CAUSE-EFFECT RELATIONSHIPS AND PARTIALLY DEFINED BOOLEAN FUNCTIONS

Yves CRAMA [1], Peter L. HAMMER [2] and Toshihide IBARAKI [3]

[1] *Department of Quantitative Economics, University of Limburg, 6200 MD Maastricht, The Netherlands*
[2] *RUTCOR - Rutgers Center for Operations Research, Rutgers University, New Brunswick, NJ 08903, U.S.A.*
[3] *Department of Applied Mathematics and Physics, Kyoto University, Kyoto, Japan 606*

Abstract

This paper investigates the use of Boolean techniques in a systematic study of cause-effect relationships. The model uses partially defined Boolean functions. Procedures are provided to extrapolate from limited observations, concise and meaningful theories to explain the effect under study, and to prevent (or provoke) its occurrence.

1. Introduction

The main problem we plan to investigate in this paper is that of identifying the small subsets of plausible causes of a given effect, among a large set of factors including all the potential causes, along with many other (irrelevant) factors. Moreover, we shall present methods to describe all the combinations of possible causes which can be identified on the basis of a limited number of observations, as well as to build procedures for provoking or preventing the occurrence of a certain effect. It should be remarked from the very beginning that the proposed methods give a complete description of the cause-effect relationship only when the effect of *all* possible combinations of factors has been observed. In the usual case when only partial observations are available, no method can provide definite answers; however, even in this case, the cause-effect relationship can be narrowed down sufficiently to provide substantial guidance to a decision-maker.

Models related to the one developed here have been considered in the artificial intelligence literature, mostly by researchers interested in machine learning and in inductive inference (see e.g. chapter XIV in [1]). In the terminology of [1], our paper deals with data driven techniques (apparented to version space techniques) for disjunctive concept learning.

In this context, the main contribution of our work is its emphasis on the use of classical concepts from switching theory, and, in particular, on the concept of partially defined Boolean function. We consider this viewpoint as extremely fruitful, since it opens the possibility to take direct advantage, in an artificial

intelligence framework, of an enormous body of known results concerning Boolean functions (see e.g. [2], [4–8], [10], etc.).

Partially defined Boolean functions are usually refered to in the literature as Boolean functions with "don't care" conditions, reflecting the concern of specialists in the area of switching circuit design to define circuits with prescribed behavior in certain 0–1 points, and with arbitrary behavior in the remaining points. In the applications we have in mind, the value of the function is known in certain observed points, and unknown, but *relevant*, in all other points (as a matter of fact, such values can always be obtained at the cost of performing additional observations). Consequently, our object, as opposed to that of circuit designers, is as much to gain an overview of the family of all functions agreeing in the observed points with the partially defined one, as to select an 'optimal' function among that family.

Our work also displays connections with Valiant's learning model [9], but our focus is again different. Indeed, Valiant is mostly interested in circumscribing classes of "easily learnable" relationships, i.e. relationships (or Boolean functions) which can be induced, almost always and almost perfectly, by performing only a small (polynomial) number of observations and of computations.

By contrast, many of the computational problems arising in our framework are NP-hard, or provably require an exponential number of steps in the worst-case (since, for example, the number of "supporting sets", of "patterns", of "thrifty theories" associated with a problem may be exponentially large–see sections 3, 4, 6). Therefore, we expect our methods to be mostly applicable in such situations where the number of explanatory factors and of observations is reasonably small (which, we believe, is the case in most applications of interest).

Still, interesting connections remain between the results presented in [9] and in the present paper. Some of the learnable classes considered in [9], for instance, also turn out to provide the most tractable relationships from our viewpoint (positive and low order relationships; see sections 8.1, 8.3).

The model on which we are going to illustrate the general problem will consist in an attempt to identify the combinations of food items which give headaches to a certain individual. The nutrient-headache relationship will be used for illustratory purpose only, and should be taken with a grain of salt. Similar illustrations could have been built using other types of situations occurring in medical practice, scientific experimentation, or engineering.

2. Headaches and Boolean formalism

We shall start this section with a very simple example.

EXAMPLE 1
A physician would like to find out the combination of food items which cause a headache to one of his patients, and requests his patient to keep a record of his

Table 1

Day	Food item								Headache
	1	2	3	4	5	6	7	8	
1		X		X		X	X		Yes
2	X		X		X		X		No
3				X	X	X			No
4	X	X		X		X		X	No
5	X	X		X	X			X	Yes
6		X		X			X	X	No
7		X	X		X			X	Yes

diet. One week later, the patient returns to the doctor and brings in the record displayed in table 1.

After a brief examination, the doctor concludes that on the days when the patient had no headache, he never consumed food #2 without food #1, but he did so on some of the occasions when he had a headache. Similarly, our clever doctor concludes that the patients has never consumed food #4 without food #6 on the days when he had no headache; but he did so once, and he had a headache. He finally concludes that the two "patterns" noticed above explain every headache, and he puts forward the "theory" that this patient's headaches can always be explained by using these two patterns.

Obviously, the doctor had to answer three questions in this process.

(a) How to come up with a short list of food items sufficient to explain the presence or absence of headaches? In our example, foods #1, 2, 4, 6 were already enough for this purpose.

(b) How to detect patterns (i.e., combinations of food items) causing headaches? In our case, the doctor found two such patterns.

(c) How to build theories (i.e., collections of patterns) explaining every observed headache?

Fortunately, his patient was familiar with Boolean terminology (see e.g. Mendelson [6], Muroga [7]), and made the following suggestions. Let $x_1, x_2, \ldots, x_n$

Table 2

Day	Food item								$f(x_1, \ldots, x_8)$
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
1	0	1	0	1	0	1	1	0	1
5	1	1	0	1	1	0	0	1	1
7	0	1	1	0	1	0	0	1	1
2	1	0	1	0	1	0	1	0	0
3	0	0	0	1	1	1	0	0	0
4	1	1	0	1	0	1	0	1	0
6	0	0	1	0	1	0	1	0	0

be 0-1 variables associated to the food items, with the interpretation that $x_i = 1$ if and only if food #i has been consumed on a particular day. Let $f(x_1, x_2, \ldots, x_n)$ be a Boolean function taking the value 1 on the days with headaches, and 0 on the other ones. Due to the limitations in record-keeping, the values of f are known only for a small subset of 0-1 vectors. In our example, we only know the values of f in 7 out of the 2^8 possible vectors. With our new notations, table 1 becomes now table 2.

If x is a variable taking values 0 and 1, then $\bar{x} = 1 - x$ is called its *complement*. If $x_1, x_2, \ldots, x_n$ are Boolean variables, then the symbols $x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_n, \bar{x}_n$ are called *literals*. A product of the form

$$C = \prod_{i \in M} x_i \prod_{j \in N} \bar{x}_j \quad (M \cap N = \emptyset)$$

is an elementary conjunction (here and in the sequel, products over the empty set are taken equal to 1). A disjunction of elementary conjunctions is called a *disjunctive normal form* (or simply disjunctive form). Every disjunctive form defines a Boolean function; the same Boolean function can be defined by several disjunctive forms.

The patterns noticed by the doctor tell that, whenever $\bar{x}_1 x_2 = 1$ or $x_4 \bar{x}_6 = 1$, the patient had a headache, and that whenever $\bar{x}_1 x_2 = x_4 \bar{x}_6 = 0$, he did not. The doctor's theory is that the function $\psi = \bar{x}_1 x_2 \vee x_4 \bar{x}_6$ explains all the headaches. Consequently, the list of variables x_1, x_2, x_4, x_6 can be called a *supporting* list of variables.

Clearly, the doctor's theory is consistent with all the observed data. It is only natural to ask whether a mathematical approach could construct such a valid theory, or possibly *all* valid theories consistent with the known facts. With this goal in mind, we shall introduce a few further definitions.

Let T and F be two disjoint subsets of n-dimensional 0-1 vectors. The *partially defined Boolean function* on $T \cup F$ is given by:

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } (x_1, x_2, \ldots, x_n) \in T \\ 0 & \text{if } (x_1, x_2, \ldots, x_n) \in F. \end{cases}$$

Partially defined Boolean functions will provide the mathematical model capturing the doctor's observations. This concept will be used throughout the paper.

Every Boolean function (i.e., defined over all 2^n 0-1 vectors) agreeing with f on $T \cup F$ and taking arbitrary 0-1 values elsewhere will be called an *extension* of f. In particular, the completely defined Boolean functions f^+ and f^- defined by:

$$f^+(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } (x_1, x_2, \ldots, x_n) \notin F \\ 0 & \text{if } (x_1, x_2, \ldots, x_n) \in F \end{cases}$$

and

$$f^-(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } (x_1, x_2, \ldots, x_n) \in T \\ 0 & \text{if } (x_1, x_2, \ldots, x_n) \notin T \end{cases}$$

are extensions of f, and any extension f' of f is such that $f^- \leqslant f' \leqslant f^+$, meaning that $f^-(x) \leqslant f'(x) \leqslant f^+(x)$ for any $x = (x_1, x_2, \ldots, x_n)$. The unknown function describing the actual headache, or cause-effect relationship, is therefore one of the many extensions of the given partially defined function.

3. Supporting sets of variables

A set of variables, say $\{x_1, x_2, \ldots, x_k\}$, is called *supporting* for the partially defined function f if f has an extension depending only on the variables $x_1, x_2, \ldots, x_k$. Equivalently, $\{x_1, x_2, \ldots, x_k\}$ is supporting if, for every pair of vectors $X' = (x_1, x_2, \ldots, x_k, x'_{k+1}, \ldots, x'_n)$ and $X'' = (x_1, x_2, \ldots, x_k, x''_{k+1}, \ldots, x''_n)$ in $T \cup F$, $f(X') = f(X'')$. A supporting set is called *minimal* if it does not properly contain any other supporting set.

In order to identify the minimal supporting sets, we shall introduce a new set of 0-1 variables, $\{y_1, y_2, \ldots, y_n\}$. The vectors $(y_1, y_2, \ldots, y_n)$ will be interpreted as the characteristic vectors of subsets of variables, and we shall derive conditions under which such a vector corresponds to a supporting set of variables; minimality will be achieved by finding the optimal solutions of some associated set-covering problems.

It is obvious that if $(x'_1, x'_2, \ldots, x'_n) \in T$, $(x''_1, x''_2, \ldots, x''_n) \in F$, and $Y = (y_1, y_2, \ldots, y_n)$ corresponds to an arbitrary supporting set, then there exists an index i for which $x'_i \neq x''_i$ and $y_i = 1$ (since otherwise, for every j, $y_j = 1$ would imply $x'_j = x''_j$, in contradiction with the definition of a supporting set). Therefore, for every pair of vectors $X' \in T$ and $X'' \in F$, denoting by $\Delta(X', X'')$ the set of indices where X' and X'' differ, the vector Y will satisfy:

$$\sum_{i \in \Delta(X', X'')} y_i \geqslant 1. \tag{1}$$

It can be seen immediately that Y is a supporting vector if and only if it satisfies (1) for every pair of vectors $X' \in T$ and $X'' \in F$.

In conclusion, the minimal supporting sets are in one-to-one correspondence with the minimal 0-1 vectors satisfying system (1). This observation can already be found in Necula [8].

By appropriately choosing the coefficients c_i in the *set-covering problem* (SCP):

$$\text{minimize } \sum_{i=1}^{n} c_i y_i$$

$$\text{st. } \sum_{i \in \Delta(X', X'')} y_i \geqslant 1 \quad \text{for } X' \in T, \ X'' \in F$$

$$y_i \in \{0, 1\} \quad \text{for } i = 1, 2, \ldots, n,$$

one can identify particular types of minimal supporting sets. For example, the minimal supporting sets involving the smallest number of variables (which have

an obvious appeal for practitioners) can be found by taking all the c_i's equal to 1. Similarly, any a priori knowledge (e.g. of a statistical nature) about the relative importance of the variables can be built into the selection of a minimal supporting set by weighting the variables accordingly in (SCP).

EXAMPLE 1 (continued)

For the example of section 2, the generated constraints would be the following:

$$y_1 + y_2 + y_3 + y_4 + y_5 + y_6 \geqslant 1 \tag{2}$$

$$y_2 \qquad\qquad + y_5 \qquad + y_7 \qquad \geqslant 1 \tag{3}$$

$$y_1 \qquad\qquad\qquad\qquad + y_7 + y_8 \geqslant 1 \tag{4}$$

$$y_2 + y_3 + y_4 + y_5 + y_6 \qquad\qquad \geqslant 1 \tag{5}$$

$$y_2 + y_3 + y_4 \qquad\qquad + y_7 + y_8 \geqslant 1 \tag{6}$$

$$y_1 + y_2 \qquad\qquad + y_6 \qquad + y_8 \geqslant 1 \tag{7}$$

$$y_5 + y_6 \qquad\qquad\qquad \geqslant 1 \tag{8}$$

$$y_1 + y_2 + y_3 + y_4 \qquad\qquad + y_7 + y_8 \geqslant 1 \tag{9}$$

$$y_1 + y_2 \qquad\qquad\qquad + y_7 + y_8 \geqslant 1 \tag{10}$$

$$y_2 + y_3 + y_4 \qquad + y_6 \qquad + y_8 \geqslant 1 \tag{11}$$

$$y_1 \qquad + y_3 + y_4 + y_5 + y_6 \qquad\qquad \geqslant 1 \tag{12}$$

$$y_2 \qquad\qquad\qquad + y_7 + y_8 \geqslant 1 \tag{13}$$

which, after the elimination of the redundant constraints (2), (5), (6), (9), (10), (12), reduces to the system (3), (4), (7), (8), (11), (13). All the minimal solutions of this system in 0-1 variables are:

$$Y_1 = (0, 0, 0, 0, 1, 0, 0, 1) \qquad Y_5 = (0, 1, 0, 0, 1, 0, 1, 0)$$
$$Y_2 = (0, 0, 0, 0, 0, 1, 1, 0) \qquad Y_6 = (0, 1, 0, 0, 0, 1, 0, 1)$$
$$Y_3 = (1, 1, 0, 0, 1, 0, 0, 0) \qquad Y_7 = (1, 0, 1, 0, 1, 0, 1, 0)$$
$$Y_4 = (1, 1, 0, 0, 0, 1, 0, 0) \qquad Y_8 = (1, 0, 0, 1, 1, 0, 1, 0)$$

which define the following minimal supporting sets:

$$S_1 = \{x_5, x_8\} \qquad\qquad S_5 = \{x_2, x_5, x_7\}$$
$$S_2 = \{x_6, x_7\} \qquad\qquad S_6 = \{x_2, x_6, x_8\}$$
$$S_3 = \{x_1, x_2, x_5\} \qquad S_7 = \{x_1, x_3, x_5, x_7\}$$
$$S_4 = \{x_1, x_2, x_6\} \qquad S_8 = \{x_1, x_4, x_5, x_7\}.$$

It can be easily seen that the variables in any of the sets above completely characterize the behavior of our function in all of the observed points. To take S_1 as an example, we notice that the function $\psi = x_5 x_8 \vee \bar{x}_5 \bar{x}_8$ is an extension of f.

4. Patterns and their detection

A combination of present or absent factors appearing only in diets associated with headaches can be suspected to be a cause of the headache. For example, in every occasion that food #2 was eaten without food #5 and food #8, a headache resulted. Therefore, the combination "#2 present, #5 absent, #8 absent" could be suspected to be conducive to a headache. However, in this explanation, food #2 plays no role, since the simple absence of foods #5 and 8 has the same property.

A minimal combination of the type described will be called a *pattern*. It is easy to notice that "the absence of foods #5 and #8" is indeed a pattern.

Mathematically speaking, a pattern is an elementary conjunction C of some of the literals $x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_n, \bar{x}_n$, so that

(i) $C(X) = 0$ for every $X \in F$;

(ii) $C(X) = 1$ for at least one vector X in T;

(iii) for every conjunction C' obtained by dropping a literal from C, there exists a vector $X \in F$ such that $C'(X) = 1$.

(Observe that, when $T = \emptyset$, there is no pattern, and when $F = \emptyset$, the constant 1 is a pattern; we shall disregard these extreme cases in what follows). For instance, the pattern identified in the above example is $C = \bar{x}_5 \bar{x}_8$.

Let us recall here that an elementary conjunction C is an *implicant* of the Boolean function $\psi(X)$ if:

$$C(X) = 1 \Rightarrow \psi(X) = 1$$

for all Boolean vectors X. An implicant C is *prime* if no implicant of ψ can be obtained by deleting literals from C.

It is now easy to see that:

THEOREM 1

Every pattern of a partially defined Boolean function f is a prime implicant of its extension f^+. □

Detecting all patterns can be achieved in various ways by applying simple, albeit time-consuming, techniques. We propose below two such procedures.

The first one is close in spirit to the procedure described in section 3 for obtaining minimal supporting sets of variables. We consider successively each vector $X \in T$. For each such vector, the list $Y_1, Y_2, \ldots, Y_k$ of minimal solutions of the system:

$$\sum_{i \in \Delta(X, X')} y_i \geq 1 \quad \text{for } X' \in F$$

is obtained. Let us associate now with each vector Y_j the elementary conjunction

$$C_j = \prod_{u \in U^j} x_u \prod_{v \in V^j} \bar{x}_v$$

where U^j (resp. V^j) is the set of indices i for which the i-th component of X is one (resp. zero) while the i-th component of Y_j is one.

One easily checks that the conjunctions obtained in that way, for all possible vectors $X \in T$, are exactly the patterns of f.

EXAMPLE 1 (continued)

For the example in table 2, when X is the vector associated with day 1, we get the system of inequalities (2), (3), (4), (5), the minimal solutions of which are:

$Y_1^* = (1, 1, 0, 0, 0, 0, 0, 0)$ $Y_6^* = (0, 0, 0, 1, 0, 0, 1, 0)$

$Y_2^* = (1, 0, 0, 0, 1, 0, 0, 0)$ $Y_7^* = (0, 0, 0, 0, 1, 0, 1, 0)$

$Y_3^* = (0, 1, 0, 0, 0, 0, 1, 0)$ $Y_8^* = (0, 0, 0, 0, 1, 0, 0, 1)$

$Y_4^* = (0, 1, 0, 0, 0, 0, 0, 1)$ $Y_9^* = (0, 0, 0, 0, 0, 1, 1, 0)$.

$Y_5^* = (0, 0, 1, 0, 0, 0, 1, 0)$

These determine the patterns:

$C_1 = \bar{x}_1 x_2$ $C_6 = x_4 x_7$

$C_2 = \bar{x}_1 \bar{x}_5$ $C_7 = \bar{x}_5 x_7$

$C_3 = x_2 x_7$ $C_8 = \bar{x}_5 \bar{x}_8$

$C_4 = x_2 \bar{x}_8$ $C_9 = x_6 x_7$.

$C_5 = \bar{x}_3 x_7$

The vectors associated with day 5 and with day 7, respectively, would generate more patterns in a similar way.

The procedure described above generates all the possible patterns associated to a problem. It will be seen in the sequel that only a small subset of patterns will be used at any given moment, and it will be also clear that the most important patterns are those involving few literals.

A second procedure can therefore be devised, which simply enumerates all the "short" elementary conjunctions and checks whether they are patterns. Various clever implementations can be designed to keep the amount of computations as low as possible. It is to be stressed that, when only patterns up to a given length (e.g. 2, or 3) are needed, then all these implementations are polynomial.

5. Theories

Due to the unavailability of complete data, the known facts can be given a variety of explanations, depending to a large degree on the interpreter's subjective viewpoint, but also respecting several generally accepted principles.

Obviously, a theory has to satisfy two main criteria:

(i) it has to be consistent with all the observations;

(ii) it has to be able to predict the outcome of any (not yet necessarily encountered) combination of factors.

From a formal point of view, a *theory* is therefore a Boolean function which agrees with the given partially defined function in every observed point. In other

words, a theory is simply an extension of the partially defined function, as introduced in section 2.

Since the number of extensions/theories can be very large, it is important to devise additional principles to help us identify particularly appealing ones. The identification can be based on commonly accepted criteria of simplicity (e.g. limitation in the number of explanatory factors, use of easily understandable causal relationships, etc.), or on observed trends in the effects of different agents (e.g. the knowledge that causal effects can be always reduced to the interplay of consistently inducive or consistently inhibitive factors, or that all factors can be meaningfully ranked as to the importance of their impact, etc.).

We shall examine in details, in the following sections, theories devised on the basis of such criteria. From a technical point of view, these sections will rely on the concepts and methods introduced in the previous sections, particularly on the detection of patterns and supporting sets of variables.

6. Thrifty theories

Every cause-effect relationship can be described by listing various combinations of present or absent factors, such that each of these combinations produces the effect in question, and the effect is produced only by the combinations in the list. Once a list is available, various transformations can be performed on it, aimed at producing "simpler" lists, which still describe valid theories. More precisely, some combinations on a list may be redundant, in the sense that the list obtained by omitting these combinations is still a theory. Other combinations may contain irrelevant factors, the elimination of which still maintains a possible theory.

A theory will be called *thrifty* if it does not involve redundancies of either of the above two types.

EXAMPLE 2

Consider a problem with the following observed situations:

Days	Food items				Headache
	1	2	3	4	
1	1	0	0	1	1
2	0	1	1	0	1
3	0	0	0	1	0

A possible theory explaining all headaches is described by the following list of headache-causing combinations:

Combinations	Food items			
	1	2	3	4
A	1			1
B	0	1		
C		1	1	

(meaning that food items #1 and #4 taken simultaneously cause a headache, food #2 without food #1 causes a headache, foods #2 and #3 together cause a headache, and the present theory assumes that these are the only headache-causing combinations).

It can be noticed that, by dropping C from the list, we still have a possible theory. Also, it can be seen that eliminating food #4 from combination A produces a simpler theory. Finally, disregarding the role of food #1 in combination B simplifies the theory to the following *thrifty* one:

Combinations	Food items			
	1	2	3	4
A'	1			
B'		1		

To rephrase more formally the intuitive definition given before the example, we can say that a thrifty theory is an extension ψ of the given partially defined function f, which has a disjunctive normal form

$$\psi = C_1 \vee C_2 \vee \cdots \vee C_k \tag{14}$$

with:

$$C_i = \prod_{j \in J_i} x_j \prod_{j \in J_i'} \bar{x}_j \quad (J_i \cap J_i' = \emptyset; \ i = 1, \dots, k) \tag{15}$$

such that the expressions obtained from ψ either by dropping one of the elementary conjunctions C_i, or by dropping a literal from one of these conjunctions, do not define extensions of f.

In order to relate this concept to the previously introduced ones, let us now call the expression

$$\Pi = P_1 \vee P_2 \vee \cdots \vee P_q \tag{16}$$

thrifty if $P_1, P_2, \dots, P_q$ are patterns, Π is an extension of f, and the disjunctive normal forms obtained by dropping one of the P_i's from Π are not extensions of f.

THEOREM 2

A theory is thrifty if and only if it has a thrifty expression. □

THEOREM 3

If $\Pi = P_1 \vee P_2 \vee \cdots \vee P_q$ is a thrifty expression, then $P_1, P_2, \dots, P_q$ are prime implicants of the function ψ defined by Π; moreover, a prime implicant P of ψ is a pattern if and only if there exists $X \in T$ such that $P(X) = 1$. □

The proofs are straightforward and therefore omitted.

EXAMPLE 3

Consider the partially defined Boolean function described by the following table:

Vectors	Variables				f
	x_1	x_2	x_3	x_4	
1	1	0	1	0	1
2	0	1	1	0	1
3	1	0	0	1	1
4	0	0	0	0	0
5	1	1	1	1	0

A thrifty extension for f is defined by the expression $\Pi = \bar{x}_1 x_2 \vee \bar{x}_2 x_3 \vee \bar{x}_2 x_4$. This extension admits the prime implicants $\bar{x}_1 x_2$, $\bar{x}_2 x_3$, $\bar{x}_2 x_4$, $\bar{x}_1 x_3$, $\bar{x}_1 x_4$, the last one of which is not a pattern of f.

An important question to be settled now is the selection of a thrifty theory, out of the large number of conceivable ones. The techniques to be described below will identify collections of patterns defining thrifty theories. We shall outline a (computationally demanding) method for generating a "shortest" thrifty expression, and a (more tractable) procedure for building at least one thrifty expression.

In order to simplify our description, we shall say that a pattern P *covers* a vector X if $P(X) = 1$. A collection of patterns $\{P_1, \ldots, P_k\}$ will be called a *minimal cover* of the set T if every vector $X \in T$ is covered by a pattern in the collection, and the collection is minimal with this property. It is easy to notice that $\{P_1, \ldots, P_k\}$ is a minimal cover if and only if $P_1 \vee \cdots \vee P_k$ is a thrifty expression. This observation shows that thrifty expressions are in one-to-one correspondence with the minimal 0-1 solutions of the system of inequalities:

$$\sum_{i \in Q(X)} z_i \geq 1 \quad \text{for all } X \in T, \tag{17}$$

where $\{P_1, \ldots, P_t\}$ is the set of all patterns, $Q(X) \subseteq \{1, \ldots, t\}$ is the set of indices of the patterns covering X, and $z_i = 1$ (0) means that pattern P_i is (not) selected in an expression. The thrifty expression involving the smallest number of patterns can therefore be selected by minimizing the objective function $\sum_{i=1}^{t} z_i$ subject to the constraints (17).

EXAMPLE 3 (continued)

Consider again the partially defined function of example 3. Its patterns are $P_1 = x_1 \bar{x}_2$, $P_2 = \bar{x}_1 x_2$, $P_3 = x_1 \bar{x}_3$, $P_4 = \bar{x}_1 x_3$, $P_5 = x_1 \bar{x}_4$, $P_6 = \bar{x}_2 x_3$, $P_7 = x_2 \bar{x}_4$, $P_8 = \bar{x}_2 x_4$, $P_9 = x_3 \bar{x}_4$, $P_{10} = \bar{x}_3 x_4$. The corresponding set covering problem is: minimize $z_1 + z_2 + \cdots + z_{10}$

s.t. $z_1 + z_5 + z_6 + z_9 \geq 1$

$z_2 + z_4 + z_7 + z_9 \geq 1$

$z_1 + z_3 + z_8 + z_{10} \geq 1$

$z_1, \ldots, z_{10} \in \{0, 1\}$.

One of its optimal solutions is given by $z_1 = z_2 = 1$, $z_3 = \cdots = z_{10} = 0$. Therefore, a thrifty theory involving only two patterns is defined by:

$$\Pi = x_1 \bar{x}_2 \vee \bar{x}_1 x_2. \tag{18}$$

A second method of building a thrifty theory, without first producing all patterns, will initially generate a pattern, and will then label all points of T covered by this pattern. At this stage, a new pattern will then be generated, covering at least one of the unlabelled points in T, and all points covered by this new pattern will be also labelled. The process will be repeated until all points of T are labelled. Since the collection of patterns produced may be redundant, patterns will then be sequentially checked for redundancy and possibly eliminated, until a minimal cover is obtained.

EXAMPLE 3 (continued)

We first produce pattern $P_6 = \bar{x}_2 x_3$ covering vector X_1, and we label x_1. Then pattern $P_2 = \bar{x}_1 x_2$ is produced, and vector X_2 is labelled. Next, $P_1 = x_1 \bar{x}_2$ is produced, and the last vector X_3 is labelled. Since the collection $\{ P_1, P_2, P_6 \}$ is not a minimal cover of T, we eliminate P_6. This results in the same minimal cover $\{ P_1, P_2 \}$, and the thrifty expression (18), found by the first procedure (although this second procedure is not guaranteed in general to yield a thrifty expression involving the smallest number of patterns).

7. Basic theories

Minimal supporting sets of variables were defined in section 3. An essential criterion in the description of cause-effect relationships is the number of factors on which they depend. This consideration leads naturally to building theories on minimal supporting sets. Such a theory will be called *basic*.

Among basic theories, a very special role will be played by basic thrifty ones. These can be obtained by first selecting a minimal supporting set of variables, say $S = \{ x_1, x_2, \ldots, x_k \}$, out of the original set of variables $\{ x_1, x_2, \ldots, x_n \}$. We shall define the *contraction* of the partially defined function $f(x_1, x_2, \ldots, x_n)$ to the set S to be the partially defined function $f_s(x_1, \ldots, x_k)$, where:

$$f_s(x_1, \ldots, x_k) = f(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \quad \text{for all } (x_1, \ldots, x_n) \text{ in } T \cup F.$$

This definition is meaningful, because of the definition of supporting sets.

Thrifty theories associated with f_s can now be built as explained in the previous section. They will use as building blocks the patterns of f_s.

THEOREM 4

Let S be a minimal supporting set of variables. Then every pattern of f_s is a pattern of f, and every thrifty theory of f_s is a basic thrifty theory of f. $\square$

This theorem allows the construction of theories for the original problem by simply working on a contraction of it, thus substantially reducing the computational effort, while maintaining the thrifty nature of the result.

Table 3

x_1	x_2	x_5	f_{s_3}
0	1	0	1
1	1	1	1
0	1	1	1
1	0	1	0
0	0	1	0
1	1	0	0

EXAMPLE 1 (continued)

We have obtained in section 3 all the minimal supporting sets of the function defined by table 2. Let us build for example the thrifty theories corresponding to $S_3 = \{x_1, x_2, x_5\}$. The associated contraction f_{s_3} of f is given by table 3. Its patterns can easily be seen to be $P_1 = \bar{x}_1 x_2$, $P_2 = \bar{x}_1 \bar{x}_5$, $P_3 = x_2 x_5$. The corresponding set covering problem, describing thrifty theories as in section 6, is:

minimize $z_1 + z_2 + z_3$

s.t. $z_1 + z_2 \quad\quad \geqslant 1$

$\quad\quad\quad\quad z_3 \geqslant 1$

$\quad z_1 \quad\quad + z_3 \geqslant 1$

$z_1, z_2, z_3 \in \{0, 1\}$.

This problem has two optimal solutions: (1, 0, 1) and (0, 1, 1), defining the thrifty theories:

$$\Pi_3^1 = \bar{x}_1 x_2 \vee x_2 x_5 \quad \text{and} \quad \Pi_3^2 = \bar{x}_1 \bar{x}_5 \vee x_2 x_5.$$

For the sake of illustration, we list below *all* the basic thrifty theories for example 1. In the list, Π_j^i stands for the basic thrifty theory i on the minimal supporting set S_j:

Table 4

$\Pi_1 = x_5 x_8 \vee \bar{x}_5 \bar{x}_8$	$\Pi_7^1 = \bar{x}_1 \bar{x}_5 \vee x_3 \bar{x}_7 \vee x_1 x_5 \bar{x}_7$	$\Pi_8^1 = \bar{x}_1 \bar{x}_5 \vee \bar{x}_4 \bar{x}_7 \vee x_1 x_4 x_5$
$\Pi_2 = x_6 x_7 \vee \bar{x}_6 \bar{x}_7$	$\Pi_7^2 = \bar{x}_1 \bar{x}_5 \vee x_3 \bar{x}_7 \vee x_1 \bar{x}_3 x_5$	$\Pi_8^2 = \bar{x}_1 \bar{x}_5 \vee \bar{x}_4 \bar{x}_7 \vee x_1 x_5 \bar{x}_7$
$\Pi_3^1 = \bar{x}_1 x_2 \vee x_2 x_5$	$\Pi_7^3 = x_3 \bar{x}_7 \vee \bar{x}_3 x_7 \vee x_1 x_5 \bar{x}_7$	$\Pi_8^3 = x_4 x_7 \vee \bar{x}_4 \bar{x}_7 \vee x_1 x_4 x_5$
$\Pi_3^2 = \bar{x}_1 \bar{x}_5 \vee x_2 x_5$	$\Pi_7^4 = x_3 \bar{x}_7 \vee \bar{x}_3 x_7 \vee x_1 \bar{x}_3 x_5$	$\Pi_8^4 = x_4 x_7 \vee \bar{x}_4 \bar{x}_7 \vee x_1 x_5 \bar{x}_7$
$\Pi_4 = \bar{x}_1 x_2 \vee x_2 \bar{x}_6$	$\Pi_7^5 = x_3 \bar{x}_7 \vee \bar{x}_5 x_7 \vee x_1 x_5 \bar{x}_7$	$\Pi_8^5 = \bar{x}_4 \bar{x}_7 \vee \bar{x}_5 x_7 \vee x_1 x_4 x_5$
$\Pi_5^1 = x_2 x_5 \vee x_2 x_7$	$\Pi_7^6 = x_3 \bar{x}_7 \vee \bar{x}_5 x_7 \vee x_1 \bar{x}_3 x_5$	$\Pi_8^6 = \bar{x}_4 \bar{x}_7 \vee \bar{x}_5 x_7 \vee x_1 x_5 \bar{x}_7$
$\Pi_5^2 = x_2 x_5 \vee \bar{x}_5 x_7$		
$\Pi_6^1 = x_2 \bar{x}_6 \vee x_2 \bar{x}_8$		
$\Pi_6^2 = x_2 \bar{x}_8 \vee \bar{x}_6 x_8$		

The fact that the list of all conceivable theories (2^{249} of them in this example) has been reduced to a manageable size (21 here) will clearly help a practitioner in selecting a meaningful theory. Obviously, when the size of the problem increases, the corresponding number of thrifty theories may increase exponentially, making the production of a complete list impossible. In the general case, additional limitations imposed by the nature of the problem will help reduce the number of theories under consideration. Several natural classes of theories will be examined in the next section.

8. Special theories

8.1. POSITIVE THEORIES

A natural assumption underlying much of the Boolean literature concerns the monotonically increasing character of the functions examined. Functions with the property that

$$X \leqslant Y \Rightarrow F(X) \leqslant F(Y)$$

where $X \leqslant Y$ means $x_i \leqslant y_i$ for $i = 1, \ldots, n$, are called *positive*.

Positive functions have an intuitively clear meaning in a variety of contexts. In the example of the headache, the interpretation will simply be that every food item has either no effect at all on the headache, or it may be a contributing factor (in certain combinations) to the headache. In other words, no food item is supposed to ever have an inhibitive effect on headache.

Let us call an elementary conjunction *positive* if it does not contain any complemented variable. It is well known that a Boolean function is positive if and only if all its prime implicants are positive. In a similar way, one easily checks that:

THEOREM 5

A partially defined Boolean function f admits a positive extension if and only if the set of positive patterns of f covers all vectors in T. □

The set of positive patterns (if any) covering a fixed vector $X \in T$ can be described easily, by a simple modification of the first procedure given in section 4 for finding all patterns. The only difference will be in the definition of the sets $\Delta(X, X')$, which will be replaced here by sets $\Gamma(X, X')$ given by:

$$\Gamma(X, X') = \{i \mid x_i = 1, x_i' = 0\}.$$

Accordingly, the positive patterns covering X will be in one-to-one correspondence with the minimal solutions of the following system of inequalities in 0-1 variables:

$$\sum_{i \in \Gamma(X, X')} y_i \geqslant 1 \quad \text{for all } X' \in F.$$

Observe that, for a fixed X, the above system is feasible if and only if $\Gamma(X, X')$ is nonempty for every $X' \in F$. Combining this remark with theorem 5, we get:

COROLLARY (Zuev [10])

A partially defined Boolean function f admits a positive extension if and only if $\Gamma(X, X')$ is non-empty for all $X \in T$, $X' \in F$. $\square$

Similar extensions of the techniques developed in the previous sections also allow us to handle minimal positive supporting sets of variables, positive thrifty extensions, etc. It has to be noticed that the existence of a positive extension guarantees the existence of a positive thrifty one, but does not guarantee the existence of a positive basic one.

EXAMPLE 4

The partially defined function given by $T = \{(1, 0, 1), (0, 1, 1)\}$ and $F = \{(1, 1, 0), (0, 0, 1)\}$ has the unique positive thrifty extension: $\Pi = x_1 x_3 \vee x_2 x_3$; it has no positive basic extension, since its unique minimal supporting set is $\{x_1, x_2\}$.

EXAMPLE 1 (continued)

To illustrate the different aspects of treating positive extensions, we return now to example 1. The minimal positive supporting sets are in correspondence with the minimal solutions of the set covering problem (compare with (2)–(13)):

$$
\begin{aligned}
y_2 &+ y_4 && + y_6 && && \geqslant 1 \\
y_2 && && && + y_7 && \geqslant 1 \\
&& && && y_7 && \geqslant 1 \\
y_2 &+ y_4 && + y_6 && && \geqslant 1 \\
y_2 &+ y_4 && && && + y_8 && \geqslant 1 \\
y_1 + y_2 && && && && + y_8 && \geqslant 1 \\
&& && y_5 && && \geqslant 1 \\
y_1 + y_2 && + y_4 && && + y_8 && \geqslant 1 \\
y_2 && && && && + y_8 && \geqslant 1 \\
y_2 + y_3 && && && + y_8 && \geqslant 1 \\
&& y_3 && + y_5 && && \geqslant 1 \\
y_2 && && && + y_8 && \geqslant 1 \\
\end{aligned}
$$

$$y_i \in \{0, 1\} \ (i = 1, \ldots, 8)$$

which, after the elimination of redundant constraints shrinks to:

$$y_2 + y_4 \quad + y_6 \qquad \geqslant 1$$

$$y_7 \qquad \geqslant 1$$

$$y_5 \qquad \geqslant 1$$

$$y_2 \qquad + y_8 \geqslant 1$$

$$y_i \in \{0, 1\} \quad (i = 1, \ldots, 8).$$

The three minimal solutions of this system define the minimal positive supporting sets $S_1^* = \{x_2, x_5, x_7\}$, $S_2^* = \{x_4, x_5, x_7, x_8\}$ and $S_3^* = \{x_5, x_6, x_7, x_8\}$. Since one of these sets (S_1^*) is also a minimal supporting set (see S_5 in the list given in section 3), it follows that our problem has the positive basic thrifty extension:

$$\Pi_5^1 = x_2 x_5 \vee x_2 x_7,$$

which had already been found in table 4.

In order to illustrate the procedure of finding a positive thrifty expansion on a minimal positive supporting set, we shall consider the supporting set S_2^*. The contraction f^* of f to S_2^* is:

x_4	x_5	x_7	x_8	f^*
1	0	1	0	1
1	1	0	1	1
0	1	0	1	1
0	1	1	0	0
1	1	0	0	0
1	0	0	1	0

The positive patterns of f^* covering the first vector in that table correspond to the minimal solutions of the problem:

$$y_4 \geqslant 1$$

$$y_7 \geqslant 1$$

$$y_4, y_5, y_7, y_8 \in \{0, 1\}.$$

The only such pattern is $P_1^* = x_4 x_7$. In a similar way, the only positive pattern covering both the second and the third vectors in T is $P_2^* = x_5 x_8$. Therefore, the unique positive thrifty theory on S_2^* is $\Pi_2^* = x_4 x_7 \vee x_5 x_8$.

Also, the (unique) positive thrifty theory on S_3^* can be checked to be $\Pi_3^* = x_5 x_8 \vee x_6 x_7$.

8.2. UNATE THEORIES

The expression $\psi = x\bar{y}$ can be transformed into the positive expression $\psi' = xy'$ by introducing a new variable $y' = \bar{y}$; no similar transformation can be found to

make the function $xy \vee \bar{y}z$ positive. A function ψ is called *unate* if, for some subset V of variables, the change of variables:

$$x' = \begin{cases} \bar{x} & \text{if } x \in V \\ x & \text{if } x \notin V \end{cases}$$

transforms ψ into a positive function.

The meaning of unateness in terms of the headache problem is that each item has a consistent effect; some items may have an inducive effect, meaning that their addition to a headache-causing diet can never prevent the headache; some items may have an inhibitive effect, so that their addition to a diet which does not provoke headaches cannot result in a headache-causing diet; some items (perhaps the vast majority) may have no effect at all. The important fact about unateness is that each food item clearly belongs to one of the three classes.

If our object is to find a unate theory, the procedure is straightforward. Given the set V, we start out by changing the variables as in the definition of unate functions, and then we simply attempt to build a positive theory for the resulting partially defined function, as explained in section 8.1.

EXAMPLE 1 (continued)

If we attempt to build theories on the minimal supporting subset $S_3 = \{x_1, x_2, x_5\}$, under the hypothesis that food #1 is inhibitive of headache, while foods #2 and #5 are conducive of headache, then we change variable x_1 to $x_1' = \bar{x}_1$, and we consider the contraction f' of the resulting function to $\{x_1', x_2, x_5\}$:

x_1'	x_2	x_5	f'
1	1	0	1
0	1	1	1
1	1	1	1
0	0	1	0
1	0	1	0
0	1	0	0

The unique positive thrifty extension of f' is:

$$\psi' = x_1' x_2 \vee x_2 x_5$$

(corresponding to the unate thrifty extension of f, $\Pi_3^1 = \bar{x}_1 x_2 \vee x_2 x_5$, listed in table 4).

From a computational point of view, the study of unateness does not appear to be more difficult than that of positivity. This is however only true when the nature (conducive or inhibitive) of every variable is known a priori. The more general question of deciding whether there exists a unate theory based on an unspecified dichotomy of variables can be approached as described below.

We first present a slight generalization of the discussions of supporting sets and positive supporting sets given in sections 3 and 8.1. To that effect, we

introduce two vectors $Y = (y_1, y_2, \ldots, y_n)$ and $Z = (z_1, z_2, \ldots, z_n)$ of 0-1 variables. Our immediate goal is to characterize pairs of vectors Y, Z with the following property: there is an extension of the given partially defined function and a disjunctive normal form of that extension, in which x_i appears uncomplemented if and only if $y_i = 1$, and x_i appears complemented if and only if $z_i = 1$.

For every pair of vectors $X \in T$, $X' \in F$, let us define:

$$\Gamma_1(X, X') = \{i \mid x_i = 1, x_i' = 0\}$$
$$\Gamma_0(X, X') = \{i \mid x_i = 0, x_i' = 1\}.$$

THEOREM 6

The vectors Y, Z have the property described in the previous paragraph if and only if they are a solution of the system:

$$\sum_{i \in \Gamma_1(X,X')} y_i + \sum_{i \in \Gamma_0(X,X')} z_i \geq 1 \quad \text{for all pairs } X \in T \text{ and } X' \in F \tag{19}$$

$$y_i \in \{0, 1\}, \ z_i \in \{0, 1\}, \ i = 1, 2, \ldots, n. \tag{20}$$

Proof

Consider an extension ψ of f and one of its disjunctive forms:

$$\psi = C_1 \vee C_2 \vee \cdots \vee C_m,$$

where $C_j (j = 1, 2, \ldots, m)$ are elementary conjunctions of some literals chosen from $x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_n, \bar{x}_n$. Every C_j satisfies $C_j(X') = 0$ for all $X' \in F$, and for each $X \in T$, there is at least one C_j with $C_j(X) = 1$. Let $y_i = 1$ if x_i occurs uncomplemented in one of the C_j's, and $y_i = 0$ otherwise; let $z_i = 1$ if x_i occurs complemented in one of the C_j's, and $z_i = 0$ otherwise. It is straightforward that the vectors Y, Z satisfy (19)–(20).

Conversely, consider any vectors Y and Z satisfying (19)–(20). By definition, for each $X^* \in T$, Y and Z satisfy

$$\sum_{i \in \Gamma_1(X^*,X')} y_i + \sum_{i \in \Gamma_0(X^*,X')} z_i \geq 1 \quad \text{for all } X' \in F.$$

This implies that the conjunction C_{X^*} defined by

$$C_{X^*} = \prod_{\substack{i \mid y_i = 1 \\ x_i^* = 1}} x_i \prod_{\substack{i \mid z_i = 1 \\ x_i^* = 0}} \bar{x}_i$$

satisfies $C_{X^*}(X^*) = 1$ and $C_{X^*}(X') = 0$ for all $X' \in F$. Therefore, the extension ψ with the following disjunctive form

$$\psi = \bigvee_{X \in T} C_X$$

has the desired property. $\quad \square$

Notice now that a function is unate if and only if it has a disjunctive normal form in which every variable appears either complemented, or uncomplemented, but not both. This leads to the following:

THEOREM 7

A partially defined Boolean function admits a unate extension if and only if the associated set of inequalities (19)–(20) has a solution (Y, Z) such that

$$y_i + z_i \leqslant 1 \quad \text{for } i = 1, 2, \ldots, n. \quad \square \tag{21}$$

It is also easy to notice that the existence of a unate extension guarantees the existence of a unate thrifty one. However, similarly to the positive case, it does not guarantee the existence of a unate basic one.

EXAMPLE 1 (continued)

Corresponding to the inequalities (2)–(13) derived for the partially defined Boolean fucntion in table 2, we now have:

$$z_1 + y_2 + z_3 + y_4 + z_5 + y_6 \qquad\qquad \geqslant 1 \tag{22}$$

$$\quad\ y_2 \qquad\qquad + z_5 \qquad + y_7 \quad \geqslant 1 \tag{23}$$

$$z_1 \qquad\qquad\qquad\qquad + y_7 + z_8 \geqslant 1 \tag{24}$$

$$\quad\ y_2 + z_3 + y_4 + z_5 + y_6 \qquad\qquad \geqslant 1 \tag{25}$$

$$\quad\ y_2 + z_3 + y_4 \qquad\quad + z_7 + y_8 \geqslant 1 \tag{26}$$

$$y_1 + y_2 \qquad\qquad\quad + z_6 \qquad + y_8 \geqslant 1 \tag{27}$$

$$\qquad\qquad\qquad\quad\ y_5 + z_6 \qquad\qquad \geqslant 1 \tag{28}$$

$$y_1 + y_2 + z_3 + y_4 \qquad\quad + z_7 + y_8 \geqslant 1 \tag{29}$$

$$z_1 + y_2 \qquad\qquad\qquad + z_7 + y_8 \geqslant 1 \tag{30}$$

$$\quad\ y_2 + y_3 + z_4 \qquad + z_6 \qquad + y_8 \geqslant 1 \tag{31}$$

$$z_1 \quad + y_3 + z_4 + y_5 + z_6 \qquad\qquad \geqslant 1 \tag{32}$$

$$\quad\ y_2 \qquad\qquad\qquad + z_7 + y_8 \geqslant 1 \tag{33}$$

Some solutions (Y, Z) also satisfying (21) are:

$$S_1' = (y_2 = y_5 = y_7 = 1; \ 0 \text{ otherwise}) \qquad S_4' = (z_1 = y_2 = y_5 = 1; \ 0 \text{ otherwise})$$

$$S_2' = (y_4 = y_5 = y_7 = y_8 = 1; \ 0 \text{ otherwise}) \quad S_5' = (z_1 = y_2 = z_6 = 1; \text{ otherwise})$$

$$S_3' = (y_5 = y_6 = y_7 = y_8 = 1; \ 0 \text{ otherwise}) \quad S_6' = (y_2 = z_6 = y_7 = y_8 = 1; \text{ otherwise}).$$

Solutions S_1', S_2' and S_3' correspond to S_1^*, S_2^* and S_3^*, respectively, in section 8.1. These all give positive extensions as discussed therein. S_4' gives the extension ψ' already introduced in the beginning of this section. S_5' and S_6' respectively provide the unate extensions:

$$\psi_5 = \bar{x}_1 x_2 \vee x_2 \bar{x}_6 \ (= \Pi_4)$$

$$\psi_6 = x_2 x_7 \vee \bar{x}_6 x_8.$$

S_1', S_4' and S_5' are basic as these extensions are found in the list of table 4. The other extensions are not basic, though all these are constructed from the solutions to (22)–(33), which are minimal in the sense that converting any $y_i = 1$ or $z_i = 1$ to 0 does not give a solution to (22)–(33).

Therefore, the question of whether a given partially defined Boolean function admits a unate extension can now be stated as follows: does the set of inequalities (19)–(20) have a solution (Y, Z) satisfying (21)? It is of course possible to answer this question by exhausting all possible 0-1 solutions (Y, Z) of (19)–(20) and checking whether some of them satisfy (21). This is not practically feasible, however, unless the number of variables and the number of vectors in $T \cup F$ are rather small. Unfortunately, the next theorem states that it is most unlikely that the above question can always be answered efficiently.

THEOREM 8

Deciding whether a partially defined function has a unate extension is NP-complete. □

The proof is rather technical and will be given in the appendix. For the concept of NP-completeness and the related complexity issues, the reader is referred, for example, to [3].

8.3. LOW-ORDER THEORIES

The order $O(C)$ of an elementary conjunction C is the number of literals occuring in it; e.g., $x\bar{y}$ is of order 2, x is of order 1, the constant 1 is of order 0, etc. The order of a disjunctive normal form $C_1 \vee C_2 \vee \cdots \vee C_k$ is simply the maximum of $O(C_i)$ for $i = 1, \ldots, k$; e.g., the order of $x \vee \bar{y}zw$ is 3.

It is known that every Boolean function can be expressed as a (not necessarily unique) disjunctive normal form. Accordingly, we define the order of a Boolean function as the order of its expression(s) of lowest order; for example, the order of the function given by the expression $\bar{z}t \vee \bar{t}x \vee xyz$ is 2, since the same function also admits the expression $\bar{z}t \vee \bar{t}x \vee xy$ (and since it obviously has no expression of order 1).

Theories of order 1 are the simplest ones; by analogy to the usual algebraic concept, such theories will be called *linear*. Similarly theories of order 2, or *quadratic* theories, can be written as unions of conjunctions involving at most two variables. A conjunction in two variables can be interpreted in a straightforward way. In general the lower the degree of a theory, the simpler its interpretation is.

The interest in restricting our attention, whenever possible, to theories of low order, is therefore obvious. It turns out that this goal can be accomplished within the framework of thrifty theories, as shown by:

THEOREM 9

If a partially defined function has an extension of order k, then it has a thrifty extension of order at most k. In particular, there exists an extension of lowest order which is thrifty. □

An extension of lowest possible order can therefore be simply built by modifying the second procedure of section 6, in the following way. At every stage, when a new pattern is chosen to cover one of the unlabelled points in T, this pattern is chosen as short as possible.

It is easy to see that this "greedy" procedure always generates a thrifty expression of minimum order (although not necessarily one involving a minimum number of patterns). For any given order k, it gives a polynomial-time constructive answer to the question of the existence of an extension of order at most k.

EXAMPLE 5

Does the partially defined function f given by:

x_1	x_2	x_3	x_4	f
1	1	1	1	1
0	1	1	1	1
1	1	0	1	1
1	0	1	1	0
0	1	1	0	0
0	1	0	1	0

have a quadratic thrifty theory? In order to answer the question, we determine a pattern of smallest order covering the first vector, e.g. $x_1 x_2$. The same pattern also covers the third vector. We try in turn to cover the second vector with a linear or quadratic pattern. Since this is impossible, we conclude that the answer to the above question is negative. However, the pattern $\bar{x}_1 x_3 x_4$ (or alternatively $x_2 x_3 x_4$) covers the second vector. Hence, the function admits thrifty theories of order 3, for example $\Pi = x_1 x_2 \vee \bar{x}_1 x_3 x_4$.

8.4. THRESHOLD THEORIES

Some cause-effect relationships display the interesting feature that a constant "weight" can be attributed to every possible cause, in such a way that the effect under study is triggered whenever the sum of the weights of the observed causes exceeds a certain "threshold".

For instance, "headaches" or other symptoms may be associated with diets involving excessive amounts of a certain nutrient, found in varying quantities in usual food items. Also, an electrical circuit may fail whenever the intensities of several currents traversing a given component add up to some critical value.

Formally, a Boolean function $\psi(x_1, \ldots, x_n)$ is called *threshold* if there exist numbers $w_1, \ldots, w_n$, q such that:

$$\psi(x_1, \ldots, x_n) = 1 \text{ if and only if } \sum_{i=1}^{n} w_i x_i \geq q. \tag{34}$$

The problem of deciding whether a partially defined Boolean function admits a threshold extension is easily settled. Indeed, this will be the case if and only if the following system has a solution $(w_1, \ldots, w_n, q)$:

$$\sum_{i=1}^{n} w_i x_i \geq q \quad \text{for all } X \in T$$

$$\sum_{i=1}^{n} w_i x_i \leq q - 1 \quad \text{for all } X \in F.$$

This system can be solved using linear programming techniques.

EXAMPLE 1 (continued)

To decide whether the partially defined function given by table 2 has a threshold extension, we set up the system:

$$
\begin{array}{llllll}
w_2 & +w_4 & +w_6 +w_7 & & \geq q \\
w_1 +w_2 & +w_4 +w_5 & & +w_8 & \geq q \\
w_2 +w_3 & +w_5 & & +w_8 & \geq q \\
w_1 & +w_3 & +w_5 & +w_7 & \leq q-1 \\
& w_4 +w_5 +w_6 & & & \leq q-1 \\
w_1 +w_2 & +w_4 & +w_6 & +w_8 & \leq q-1 \\
& w_3 & +w_5 & +w_7 & \leq q-1
\end{array}
$$

A solution of this system is for instance: $w_2 = 2$, $w_5 = w_7 = 1$, $w_1 = w_3 = w_4 = w_6 = w_8 = 0$, $q = 3$. Via the relationship (34), these values define the Boolean function $\Pi_5^1 = x_2 x_5 \vee x_2 x_7$ (see table 4). Another solution of the above system would be $w_1 = w_6 = -1$, $w_2 = 2$, $w_i = 0$ for all other i, $q = 1$, corresponding to the threshold extension $\Pi_4 = \bar{x}_1 x_2 \vee x_2 \bar{x}_6$.

Notice that, even when the procedure described above identifies a threshold extension, it does not produce a disjunctive normal form of that extension, but only an implicit representation described by the weights $w_1, \ldots, w_n$, the threshold value q, and the relation (34). Using that information to build a disjunctive form of the extension may still be a laborious task.

Finally, it should be remarked that the existence of a threshold theory does not guarantee the existence of a thrifty, or of a basic, threshold theory.

EXAMPLE 6

Consider the partially defined Boolean function given by $T = \{(1, 0, 1, 1), (1, 1, 0, 0), (1, 1, 0, 1), (1, 1, 1, 0), (1, 1, 1, 1)\}$ and $F = \{(0, 0, 0, 0), (0, 0, 0, 1), (0, 0, 1, 0), (0, 1, 0, 0), (0, 1, 0, 1), (0, 1, 1, 0), (1, 0, 0, 0), (1, 0, 0, 1), (1, 0, 1, 0)\}$. The four extensions of this function are:

$$\psi_1 = x_1 x_2 \vee x_1 x_3 x_4,$$
$$\psi_2 = x_1 x_2 \vee x_1 x_3 x_4 \vee \bar{x}_2 x_3 x_4,$$
$$\psi_3 = x_1 x_2 \vee x_1 x_3 x_4 \vee x_2 x_3 x_4,$$
$$\psi_4 = x_1 x_2 \vee x_3 x_4.$$

Of these functions, only ψ_1 and ψ_3 are threshold. In particular, ψ_4, the unique thrifty extension of f, is not threshold.

EXAMPLE 4 (continued)

For the function given in example 4 (section 8.1), the extension $\Pi = x_1x_3 \vee x_2x_3$ is threshold. But the unique basic extension, $\Pi' = x_1\bar{x}_2 \vee \bar{x}_1x_2$, is not threshold.

9. Prevention

Once a theory has been tentatively accepted, the next decision concerns its use for the prevention of further headaches (obviously, in other situations, our goal may be to provoke the phenomenon rather than to prevent it; in this case, the method proposed below works, after some obvious Boolean adjustments).

A *preventor* associated with a theory ψ is a prime implicant of the complement $\bar{\psi}$ of the theory. As usual, $\bar{\psi}$ is the Boolean function which takes value 1 in exactly those vectors were ψ takes value 0. An implicant of $\bar{\psi}$ is then an elementary conjunction with the property that, when it is 1, then $\bar{\psi}$ is 1, i.e., $\psi = 0$. A prime implicant of $\bar{\psi}$ is an implicant with no redundant literals; intuitively, it can be viewed as a minimal list of instructions, listing food items to be eaten or to be avoided in order to prevent the headache.

EXAMPLE 1 (continued)

If the theory chosen for this example is $\Pi_4 = \bar{x}_1x_2 \vee x_2\bar{x}_6$, then $\overline{\Pi}_4 = \bar{x}_2 \vee x_1x_6$, having as unique prime implicants $\bar{x}_2$ and x_1x_6. The first preventor will simply forbid the consumption of food #2. The second preventor will not forbid any food, but will recommend the consumption of both foods #1 and #6. It is up to the doctor to choose between the two preventors, or even to recommend his patient to alternate between a diet not containing food #2, and one containing both foods #1 and #6.

This is a good place to reiterate the fact that the unavailability of complete data makes the elaboration of foolproof theories impossible. A preventor will have a guaranteed effect if the chosen theory happens to describe precisely the phenomenon at hand. However, even when this is not the case, the preventor may prove effective in eliminating, or at least in reducing the occurence of headaches.

EXAMPLE 1 (continued)

We have seen that $\bar{x}_2$ is one of the preventors derived from theory Π_4. However, we can see from the list in table 4 that $\bar{x}_2$ will completely prevent headaches even if the correct theory is not Π_4, but is one of the theories Π_3^1, Π_5^1 or Π_6^1. Finally, if the correct theory turns out to be Π_3^2, Π_5^2 or Π_6^2, then $\bar{x}_2$ will still reduce the number of headaches, by preventing the occurrence of some of the patterns considered in these theories.

If a preventor turns out not to prevent all headaches, that obviously signals the fact that the wrong theory was chosen; in this situation, the addition of the new observations to the previous list can be used to build a new theory.

Actually, in some situations, the decision maker may be unwilling to commit himself to a particular theory of the phenomenon at hand, and would therefore not be able to determine preventors as defined above. In such cases, an alternative approach to the prevention problem can be described as follows.

With a partially defined function f, taking value 1 on T and 0 on F, let us associate the partially defined function $\bar{f}$ (*complement* of f) taking value 1 on F, 0 on T, and unspecified elsewhere. Clearly, every extension of $\bar{f}$ is the complement of an extension of f. Combining this observation with theorem 1, it follows immediately that every pattern of $\bar{f}$ is a preventor for some theory associated with f. The list of patterns of $\bar{f}$ constitutes therefore a list of preventors, which can be obtained independently of the choice of a theory for f.

Similarly, minimal supporting sets of variables, thrifty theories, basic theories,... can now be associated with $\bar{f}$, just as they were earlier with f. These tools allow a direct study of the prevention problem, as opposed to the "dual" approach outlined at the beginning of this section.

10. Conclusions

We would like to reiterate here, and emphasize, our ignorance in medical matters, and our awareness of the naivity of our analysis of the headache problem. We also want to emphasize the fact that headache served in this paper simply as a model to illustrate our view of cause-effect relationships. Our aim has been to analyze such relationships when only partial information was available, and to ignore characteristic features of human headaches, elevator failures, or other specific situations arising in chemistry, ecology, political science, etc. We do think that, in many of these situations, the ideas of minimal supporting sets of variables, of patterns, of basic, thrifty and other theories, of preventors, etc., can have a meaningful interpretation.

We also feel that, among the general features of cause-effect relationships, there are many important ones which have not been examined in this paper. They can, and hopefully will, constitute the object of future research. Two of the most obvious directions in which further research is needed are the following.

First, the underlying assumption of our work is that no information is available about the phenomenon under study, except for the list of observed situations. However, in most types of real problems, statistical and other exogeneous evidence can be, and is, used in elaborating theories. The design of general procedures incorporating such additional information may lead to substantial improvements in our current approach.

Second, any theory built on the basis of limited observations can only be accepted tentatively, and its validity has to be assessed by further experimentation. An efficient design of experiments has to aim at the rapid refutation of as many false theories as possible. It should also allow us to discriminate easily between alternative potential theories. Deriving such testing procedure is an important direction of research.

However, the bottom line remains the same: a theory can sometimes be refuted, but almost never be fully confirmed. The present paper attempts to give a further confirmation of this theory.

Appendix

PROOF OF THEOREM 8

We assume here that the reader is familiar with the concept of NP-completeness, and give only technical details. The following problem is known to be NP-complete.

Satisfiability: Given a set of clauses C_j, $j = 1, 2, \ldots, m$, each of which is a subset of $L = \{ x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_n, \bar{x}_n \}$ containing at most one of x_i and $\bar{x}_i$ for each i, decide whether there is a subset C^* of L with the following properties; for each i, exactly one of x_i and $\bar{x}_i$ belongs to C^*, and

$$C^* \cap C_j \neq \emptyset, \quad j = 1, 2, \ldots, m.$$

Given an arbitrary instance of *satisfiability*, we now construct the corresponding partially defined function f, such that the set of inequalities (19)–(20) derived from f has a solution (Y, Z) satisfying (21) if and only if the original instance of *satisfiability* has a solution. This shows that *satisfiability* is reducible to our problem of deciding the existence of unate extensions. Since it is trivial to see that our problem belongs to the class NP, this implies that our problem is NP-complete.

Now construct two vectors $X^j = (x_1^j, x_2^j, \ldots, x_n^j, x_{n+1}^j, \ldots, x_{n+m}^j)$ and $Y^j = (y_1^j, \ldots, y_{n+m}^j)$ corresponding to C_j, $j = 1, 2, \ldots, m$, as follows:

$x_i^j = 1$ and $y_i^j = 0$ if $x_i \in C_j$

$x_i^j = 0$ and $y_i^j = 1$ if $\bar{x}_i \in C_j$

$x_i^j = 1$ and $y_i^j = 1$ if $1 \leq i \leq n$ and $x_i, \bar{x}_i \notin C_j$

$x_{n+j}^j = y_{n+j}^j = 1$

$x_{n+i}^j = y_{n+i}^j = 0$ if $i \geq 1$ and $i \neq j$.

Similarly, for each $j = 1, \ldots, m$, we introduce a vector Z^j, with $z_i^j = x_i^j$ for $i \neq n + j$ and $z_{n+j}^j = 0$.

Then, define

$$T = \{ X^j \mid j = 1, 2, \ldots, m \}$$

$$F = \{ Y^j, Z^j \mid j = 1, 2, \ldots, m \}.$$

The set of inequalities (19)–(20) derived from these T and F can be classified as follows:

(A) For $X^j \in T$ and $Y^j \in F$ (for the same j),

$$\sum_{x_i \in C_j} y_i + \sum_{\bar{x}_i \in C_j} z_i \geqslant 1, \quad j = 1, 2, \ldots, m.$$

(B) For $X^j \in T$ and $Z^j \in F$ (for the same j),

$$y_{n+j} \geqslant 1, \quad j = 1, 2, \ldots, m.$$

(C) Others.

From the set of inequalities (B), we see that $y_{n+j} = 1$ must hold for $j = 1, 2, \ldots, m$. This implies that the inequalities in (C) are all redundant since any inequality in (C) contains at least one y_{n+j} in its left hand side. Therefore our problem becomes equivalent to deciding whether the set of inequalities (A) has a solution satisfying in addition the constraint (21). It is now obvious that such a solution exists if and only if the original instance of *satisfiability* has a solution. □

References

[1] P.R. Cohen and E.A. Feigenbaum, eds., *The Handbook of Artificial Intelligence, Vol. III* (W. Kaufmann, Inc., Los Altos, 1982).
[2] R.B. Cutler and S. Muroga, Useless prime implicants of incompletely specified multiple-output switching functions, International Journal of Computer and Information Sciences 4 (1980) 337–350.
[3] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman and Co., San Francisco, 1979).
[4] Yu. I. Juravliov, On the separability of subsets of vertices of the *n*-dimensional unit cube, Trudy Matematičeskovo Instituta Imeni V.A. Steklova 51 (1958) 143–157 (in Russian).
[5] F. Lapscher, Fonctions Booléennes très incomplètes: représentation par des sommes de monômes, Revue d'Informatique et de Recherche Opérationnelle 4 (1967) 73–88.
[6] E. Mendelson, *Theory and Problems of Boolean Algebra and Switching Circuits*, Schaum's Outline Series (McGraw-Hill, New York, 1970).
[7] S. Muroga, *Logic Design and Switching Theory* (Wiley, New York, 1979).
[8] N.N. Necula, O metodă pentru reducerea numărului de variabile al functiilor Booleene foarte slab definite, Studii Cercetari Matematice 24 (1972) 561–566.
[9] L.G. Valiant, A theory of the learnable, Communications of the ACM 27 (1984) 1134–1142.
[10] Yu.A. Zuev, Approximation of a partial Boolean function by a monotonic Boolean function, U.S.S.R. Computational Mathematics and Mathematical Physics 18 (1979) 212–218.