to (A1) and (A2), we obtain 22 classes of equivalent responses some of which have a rather complicated structure [24]. This makes it difficult to analyse them in terms of their joint action nature. For example, it is unclear which of the following responses could be considered to have a stronger joint effect of factors

$$x_1x_2x_3 \vee \overline{x}_1\overline{x}_2\overline{x}_3 \text{ or } x_1x_2x_3 \vee \overline{x}_1\overline{x}_2x_3$$

We will propose below an integer invariant μ which allows us to order classes of equivalent responses in such a way that responses that do not have a joint action at all, such as constant functions $f = 0$, $f = 1$ or one-factor functions $f = x_1$, $f = x_2$, ..., $f = x_n$, have zero value μ, and the response $x_1x_2\ldots x_n$ (conjunction of variables $x_1, x_2, \ldots, x_n$) has a maximum value of μ.

B. Boolean versions of some notions of sufficient causes theory

In order to get an adequate apparatus for analyzing the problem posed, it is necessary to supplement the general formalism described in the previous section with some new concepts, which are mathematical expressions for more complex concepts of the sufficient causes theory. The following definitions are known in the theory of Boolean functions [26–28]. We recall these definitions together with the corresponding notions from the theory of sufficient causes.

Definition 1. A Boolean function g from n variables is called *an implicant* of a Boolean function f from n variables if for all vector $\boldsymbol{\alpha} \in \mathbb{B}^n$ the equality $g(\boldsymbol{\alpha}) = 1$ implies the equality $f(\boldsymbol{\alpha}) = 1$. Conjunction of a subset of literals. i.e. Boolean variables $x_1, \ldots, x_n$ or their negation, which is an implicant of a Boolean function f is called its *prime implicant*, if removal of any literal from this conjunction results in a non-implicant for it.

Prime implicant is a mathematical expression of a notion of *minimal sufficient cause* [20,21] of a response f, since by definition [20,21], there is no excessive variable in the record of a minimum sufficient cause, i.e. there is no such literal that can be removed and the resulting conjunction remains a sufficient cause. The set of minimal sufficient causes for the response f, (i.e., the prime implicants of the Boolean function f) those disjunction is equal to f, is called *determinative set of sufficient minimal causes* [20,21]. It is well known that any Boolean function f can be represented as a disjunction of all its prime implicants (so-called *complete DNF*, see [27]). However, this representation may contain *redundant* prime implicants. Such implicants can be removed (not all at the same time) from the representation f as a disjunction of prime implicants without violating the equality. Representation of a response f as a disjunction of prime non-redundant implicants is called *irredundant* [27], and the set of corresponding prime implicants in [20,21] is called *non-redundant determinative set of minimal sufficient causes*.

In [21,30,31] the concept of a *sufficient cause exhibiting a sufficient cause interaction* in the response f is introduced. This concept can also be formulated in the Boolean functions language. Below we call it *joint or combined action* of given factors (this terms are common in biomedical sciences).

Definition 2. We say that there is *joint action* of factors $x_1, \ldots, x_n$ in a response f depending on n variables $x_1, \ldots, x_n$ if such a vector $\boldsymbol{\alpha} \in \mathbb{B}^n, \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ exists that a conjunction $\boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1}x_2^{\alpha_2}\ldots x_n^{\alpha_n}$ presents in every irredundant representation of the Boolean function f. We also say that in this case the joint action in the response f *attains at* $\boldsymbol{x} = \boldsymbol{\alpha}$.

It should be stressed that in Definition 2 a specific type of general notion of joint action is introduced. Arbitrary function $f(x_1,\ldots,x_n)$ exhibits, in a sense, some kind of joint action though Definition 2 might not be fulfilled. Below we consider joint action only in the sense of Definition 2.

We introduce *support* C_f of a Boolean function f by the following equality

$C_f = \{\boldsymbol{\alpha} \in \mathbb{B}^n \mid f(\boldsymbol{\alpha}) = 1\}$. It is obvious then that if the joint action of factors attains at $\boldsymbol{x} = \boldsymbol{\alpha}$ in the response *f*, then $\boldsymbol{\alpha} \in C_f$.

It should also be noted that in the theory of Boolean algebras, geometric reasoning is widely used based on the representation of a Boolean cube as a graph. Namely, in the graph $\mathbb{B}^n$ vertices are all binary vectors of the length *n*, and edges connect those vertices Hamming distance between which equals to 1. Denote by Γ_f a subgraph of the graph $\mathbb{B}^n$ those vertices are points in C_f and edges are corresponding edges in $\mathbb{B}^n$. It is clear that the graph Γ_f and the set C_f are different forms of presentation of a Boolean function (i.e. response or outcome) *f*.

Another means widely used in the study of DNF is covering of the set C_f by faces of the Boolean cube $\mathbb{B}^n$ [27]. Since the theory of sufficient causes is based on representation of a response in DNF, the coverings serve a very effective tool for this theory as well. Recall, that a set $B_I^{\boldsymbol{\alpha}} = \{\boldsymbol{\beta} \in \mathbb{B}^n \mid \boldsymbol{\beta}_I = \boldsymbol{\alpha}\}$ is called *a face* (or *a subcube*) of the Boolean cube $\mathbb{B}^n$ where

$I = \{i_1, \ldots i_k\}, 1 \le i_1 < i_2 < \ldots < i_k \le n, \boldsymbol{\alpha} \in \mathbb{B}^k,$

$\boldsymbol{\beta}_I = (\beta_{i_1}, \ldots, \beta_{i_k}).$

This face is also called $n-k$-face since it has dimension $n-k$. A face $B_I^{\boldsymbol{\alpha}}$ is called a *face of Boolean function f* if $B_I^{\boldsymbol{\alpha}} \subseteq C_f$. All faces of a Boolean function *f* form a partially ordered set which maximal elements are called *maximal face of the function f*. A family of faces of a function f whose union is equal to the set C_f is called a *covering* of the set C_f or a *covering* of the Boolean function *f*. A covering of a Boolean function f with maximal faces which is minimal by inclusion is called its *irredundant covering*.

A relation between the implicants of the Boolean function *f* and its faces is established by the following Lemma.

Lemma 1 [27]. Conjunction $\boldsymbol{x}_I^{\boldsymbol{\alpha}} = x_{i_1}^{\alpha_{i_1}} x_{i_2}^{\alpha_{i_2}} \ldots x_{i_k}^{\alpha_{i_k}}$ is a (prime) implicant of a Boolean function *f* if and only if the $n-k$-face $B_I^{\boldsymbol{\alpha}}$ is a (maximal) face of the function *f*.

Thus, each representation of the response *f* in the form of disjunction of conjunctions of literals that are (prime) implicants of f corresponds to a certain covering of the set C_f with (maximal) faces of this function. Similarly, each irredundant representation of the function *f* corresponds to an irredundant covering of this function.

Therefore, we can establish the following correspondence between concepts of sufficient causes theory and geometric concepts of the theory of Boolean functions. Each (minimal) sufficient cause of a response *f* is geometrically represented by a (maximal) face of the corresponding Boolean function *f*. Each of the determinative set of (minimal) sufficient causes is represented by a covering with (maximal) faces of the function *f*. Each of the non-redundant determinative set of minimal causes is represented by an irredundant covering of the function *f*.

III. MAIN RESULTS

The criterion of the presence of joint action in the binary sufficient causes theory given in [20,21] can be formulated in terms of Boolean functions as follows.

Theorem 1. There is joint action of factors $x_1,\ldots,x_n$ in an response f which attains at $\boldsymbol{x} = \boldsymbol{\alpha}$, $\boldsymbol{\alpha} \in \mathbb{B}^n$ if and only if the conjunction $\boldsymbol{x}^{\boldsymbol{\alpha}}$ presents in every DNF equivalent to the Boolean function f.

In terms of sufficient causes this means that $\boldsymbol{x}^{\boldsymbol{\alpha}}$ is a minimal sufficient cause presented in every determinative set of sufficient causes for the response *f*.

Another criterion for the presence of join action of factors is

Theorem 2. There is joint action of factors $x_1,\ldots, x_n$ in a response f which attains at $\boldsymbol{x} = \boldsymbol{\alpha}$ if and only if the conjunction $\boldsymbol{x}^{\boldsymbol{\alpha}}$ is a prime implicant of the Boolean function *f*.

From Lemma 1 and Theorem 2 we obtain the following joint action criterion expressed in geometric terms.

Theorem 3. There is joint action of factors $x_1,\ldots, x_n$ in a response f which attains at $\boldsymbol{x} = \boldsymbol{\alpha}$ if and only if any single point set $\boldsymbol{\alpha}$ is a maximal 0-face of the Boolean function f.

From computational viewpoint, it is interesting how and what time is required to determine whether the joint action of n factors in a given response attains at $\boldsymbol{x} = \boldsymbol{\alpha}$. The answer to this question follows from Theorem 4.

Theorem 4. There is joint action of factors $x_1,\ldots, x_n$ in a response f which attains at $\boldsymbol{x} = \boldsymbol{\alpha}$ if and only if the point $\boldsymbol{\alpha}$ is an isolated vertex in the graph Γ_f.

Corollary 1. Running time (time complexity) of checking for presence of joint action of *n* factors at $\boldsymbol{x} = \boldsymbol{\alpha}$ in the response f is of order $O(2^n \cdot n)$.

Proof. It is obvious that the graph $\mathbb{B}^n$ has $N = 2^n$ vertices. Since the graph Γ_f is a subgraph of the graph $\mathbb{B}^n$, it can be defined by a list of vertices from the support C_f. Each of these vertices can be represented by a binary vector of length *n*. Check if a given vertex from C_f has zero degree takes a time no more than $O(2^n \cdot n)$. Indeed, let us act on each vertex of the set C_f by an automorphism *t* of the Boolean cube $\mathbb{B}^n$ defined as follows: $t(x) = x \oplus \boldsymbol{\alpha}$. Here the symbol $a \oplus b$ means modulo 2 addition. In the resulting set $C = t(C_f)$ the number of vertices of unit Hamming weight coincides with the degree of the vertex $\boldsymbol{\alpha}$ in the graph Γ_f. Checking for the absence of such vertices in C can be done in time no more than $O(2^n \cdot n)$ as the sets C and C_f have equal cardinality.

The estimation $O(2^n \cdot n)$ of the time complexity cannot be improved as in the worst case the graph Γ_f has 2^n vertices each of which is coded with a binary vector of length n.

Note, that the estimate $O(2^n \cdot n)$ is of *exponential* order with respect to the number n of the acting factors, and it is of order $O(N)$, with respect to the largest input size $N = 2^n \cdot n$.

From Definition 2 and Corollary 1 we obtain

Corollary 2. Running time of checking for the presence of joint action of n factors in the response f is of order no more than $O(4^n \cdot n)$.

Proof. Indeed, applying to each vertex from the support C_f algorithm from the proof of Corollary 1, we obtain the desired algorithm for checking for the presence of joint action of n factors in the response f. Therefore, its time complexity is no more than $O(4^n \cdot n)$.

In the theory of Boolean functions, the Quine algorithm for finding the set of all prime implicants (respectively, maximal faces) of the Boolean function f is known [28].

The above algorithm is a modification of the first stage of the Quine algorithm for finding implicants of length $n - 1$ (respectively, 1-faces) of Boolean function f.

As already noted, a type of joint action of factors in a response should not depend on the coding of acting factors' levels and their order (see, Section II.B). This statement is formalized by defining the group G of automorphisms (symmetries) on the Boolean algebra of responses [22–24]. In usual epidemiological theory of sufficient causes with n factors the automorphism group G is isomorphic to the group of all symmetries of the n-dimensional cube (i.e. hypercube) [24]. Action of this group on the Boolean algebra of responses generates a partition of this algebra into equivalence classes. It easily follows from Theorem 2 that the presence of joint action of factors holds or does not hold simultaneously for all Boolean functions from the same class, i.e. it is a property of the whole class of equivalent functions. Therefore, each such a class can be considered *a class*, or *a type of joint action of factors* (in a more general sense than it was defined in Definition 2).

From Theorem 2–4 follows

Theorem 5. The class $\langle f \rangle$ is a class representing a type of joint action of factors $x_1,\dots, x_n$ if and only if any of the following equivalent conditions is satisfied

(1) Boolean function f has a prime implicant consisting of exactly n literals;

(2) there is a 0-face among maximal faces of Boolean function f;

(3) there is an isolated vertex in the graph Γ_f.

Example 1. For $n = 2$ classes representing nontrivial joint action of n factors are only the following $\langle x_1 x_2 \rangle, \langle x_1 x_2 \vee \overline{x}_1 \overline{x}_2 \rangle$.

Example 2. For $n = 3$ classes representing nontrivial joint action of n factors are only the following classes $\langle x_1 x_2 x_3 \rangle, \langle x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 \overline{x}_3 \rangle, \langle x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 \rangle$

$\langle x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 x_3 \vee \overline{x}_1 x_2 \overline{x}_3 \vee x_1 \overline{x}_2 \overline{x}_3 \rangle$,

$\langle x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 x_3 \vee \overline{x}_1 x_2 \overline{x}_3 \rangle, \langle x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 x_3 \rangle$,

$\langle x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 \vee \overline{x}_2 \overline{x}_3 \vee \overline{x}_1 \overline{x}_3 \rangle, \langle x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 \vee \overline{x}_1 \overline{x}_3 \rangle$.

As can be seen from these examples, it is sometimes difficult to understand a specificity of joint action (i.e. interaction) of factors and to compare classes among themselves by the structure of minimal DNF representing these classes. In this regard, the following statement is useful.

Definition 3. We call *degree of joint action of factors* $x_1,\dots, x_n$ in a response f at values of factor levels $\boldsymbol{x} = \boldsymbol{\alpha}$ a number

$$\mu_f(\alpha) = \begin{cases} \min\{d(\alpha,\beta) \mid \beta \in C_f, \beta \neq \alpha\} - 1, & \text{if } |C_f| > 1 \\ n, & \text{if } |C_f| = 1 \end{cases}$$

where $|C_f|$ is the cardinality of C_f and $d(\alpha,\beta)$ is Hamming distance between vectors α and β. As follows from the Definition 3, the inequality $0 \le \mu_f(\alpha) \le n$ holds.

It can be shown that the time complexity of calculating $\mu_f(\alpha), \alpha \in \mathbb{B}^n$ is equal to $O(2^n \cdot n)$.

The following Theorem 6 permits a better insight into the joint action concept.

Theorem 6. There is joint action of n factors in a Boolean function f at $\boldsymbol{x} = \boldsymbol{\alpha}$ if and only if $\mu_f(\alpha) \ge 1$.

Proof. If there is joint action in the response f at $\boldsymbol{x} = \boldsymbol{\alpha}$ then by Theorem 4 the point $\boldsymbol{\alpha}$ is an isolated vertex in the graph Γ_f. Hence, in particular, $\alpha \in C_f$. If $|C_f| = 1$ then $\mu_f(\alpha) = n \ge 1$. Let be $|C_f| > 1$. Since the point $\boldsymbol{\alpha}$ is isolated in the graph Γ_f then the inequality $d(\alpha,\beta) \ge 2$ holds for any point $\beta \in C_f$, i.e. $\mu_f(\alpha) \ge 1$. The reverse statement is proved similarly.

The following geometric meaning of the number $\mu_f(\alpha)$ can be noted. From Definition 3, it is clear that value $\mu_f(\alpha) + 1$ is the minimum distance from the point $\boldsymbol{\alpha}$ to the set of faces that, together with 0-face $\{\boldsymbol{\alpha}\}$, enter in every covering of the Boolean function f. In terms of sufficient causes, this means that this value is a "distance" from the minimal sufficient cause $\boldsymbol{x}^\alpha$ to any other sufficient cause of the response f. This means that the minimal sufficient cause $\boldsymbol{x}^\alpha$ has "significant influence" at a distance of no more than $\left\lfloor \frac{\mu_f(\alpha)}{2} \right\rfloor$ if $|C_f| > 1$, and at a distance n if $|C_f| = 1$.

The following property is also useful for understanding of the number $\mu_f(\alpha)$. As follows from Theorem 4, a necessary and sufficient condition for the presence of the joint action of n factors in the response f which attains at $\boldsymbol{x} = \boldsymbol{\alpha}$ is changing the value of the response f from 1 to 0 when changing the value of *any* factor. This fact was also noted in [20]. It follows from Definition 3 that the response f changes its value from 1 to 0 when changing values of *any number* of factors but no more than $\mu_f(\alpha)$.

In connection with these properties of the degree of joint action we can add a characterization of the number $\mu_f(\alpha)$ in terms of derivatives of Boolean functions [32]. Recall the definition of that derivative.

Definition 4. *The derivative* of the Boolean function *f in direction* of a vector $\boldsymbol{a} \in \mathbb{B}^n$ is called a Boolean function $f'_{\boldsymbol{a}}(\boldsymbol{x}) = f(\boldsymbol{x}) \oplus f(\boldsymbol{x} \oplus \boldsymbol{a})$.

Let be $f(\alpha) = 1$ for an $\alpha \in \mathbb{B}^n$. Then $f'_{\boldsymbol{a}}(\alpha) = 1$ if $f(\alpha \oplus \boldsymbol{a}) = 0$ and $f'_{\boldsymbol{a}}(\alpha) = 0$ if $f(\alpha \oplus \boldsymbol{a}) = 1$.

Theorem 7. The degree $\mu_f(\alpha)$ of joint action equals to a maximum of those numbers *m* for which the equality $f'_{\boldsymbol{a}}(\alpha) = 1$ holds for any nonzero vector $\boldsymbol{a}$ with Hamming weight no more than *m*.

Proof. Let $\boldsymbol{a} \in \mathbb{B}^n$ be a nonzero vector with Hamming weight no more than $\mu_f(\alpha)$. Then for a point $\beta = \alpha \oplus \boldsymbol{a}$ the equality $d(\alpha, \beta) = w(\alpha \oplus \beta) = w(\boldsymbol{a})$ holds and, therefore, we obtain an inequality $d(\alpha, \beta) \le \mu_f(\alpha)$ for $\beta \ne \alpha$ where $w(\gamma)$ denotes Hamming weight of a point $\gamma \in \mathbb{B}^n$. By Definition 3, it follows that $\beta \notin C_f$, i.e. $f(\beta) = 0$ and $f'_{\boldsymbol{a}}(\alpha) = 1$.

Now let us have $\mu_f(\alpha) \le n - 1$ and $m = \mu_f(\alpha) + 1$. Then a point $\beta \in C_f$ exists such that $d(\alpha, \beta) = m$, i.e. Hamming weight of the vector $\boldsymbol{a} = \alpha \oplus \beta$ equals to *m*. Thus, for a nonzero vector $\boldsymbol{a}$ with Hemming weight greater than $\mu_f(\alpha)$ the equalities $f(\alpha \oplus \boldsymbol{a}) = f(\beta) = 1$ hold, i.e. $f'_{\boldsymbol{a}}(\alpha) = 0$.

From Theorem 6 and 7 it follows

Corollary 3. There is joint action of the factors $x_1, \ldots, x_n$ in the response *f* which attains at $\boldsymbol{x} = \alpha$ if and only if $f'_{\boldsymbol{a}}(\alpha) = 1$ for any vector $\boldsymbol{a} \in \mathbb{B}^n$ of unit Hemming weight.

The Definition 3 introduces the degree of joint action of factors in a given Boolean function for *given values* of its arguments. However, the same concept can be defined for a Boolean function as such.

Definition 5. We call the degree of joint action of factors $x_1, \ldots, x_n$ in a response *f* a number $\mu_f = \max\{\mu_f(\alpha) \mid \alpha \in C_f\}$.

Obviously, the degree of joint action for *n* factors is invariant with respect to the action of the hypercube symmetry group, i.e. this value is correctly defined for a class of equivalent responses.

As mentioned above, the time complexity of calculation of $\mu_f(\alpha)$ for a given $\alpha \in \mathbb{B}^n$ is of order $O(2^n \cdot n)$. Therefore, the time complexity of calculating the value μ_f is not more than $O(4^n \cdot n)$.

From Theorem 6 follows

Theorem 8. There is joint action of factors $x_1, \ldots, x_n$ in a response *f* if and only if $\mu_f \ge 1$.

Corollary 4. Class $\langle f \rangle$ represents joint action of the factors $x_1, \ldots, x_n$ if and only if the inequality $\mu_f \ge 1$ holds.

From Definitions 3 and 5 we now obtain

Corollary 5. The greatest degree $\mu_f = n$ of joint action of factors $x_1, \ldots, x_n$ has only the conjunction class $\langle x_1 x_2 \ldots x_n \rangle$.

Since the conjunction of all variables $x_1 x_2 \ldots x_n$ can be considered a response with the strongest joint action (an analogue of the product of predictors in regression analysis), and responses without a joint action are characterized by the condition $\mu_f = 0$, it follows from Corollary 5 that the number μ_f is indeed a feature of joint action power.

Example 3. For *n* = 2, the degree of joint action for the classes from Example 1 is: $\mu_f = 2$ for $f = x_1 x_2$ and $\mu_f = 1$ for $f = x_1 x_2 \vee \overline{x}_1 \overline{x}_2$. The responses *f* for which $\mu_f = 0$ (representing the absence of joint action of factors x_1 and x_2), are included in the classes $\langle 0 \rangle, \langle 1 \rangle, \langle x_1 \rangle, \langle x_1 \vee x_2 \rangle$.

Example 4. For *n* = 3, the degree of joint action for the classes from Example 2 is: $\mu_f = 3$ for $f = x_1 x_2 x_3$, $\mu_f = 2$ for $f = x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 \overline{x}_3$ and $\mu_f = 1$ for the remaining classes. Particularly, now can answer the question posed in the Introduction. Since $\mu_f = 1$ for $f = x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 x_3$ and $\mu_f = 2$ for $f = x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 \overline{x}_3$ the joint action in the response $f = x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 \overline{x}_3$ should be considered *stronger* then joint action in the response $f = x_1 x_2 x_3 \vee \overline{x}_1 \overline{x}_2 x_3$.

Responses without a joint action of the factors x_1, x_2, x_3 (for which $\mu_f = 0$) are included in the following classes $\langle 0 \rangle, \langle 1 \rangle, \langle x_1 \rangle, \langle x_1 \vee x_2 \rangle, \langle x_1 \vee x_2 \vee x_3 \rangle$, and in other classes whose representatives recorded in minimal DNF have prime implicants containing no more than two literals.

A comparison of the responses with $\mu_f = 0$ for *n* = 2 and *n* = 3 allows one to suggest that in general for the presence of a joint action of all factors $x_1, \ldots, x_n$ in a given response it is necessary that in the minimal DNF representing this response at least one conjunction has to contain all these factors or their negations. As shown in Theorem 1, this is indeed so and it agrees well with the term joint action of *all factors*. In this connection, the need arises for additional analysis of those responses *f* for which $\mu_f = 0$ and , however, there are conjunctions of at least two literals, for example, $x_1 x_2 \vee x_2 x_3 \vee x_3 x_1$. Obviously, such a response has a stronger joint action than, for example, a constant response equals to 0 or 1. However, for all these functions $\mu_f = 0$.

IV. CONCLUSION

We examined the Boolean formalism of the epidemiological binary theory of sufficient causes in the context of constructing an invariant to describe the strength of joint action. It is shown that this problem can be solved within the Boolean framework by introducing a new notion of the *degree of joint action*. We also provide precise mathematical concepts that express known epidemiological notions.

It is shown that in the formal description and mathematical analysis of the theory of sufficient causes graph theory methods can be effectively applied in addition to the formalism of Boolean functions theory.

We discussed in detail the very concept of joint action, giving its strict definition as well as several criteria for its presence in a given response. This allows one to obtain a rigorous foundation for a formal analysis of the binary sufficient causes theory.

In connection with the demonstration that the epidemiological binary theory of sufficient causes can be considered as a specific application of the theory of Boolean functions, we present an algorithm for verification whether a given class of equivalent responses is a type of joint action and evaluate its time complexity. The formalism of Boolean functions also allows us to estimate the time complexity of the degree of joint action calculation.

In general, it is shown that Boolean algebra theory and graph theory is an effective means for formalizing and analysis of the binary sufficient causes theory. In particular, it allows one not only to obtain exact concepts for the fundamental constructions of this theory, but also to pose new problems and propose solutions to them.

We note, however, that at the current formalization stage, it is not possible to formulate concept of such a response function, which could be considered an analogue of the summing effects function known in epidemiology [9, 33–35]. With respect to this function, the observed dose-response dependences are compared, and on the basis of that comparison a conclusion is drawn about the manifestation of a sub- or superadditive joint action.

Nonetheless, the above classification of the types of joint action by the magnitude of μ_f makes it possible to assess the strength of the joint influence of factors and, based on the available expert information, suggest a reasonable conclusion about the appropriateness of applying one or another scenario of regulatory measures.

REFERENCES

1. M. L. Brusseau, I. L. Pepper, Ch. P. Gerba, *Environmental and Pollution Science*. 3rd ed. Academic Press, Elsevier, 2019.
2. F. R. Spellman, *The Science of Environmental Pollution*. CRC Press, 2017.
3. J. Peterson, *Environmental Pollution Caused by Factories*. https://sciencing.com/environmental-pollution-caused-by-factories-12071269.html.
4. *Taxonomy of Environmental Pollution*. https://www.reference.com/science/explore/pollution?qo=categoryTaxonomyLeft.
5. E. J. Calabrese, *Multiple Chemical Interactions (Toxicology and Environmental Health Series)*. CRC Press, 1991.
6. M. A. Subramanian, *Toxicology: Principles and Methods*. 2nd ed. MJP Publishers, 2015.
7. W. B. Pratt, P. Taylor, *Principles of Drug Action: The Basis of Pharmacology*. 3rd ed. Churchill Livingstone, 1990.
8. G. Pöch, *Combined Effects of Drugs and Toxic Agents. Modern Evaluation in Theory and Practice.* Vienna: Springer, 1993.
9. W. R. Greco, G. Bravo, J. C. Parsons, "The Search for Synergy: A Critical Review from a Response Surface Perspective". *Pharmacological Review*. Vol. 47 (2), pp. 331–385.
10. *Combined Actions and Interactions of Chemicals in Mixtures The Toxicological Effects of Exposure to Mixtures of Industrial and Environmental Chemicals*. Danish State Information Centre. Danish Veterinary and Food Administration. Federal Report. 2003.
11. *Assessment of Combined Exposures to Multiple Chemicals: Report of a WHO/IPCS International Workshop on Aggregate/Cumulative Risk Assessment*. WHO, 2009.
12. M. Doble, A. Kumar, *Biotreatment of Industrial Effluents*. Elsevier Inc., 2005.
13. *Edetate calcium disodium*. Cerner Multum, Inc. 2018. Available: https://www.drugs.com/mtm/edetate-calcium-disodium.html.
14. K. J. Rothman. "Causes". *Am. J. of Epidemiology*, 1976, vol. 104(6), pp. 587–592.
15. O. S. Miettinen. "Causal and preventive interdependence: Elementary principles". *Scand. J. Work. Environ. Health*, 1982, vol. 8, pp. 159–168.
16. O. S. Miettinen. *Theoretical Epidemiology*. John Wiley & Sons Inc., 1985.
17. S. Greenland, C. Poole. "Invariants and noninvariants in the concept of interdependent effects". *Scand. J. Work. Environ. Health*, 1988, vol. 14, pp. 125–129.
18. S. Greenland, B. Brumback. "An overview of relations among causal modelling methods". *Int. J. Epidemiol.*, 2002, vol. 31, pp. 1030–1037.
19. T. J. VanderWeele. "Sufficient cause interactions and statistical interactions". *Epidemiology*. 2009, vol. 20(1), pp. 6–13.
20. T. J. VanderWeele, T. S. Richardson. "General theory for interactions in sufficient cause models with dichotomous exposures". *Ann. Statistics*. 2012, vol. 40, pp. 2128–2161.
21. T. J. VanderWeele, J. M. Robins. "The identification of synergism in the sufficient-component-cause framework". *Epidemiology*. 2007, vol. 18(3), pp. 329–339.
22. V. G. Panov, J. V. Nagrebetskaya. "An algebraic treatment of two-factor binary sufficient causes theory". *Proceed. SPIIRAN*. 2013, vol. 26, pp. 277–296. (in Russian).
23. V. G. Panov, J. V. Nagrebetskaya. "Boolean algebras and classification of interactions in sufficient-component cause model". *Int. J. Pure Appl. Math.* 2015, vol. 98(2), pp. 239–259.
24. V. G. Panov, J. V. Nagrebetskaya. "Classification of combined action of binary factors and Coxeter groups". *J. Discrete Math. Sci. & Crypt.* 2018, vol. 21(3), pp. 661–677.
25. S. L. Morgan, Ch. Winship. *Counterfactuals and Causal Inference.* Cambridge University Press, 2007.
26. "Handbook of Boolean algebras". Vols 1–3. J. D. Monk, R. Bonnet, Eds. Amsterdam: Elsevier, 1989.
27. Y. Crama, P. L. Hammer. "Boolean functions. Theory, Algorithms, and Applications". Cambridge: Cambridge University Press, 2011.
28. R. Lidl, G. Pilz. "Applied abstract algebra". 2nd ed. New York: Springer, 1998.
29. N. Christofides. *Graph Theory. An Algorithmic Approach.* 3rd ed. London: Academic Press, 1978.
30. T. J. VanderWeele, J. Robins. (2006). A Theory of Sufficient Cause Interactions. *COBRA Preprint Series, Article 13*, Available: http://biostats.bepress.com/cobra/ps/art13.
31. T. J. VanderWeele, J. Robins. "Empirical and counterfactual conditions for sufficient cause interactions". *Biometrika*, 2008, vol. 95, pp. 49–61.
32. B. Steinbach, Ch. Posthoff. "Boolean Differential Calculus". Synthesis Lectures on Digital Circuits and Systems. Morgan&Claypool Publishers. 2017.
33. G. J. Howard, T. Webster. "Contrasting Theories of Interaction in Epidemiology and Toxicology". *Environ. Health Perspect.*, 2013, vol. 121, 1–6.
34. U.S. EPA (U.S. Environmental Protection Agency). *Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures*. Washington, DC: U.S. EPA, 2000. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=20533#Download.

35. M. C. Berenbaum. "What is synergy?" *Pharmacol. Rev.*, 1989, vol. 41(2), pp. 93–141.

AUTHORS PROFILE

Dr. Julia V. Nagrebetskaya is working as an Associate Professor in the Department of Mathematics, Mechanics and Computer Science, Institute of Natural Sciences and Mathematics, Ural Federal University, Ekaterinburg. She received a mathematical education in mathematical logic, theory of algorithms and computer science. The first works were devoted to model theory and theory of algorithms. Today, her research interests and current activities are abstract and applied algebra, mathematical logics, discrete mathematics, probability, statistics and their applications to epidemiology, biology and medicine.

Dr. Vladimir Panov is a leading researcher at the Institute of Industrial Ecology, Ekaterinburg. He was engaged in research in functional analysis, topology, algebra, probability theory, mathematical and applied statistics at different times. His main subjects of interest currently are mathematical modelling in toxicology, epidemiology, and medicine, formal analysis of multifactorial combined action, algebraic and statistical analysis and modelling of biological and epidemiological phenomena.