ELSEVIER
Artificial Intelligence 91 (1997) 205–223
Artificial Intelligence

Induction and the discovery of the causes of scurvy: a computational reconstruction

Vincent Corruble *, Jean-Gabriel Ganascia [1]

LAFORIA-IBP, Université Paris VI, 4 Place Jussieu, Boîte 169, 75252 Paris Cedex 05, France

Received December 1995; revised November 1996

Abstract

The work presented here addresses the problem of inductive reasoning in medical discoveries. The discovery of the causes of scurvy is studied and simulated using computational means. An inductive algorithm is successful in simulating some essential steps in the progress of the understanding of the disease and also allows us to simulate the false reasoning of previous centuries through the introduction of some *a priori* knowledge inherited from pre-clinical medicine. These results confirm the good results obtained by other AI researchers with an inductive approach of discovery, and illustrate the importance of the social and cultural environment on the way the inductive inference is performed and on its outcome. © 1997 Published by Elsevier Science B.V.

Keywords: Scientific discovery; Induction; Medicine; Scurvy; Inductive bias

1. Introduction

This article explores the potential of symbolic induction for scientific discovery. As such, it is related to the data-driven line of research developed since the 1970s in the study of computational models of scientific discovery (see for example [15] for a general presentation of this approach, and [13] for a presentation of the system BACON). However, the specific problem addressed here is original both with regards to the *application domain*, and to the actual *task* studied.

Work on computational models of scientific discovery has traditionally focused on hard science. Among them, the most important ones are undoubtedly physics and

* Corresponding author. E-mail: corruble@laforia.ibp.fr.
[1] E-mail: ganascia@laforia.ibp.fr.

PII S0004-3702(97)00014-3

chemistry. A characteristic of these sciences is that they rest, at least over the last century, on relatively well-established formalisms. At the opposite, work on medicine is in large part based on the interpretation of symptoms, which only point toward the underlying physiological phenomena. Medicine remains a rich domain to be explored by researchers interested in computational models of scientific discovery. It has been until now largely neglected by the AI community, with a few notable exceptions (see for example [22,25]).

The second novelty is in the type of inference studied. As said above, induction has been studied extensively in the research on computational models of scientific discovery. However, the main tasks modeled *with an inductive inference* have been the discovery of numerical laws with the BACON family of systems, or the formation of taxonomies with systems such as GLAUBER [14]. The work presented here focuses on the induction of *causal hypotheses* applied to the field of medical etiology (the field interested in finding the causes of diseases). We return later to the relation between induction and causality in the context of medicine. One of the rare computational applications of induction of causal hypotheses can be found in the system TETRAD [23], yet this system manipulates *numeric* data and is deeply entrenched in a *probabilistic* framework; it is therefore only slightly related to our focus on *symbolic* induction.

Another important issue for any work based on historical data is the choice of focus on *normative* (what *should* have been done by the scientist) or *descriptive* models (what was actually done by the scientist). This distinction is explored in detail in [24]: the normative focus was traditionally chosen in philosophy of science (until the development of historical philosophy of science), logic and also generally in artificial intelligence. Descriptive approaches are more common in history of science. We find it useful to introduce a third, intermediary, type of focus, named *prescriptive*—interested in what a scientist *could* have done (given the same situation). While a normative focus can be interpreted as interested in defining right science, a prescriptive focus would in this context be interested in proposing methods for good science, with no claim to exclusivity. Our approach in this article is to start from a prescriptive focus in a first stage, and then to move toward a descriptive one. Specifically, in a second stage, we are interested in a computational reconstruction of the *false* reasoning performed in history.

The next section gives some background on the role of induction in medical discovery by outlining three major periods, pre-clinical, clinical, and experimental medicine. Section 3 presents experiments carried out on the simulation of the discovery of the causes of scurvy, both from a prescriptive and descriptive point of view using the general inductive algorithm CHARADE. Section 4 concludes by discussing the perspectives opened by our work for the aid to discovery in contemporary medicine and introduces some projects going in this direction.

2. Discoveries and induction in medicine

Induction, the inference aiming at the general from the particular, has had a rather tumultuous relationship with science over the centuries. The role of induction in medicine is possibly even more problematic than in other sciences. This can be surprising to the modern mind for whom basing theories on observation sounds like the only natural and correct way of doing science. During a long period culminating in the Middle Ages, the very idea of relying mainly on observation to reach a diagnosis or a new medical theory was seen as anti-scientific, because "*scientific*" practice was supposed to be based on a number of abstract theoretical systems, these systems being a direct reflection of the culture of the scientist. In this article, we will frequently use the term *system*, as it has often been used in the history of medicine, to refer to a type of conceptual framework. One instance of such a system is Galen's fluid theory which was inherited from Hippocrates. In modern medicine, genetic determinism constitutes another such explanatory system. Even though the notion of explanatory system is related to the notion of paradigm proposed by Kuhn, they have two features which are not typical of paradigms. First, explanatory systems can be implicit, as an element of the culture of the scientist which is not necessarily acknowledged. Second, and most important, explanatory systems can compete with each other at a given time, but do not have to exclude one another; [2] two different systems can coexist within the same community and sometimes even for the same researcher.

The importance of explanatory systems in medicine does not preclude references to observations. The role of observations has evolved through history, ranging from the status of anecdotal illustration to the one of evidence. In order to have a better idea of this evolution, and thus to make clear the role of induction in Western medicine, let us see the way it was considered in the different periods which Foucault [8] analyzed.

In *pre-clinical* medicine (which developed from the Middle Ages, after the medicine of ancient Greece and Rome [5]), the use of direct observation was seen as a definite lack of culture, characteristic of a layman. Observations were useful, but only if they fell within the framework proposed by the systems inherited from Antiquity. In parallel to this official, academic medicine, it is important to realize that other types of medicine coexisted. Some branches of medicine were not directly influenced by these theoretical systems. This was particularly the case of surgeons who, instead of receiving a theoretical teaching in universities, were trained as apprentices under a senior surgeon.

A key shift in the history of medicine occurred when the medical community developed *clinical medicine*. One of the central ideas in this shift was to raise the status of direct observation to such a level that it became the only acceptable element on which medicine should be based, thus rejecting any theory or system that was not the direct result of observation. This shift is well illustrated by Corvisart claiming: "Theories are silent or vanish at the sick-bed" (quoted in [8]). The scientific progress resulting from this new approach should not be underestimated. Medical expertise which, until then, was often evaluated according to the number of references to previous research produced by the so-called expert, was now evaluated according to clinical efficacy:

[2] As defined by Kuhn in [12, p. 294], "a paradigm is what the members of a scientific community, and they alone, share. Conversely, it is their possession of a common paradigm that constitutes a scientific community of a group of otherwise disparate men." This definition therefore does not allow for the coexistence of two competing paradigms in the same community for a period of time.

could the patient be saved, could a cure be found? This resulted in many of the astounding progresses of medicine during the last century.

Soon after the development of clinical medicine, the field saw the development of *experimental medicine*. Experimental medicine is not opposed to clinical medicine because it is also based on observation. The main difference (beautifully analyzed in [2]), is that experimental medicine generates the observations to test a hypothesis already formulated, while clinical medicine is more passive (cognitive activity aside). Experimental medicine has been very successful and thus soon overshadowed the clinical revolution. By generating observations, experimental medicine does not directly reach its conclusion through induction. It is worth noting that experimental medicine does use induction, but mainly at the stage of hypothesis formation.

This historical perspective raises important questions for machine discovery about the potential scope of purely data-driven approaches. Reciprocally, computational methods can be used to study experimentally these questions, which are essential to epistemology and history of medicine. Nevertheless, medical discovery has been largely neglected by the AI community. In the next section, we present our experiments on an inductive reconstruction of the discovery of the causes of scurvy.

3. The discovery of the causes of scurvy

A brief history of the discovery of the causes of scurvy is given in Section 3.1. Afterwards, Section 3.2 presents some experiments on the rational reconstruction of the inductive reasoning performed by 18th and 19th century physicians, using computational means. Historical observations of cases of scurvy have been collected in the medical literature and given as inputs to the symbolic induction system CHARADE. The rules which are produced by the system are similar to the explanations given for scurvy in the 18th and 19th centuries. However, physicians had great difficulties in perceiving and understanding the causes of scurvy. As seen in Section 3.3, our results show the importance of the implicit knowledge of the scientists—related to the notion of inductive bias studied in the field of machine learning—to explain these difficulties.

3.1. A brief history of scurvy

Scurvy has been the cause of over a million of deaths aboard commercial and navy ships, and also on land, though to a lesser extent [1,20]. The disease took on an increased importance in the 15th century, with the development of long circumnavigations [4], and also in the 17th and 18th century, with the development of long missions in European navies, which involved numerous seamen. It is striking to note that scurvy is said to have caused more deaths in the French navy than combat with other European navies.

Consequently, research on the causes of scurvy attracted the brightest minds of the time. Among them, James Lind is famous for his remarkable *Treatise of the Scurvy* [16]. From the 17th to the 20th century (when the actual cause of the disease, i.e., the lack of vitamin C, was discovered), dozens of theories were put forward on the origins of the disease. Many of these were totally disconnected from the real cause such as those referring to the psychological effect of being at sea, far away from home. Other theories were quite close to finding the real cause, especially the one that became widely accepted, which said that scurvy was the result of the conjunction of the humidity in the air and of the lack of fresh fruits and vegetables in the diet.

The lack of fresh fruits and vegetables was eventually accepted as the only cause of scurvy in the early 20th century. A first explanation for this late discovery is the lack of a concept necessary for a global understanding of the disease. The concept of vitamin, i.e., the idea that a small quantity of a chemical has a great influence on the functioning of the human body, is a key to the comprehension of the mechanism leading to scurvy. However, knowing that seamen were acquainted with the importance of fresh fruits and vegetables as early as the 15th century, it is surprising that a practical cure was not widely accepted earlier. To put it into more formal terms, though the lack of an *explanatory adequate theory* of scurvy is easily understandable due to the necessity of the concept of vitamin, the reason for the absence of a *descriptively adequate theory* (i.e., a theory establishing only the conditions of development of the disease) is unclear.

This last question provided a good motivation to try modern inductive techniques on observational data available in the 19th century, before the actual discovery was made. This attempt is described in the following sections.

3.2. Induction on scurvy cases

3.2.1. Introduction

A first experiment makes the assumption of pure inductivism: only the descriptions of cases are given to the inductive system. The rules induced are then analyzed according to a number of steps. The next sections detail the process of data collection, and the algorithm used. A third section focuses on the method used for the analysis of results proposed by the algorithm.

3.2.2. Data collection

In order to give our work a real simulation value, it has been necessary to use as training examples case descriptions that were as close as possible to the descriptions that were made before the discovery of the causes of scurvy. Therefore, the examples used all come from the 1880 *Dictionnaire Encyclopédique des Sciences Médicales* [17], which provides relatively detailed descriptions of 25 cases of scurvy. [3] During the necessary translation of these natural language descriptions into the description language of the inductive system, one main objective was to remain as close as possible to the original, without modifying the description by using our own knowledge of the disease.

Ten features which constitute the description language given to the inductive system were found in most historical cases. They are summarized in Fig. 1. Besides the date and location of the case, they include the temperature, the humidity in the air, the hygiene

[3] The database of 25 cases used for these experiments can be obtained by following pointers from the first author's web page (http://www-laforia.ibp.fr/~corruble) or by sending e-mail (corruble@laforia.ibp.fr).

Attribute	Type	Domain
year	integer	N^+
location	string	NA
temperature	ordered set	severe-cold < cold < average < hot < very-hot
humidity	ordered set	low < high < very-high
food-quantity	ordered set	starvation<severe-restrictions<restrictions<OK
food-variety	ordered set	low < average < high
hygiene	ordered set	very-bad < bad < average < good < very-good
type-of-location	unordered set	land, sea
fresh_fruits/vegetables	Boolean	yes, no
disease-severity	integer	(0,1,2,...,5)

Fig. 1. List of attributes with their characteristics.

level, the quantity of food, its variety, the use or absence of fresh fruits and vegetables, the type of location (at sea or on land), and, lastly, the severity of the disease. Each of these attributes is defined by a type (ordered or unordered set, integer, etc.) and domain, thus enabling the induction process to take full advantage of the structure of the data. It is important to notice that the choice of features used for our representation is done based on a syntactical analysis of the historical text. If a feature is present in the historical descriptions of more than a very few cases, then it is included in the description language of all the cases. Also, the induction system used does not require a value for every attribute and each case, so that missing values did not have to be filled arbitrarily with estimates.

Note that no attribute in the description language refers to the time spent at sea. The main reason for this is that the original descriptions found in the literature did not usually contain this information, or when they did, it was in a very qualitative way, difficult to reuse. These descriptions however often mention that the disease severity has developed over time, in which case the value chosen for the disease severity attribute is the one observed at the end (the beginning being seen as a transition phase). In some other cases, the disease can develop in a first phase, and then decrease or even disappear in a second phase. Since this second phase usually corresponds to new environmental conditions (new values for the other attributes), the phenomenon has been represented by creating two training cases: one corresponding to the first phase, and one for the second phase. Theses cases of remission are particularly useful in providing "negative" examples of the disease to the system, necessary for any supervised learning (Fig. 2 contains such a negative example). Also, by providing two training cases which are relatively similar except for a few attributes, each case of remission greatly helps the inductive process.

Examples of training cases will be given in the following sections but it is necessary first to provide some details on the inductive system used.

3.2.3. The induction system CHARADE

CHARADE [9,10] is a symbolic inductive system which extracts logical rules expressing empirical regularities between attributes in a set of examples. More precisely,

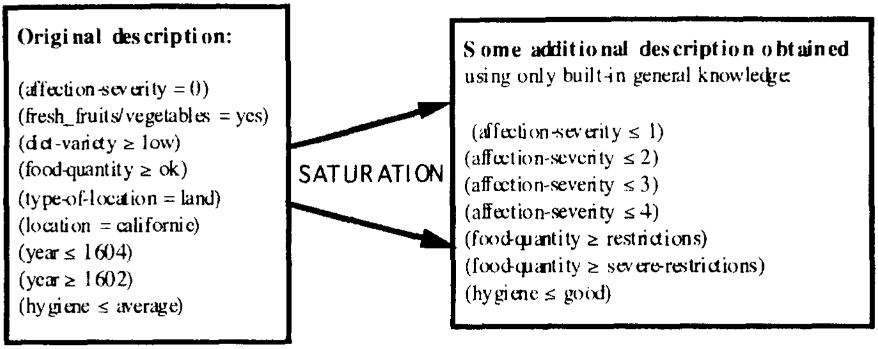

Fig. 2. An example of saturation on one training case.

CHARADE can be classified as a k-DNF learner, which means that it generates sets of formulae of the type $d1$ & $d2$ & $\cdots$ & $dn \Rightarrow c$. These rules correspond to regularities observed on a set of training examples. CHARADE is able to generate all non-redundant regularities, but it can also be restricted to generate rules verifying some particular properties corresponding to what is called a *learning bias* in the area of machine learning (see for instance [10,21,27]). The aim of this paper is not to explain the theory of CHARADE; those interested are referred to the previous papers to obtain more details. However, some interesting features of CHARADE are highlighted in this section.

One of the main advantages of CHARADE is that most of its learning bias is explicit. CHARADE takes as input a description language which defines all attributes, their type (unordered set, ordered set, etc.), and their domain. Another input is the set of training examples. Also, a feature of particular interest here is the possibility of formalizing some domain knowledge using axioms. This knowledge (entered as production rules), as well as some built-in general knowledge about the types of attributes, is taken into account during a first phase, named saturation. This phase is used to enrich the description of each training case, as illustrated in Fig. 2 using example #6 from the database.

The inductive process takes advantage of the lattice structure of the search space to produce all the logical IF-THEN rules consistent with the training examples that can be expressed with some conjunctions of elements of the descriptions. Some algebraic properties of this lattice are taken into account to cut out of the search some large parts of this lattice (an exponential search space), which are known not to contain any new rules. The algorithm finds the rules that do not have any exceptions: a rule does not need to cover [4] all the examples to be produced, but if at least one of the examples contradicts the candidate rule, the rule will not be produced. Even if a training case has

[4] A rule *covers* an example when all the premises of the rule are matched by the description of the example. An example *contradicts* a rule if and only if the example is covered by the rule and the conclusion of the rule is not matched by the description of the example. The cover of the rule is, depending on the context, the set of cases it covers, or the number of cases in this set.

been successfully classified by a rule, the system will continue searching for other rules for this example, hence providing alternative rules for each case. This last feature of CHARADE is useful to our study here, which is based on the empirical comparison of competing induced theories.

The last type of bias which can optionally be defined by the user is the desired structure of the output rules. It is possible to restrict the search to all the rules that conclude on one group of attributes. For instance, in the case of scurvy, the rules which are searched for are the ones that conclude on the severity of the disease, because this is what we want to predict or explain.

3.2.4. Methodology for analysis of results

What the inductive algorithm outputs is a rule base. These rules do not have a specific ordering or structure; they are proposed in the order in which they were found. This section presents the three steps of analysis performed manually in our experiments.

- *Regrouping of rules.*
  The rule set obtained is partitioned in a number of subsets. Two rules are put in the same subset, if (1) their premises use the same factor for prediction, and (2) they are consistent (they should not contradict each other). Note that a rule using two different attributes in its condition could potentially be assigned to two subsets according to this procedure. Then, it is possible either to assign the rule to both subsets (in which case this step does not lead to a partition), or to assign the rule only once to a new subset of rules combining these two factors. This situation did not occur in our experiments on scurvy.
  Each subset of rules obtained this way can then be considered a proto-theory because of its coherence (in the factor put forward, and in the predictions of the rules). An application of this step is found in Fig. 3.
- *Evaluation of explanatory power.*
  The evaluation here is not done according to the formal method advocated by statistics or most machine learning work (prediction on new data). We are here more interested in an informal, exploratory prior-assessment as done in medical research to decide which hypothesis to pursue (the concepts of prior-assessment and pursuit are well explained in [22]). Therefore, we use the number of individual cases explained by each *subset* of rules as a criterion. This is indirectly related to the concept of *consilience* [24] and *verisimilitude* [19] except (this is a significant difference) that these refer to *classes* of facts explained rather than *individual* cases. On the practical side, it is important to see that the cover of a subset is not obtained by adding the covers of each individual rule of the subset (this would count more than once cases covered by more than one rule). Instead, the cover is the cardinal of the union of the sets of examples covered by each rule in the subset.
- *Correspondence with history.*
  The last step is to look at the correspondence with history. Does each proto-theory correspond to a theory proposed in history? Reciprocally, does each theory proposed in history find a corresponding theory in our reconstruction? This notion

```
Set I: Rules 3,4,8 use in their premises the variety of the diet.
R3:  IF diet-variety ≥ high          THEN disease-severity ≤ 0.  [5]
R4:  IF diet-variety ≤ average       THEN disease-severity ≥ 3.  [4]
R8:  IF diet-variety ≥ average       THEN disease-severity ≤ 2.  [11]

Set II: Rules 7, 10 use in their premises the presence (or absence) of
fresh fruits and vegetables in the diet.
R7:  IF fresh_fruits/vegetables = no  THEN disease-severity ≥ 2.  [5]
R10: IF fresh_fruits/vegetables = yes THEN disease-severity ≤ 2.  [13]

Set III: Rule 2 uses in its premises the quantity of food available.
R2:  IF food-quantity ≥ ok           THEN disease-severity ≤ 0.  [4]

Set IV: Rules 5,6,9,12 use in their premises the level of hygiene.
R5:  IF hygiene ≤ bad                THEN disease-severity ≥ 3.  [3]
R6:  IF hygiene ≤ average            THEN disease-severity ≥ 2.  [4]
R9 : IF hygiene ≥ average            THEN disease-severity ≤ 2.  [7]
R12: IF hygiene ≥ good               THEN disease-severity ≤ 1.  [6]

Set V: Rules 1, 11 use in their premises the temperature.
R1:  IF location = land,
     temperature ≥ hot               THEN disease-severity ≤ 0.  [4]
R11: IF temperature ≤ severe-cold    THEN disease-severity ≥ 1.  [5]
```

Fig. 3. Rules proposed by CHARADE regrouped according to the factor they put forward. Note that number between brackets at the right of each rule stating how many examples it covers (i.e., the number of examples that match the premises of the rule). *Disease severity* is a ranking from 0 (no disease at all) to 5 (most severe).

of correspondence, essential to the enterprise of rational reconstruction, is peripheral to our first experiment, and becomes central in the second one. Here correspondence does not aim at a full isomorphism between history and our simulation. Our reconstruction, in our first and also in our second experiment, does not take into account *all* the facts, knowledge and beliefs about scurvy available to physicians in the 18th and 19th century. Our reconstructions are therefore a simplification from a psychological standpoint. Yet they are powerful means for evaluating hypotheses about the nature of the discovery process, both prescriptively (e.g. how inductive could this discovery have been?) and descriptively (e.g. why was this wrong theory defended for so long?).

3.3. Experimentation

3.3.1. First results and analysis

Our first experiment was done without giving any domain knowledge to the system. This "pure" induction produced 12 logical rules which are provided in Fig. 3 with some analysis. These 12 rules were then manually regrouped according to the factor they put

DIET (food *variety*):

It was J.F. Bachström (1734) who first expressed the opinion that, "Abstinence of vegetables is the only, the true, the first cause of scurvy." [17]

Set I: Rules 3,4,8 use in their premises the variety of the diet.

```
R3: IF diet-variety ≥ high        THEN disease-severity ≤ 0.   [5]
R4: IF diet-variety ≤ average     THEN disease-severity ≥ 3.   [4]
R8: IF diet-variety ≥ average     THEN disease-severity ≤ 2.   [11]
```

Set II: Rules 7,10 use in their premises the presence (or absence) of fresh fruits and vegetables in the diet.

```
R7: IF fresh_fruits/vegetables = no THEN disease-severity ≥ 2.   [5]
R10 IF fresh_fruits/vegetables = yes THEN disease-severity ≤ 2.  [13]
```

Fig. 4. Excerpts from the 1880 medical encyclopedia followed by the corresponding rules proposed by CHARADE.

forward to predict disease severity. Each group constructed this way can be interpreted as a proto-theory on the cause of the disease.

The first point of interest is the parallel between the rules produced by CHARADE and the competing explanations proposed for the disease until the end of the 19th century and provided in [17]. Every group of rules produced by the system corresponds to an explanation from the medical literature. In Figs. 4–7, the corresponding explanation from history is given just above each group of rules.

The second striking observation is obtained by looking at the number of examples covered by each subset of rules. The three main factors causing scurvy are, according to CHARADE, the presence of fresh fruits and vegetables in the diet, the variety of the diet (closely linked to the previous factor), and the hygiene level. A bar graph plotting the number of examples covered by each proto-theory is given in Fig. 8. The lack of fresh fruits and vegetables covers $13 + 5 = 18$ examples, i.e., more than two thirds of the total. So, THE factor set forth by CHARADE (according to this criterion) is the real cause of scurvy. It is important to remember that these results have been obtained

DIET (food *quantity*):

We are lead to conclude that a decrease in quantity of food, or to speak clearly, starvation, can occasionally serve the cause of scurvy, but it cannot produce it by itself. [17]

Set IV: Rule 2 uses in its premises the quantity of food available.

```
R2: IF food-quantity ≥ ok         THEN disease-severity ≤ 0.   [4]
```

Fig. 5. Excerpts from the 1880 medical encyclopedia followed by the corresponding rules proposed by CHARADE.

HYGIENE:

If Cook's crews were entirely spared from scurvy, in a relatively large extent considering the times, it is thought that these great results were precisely the happy consequence of the care given to the cleanliness and drying of the ships. [17]

Set II: Rules 5,6,9,12 use in their premises the level of hygiene.

```
R5:  IF hygiene ≤ bad          THEN disease-severity ≥ 3.   [3]
R6:  IF hygiene ≤ average      THEN disease-severity ≥ 2.   [4]
R9 : IF hygiene ≥ average      THEN disease-severity ≤ 2.   [7]
R12: IF hygiene ≥ good         THEN disease-severity ≤ 1.   [6]
```

Fig. 6. Excerpts from the 1880 medical encyclopedia followed by the corresponding rules proposed by CHARADE.

CLIMATE:

Spring and winter are obviously the seasons of predominance for scurvy. [17]

Set III: Rules 1, 11 use in their premises the temperature.

```
R1:  IF location = land,
     temperature ≥ hot         THEN disease-severity ≤ 0.   [4]
R11: IF temperature ≤ severe-cold THEN disease-severity ≥ 1. [5]
```

Fig. 7. Excerpts from the 1880 medical encyclopedia followed by the corresponding rules proposed by CHARADE.

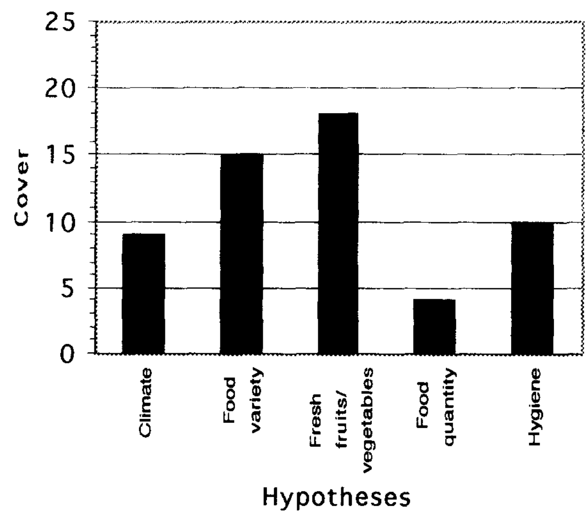

Fig. 8. Absolute covers of causal hypotheses.

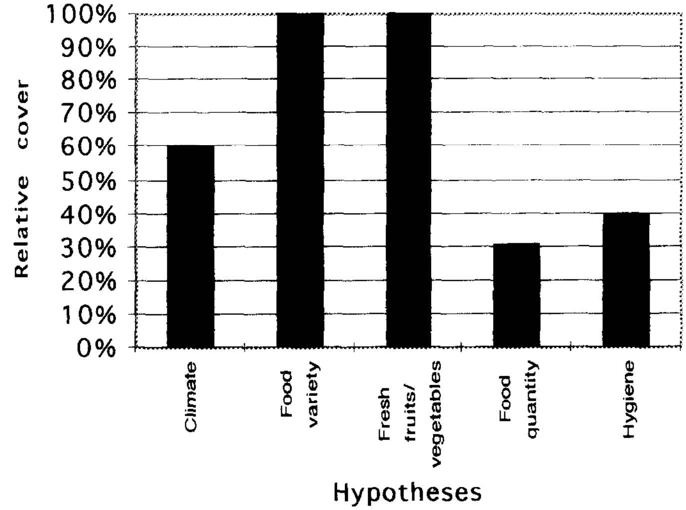

Fig. 9. Relative cover of causal hypotheses.

without any domain knowledge. Nevertheless, in this case, one can say that CHARADE obtains better results than the scientists of the past centuries, in the sense that it puts forward the real cause of scurvy while physicians were misled until the end of the last century toward other theories.

Furthermore, it is useful to realize the existence of a bias in the previous prior-assessment of the various hypotheses. Since the data used for the experiment are real data, the descriptions are not all standardized. Hence, some cases are not described with the same level of detail, some attribute values are missing. A direct consequence is that an hypothesis putting forward a factor cannot explain a case for which this factor is not known. This evaluation procedure is therefore biased in favor of well-informed attributes. An alternative prior-assessment mechanism which does not have this bias is proposed next: it is based on the cover as the previous one, but is defined as a percentage: the ratio of the former cover over the *potential* cover (i.e., the number of examples for which the factor is known). The results of this relative prior-assessment are given in Fig. 9. It can be seen that the hypotheses putting forward the role of the diet (either its variety, or more specifically the lack of fresh fruits and vegetables) stand out even more: 100% of the cases *potentially explainable with these factors* are covered by these two hypotheses.

Though all the rules proposed by CHARADE correspond to some explanations of the scurvy found in the medical literature, there are some explanations from the literature which are not produced by CHARADE. Among these, the most important one is undoubtedly the explanation referring to the humidity of the air as the main predisposing cause of scurvy. This theory was for a long time the most widely accepted, and was defended by people such as Lind. It is also the theory which is favored by the authors of the medical encyclopedia from which the examples are extracted. Therefore, it is surprising, considering the overall similarity, that it does not appear, at least marginally, in the rules produced by CHARADE. The next section attempts to provide a computational explanation for this phenomenon through a simulation.

3.4. The problem of humidity

> *The influence of a cold and humid atmosphere has been said to be the key factor for the apparition of scurvy. "Air humidity is the main predisposing cause of this disease", according to Lind* [17].

CHARADE does not conclude that the humidity of the air has any impact on the presence or absence of the disease, even though this was the most widely accepted theory in the 18th and 19th century. A possible explanation could be that, to reach this conclusion, the scientists had an inductive bias: maybe they had some *a priori* knowledge on the issue which biased their judgment on the origin of the disease.

A good way to test this hypothesis would be to have the inductive system reproduce the "wrong" induction of the scientists by formalizing the implicit knowledge they used while working on the subject. This task is greatly assisted by the work of historians of medicine who have tried to reconstitute the conceptual and reasoning framework of physicians like Lind. In [4] for instance, it appears that the system of *blocked perspiration* was very widely accepted by the medical community at Lind's times (see Fig. 10). In this system, the body is made mainly of solid tissues and fluids. The fluids naturally tend to become corrupted. An important function of all the excretions, and especially of perspiration, is to evacuate these corrupted fluids from the body to keep only healthy fluids inside. If the perspiration is blocked, the corrupted fluids act as a poison and produce diseases. One can fight against the poisonous effect of the corrupted fluids by eating fruits whose acidity acts as a "detergent". This being accepted as a reasoning framework, the explanation of the role of humidity becomes clearer: humidity tends to block the pores of the skin; therefore it prevents good perspiration and is the main cause of scurvy.

The major steps of the reasoning framework in which the blocked perspiration theory could be articulated have been formalized into a set of axioms presented in Fig. 11. As seen in the figure, two new concepts (the perspiration quality, and the fluids quality) had to be added in the description language. Indeed, these two concepts, presented in Fig. 12 are not obtained through direct observations. The axioms state how the two theoretical concepts are logically linked to the observable features according to the theory.

Also, we want to point out that these axioms do not constitute a global theory of scurvy, since they do not express how the disease severity is logically linked to the observable features by a series of deductions. They introduce only the abstract concepts proposed by the blocked perspiration theory. The role of fresh fruits and vegetables appears in the conditions of two rules, because, as noted above, the blocked perspiration theory considered that the acidity from the fruits could have a cleaning effect on corrupted fluids. Also the last two axioms are related to the blocked perspiration theory and express some 19th century common-sense knowledge left implicit in the literature. They concern the link between humidity and the level of hygiene in a ship: one of the main hygiene measures taken aboard ships consisted, especially since Cook, in keeping them dry.

The "Blocked Perspiration" theory

«Lind's theory was based on the concept that a cold, wet climate (and also an unhappy psychological state and inactivity) could result in either a constriction or clogging of the pores in the skin and a consequent reduction in insensible perspiration. The idea that the skin was a major route of excretion of undesirable "vapors and humors" from the body dates back to Galen's time, but was very much developed by Sanctorius in the early 1600's. The idea that obstruction, caused by cold and damp, could result in a variety of putrid diseases became increasingly popular in the mid-eighteenth century and was put forward as the cause of fevers and cholera in military units. Another Edinburgh physician wrote in 1759: "There is no discovery next to that of circulation of blood, that has so much affected our reasoning in Medicine as that of insensible perspiration. The origin of most diseases and the operation of most medicines are accounted for it."

To return to Lind, he began his argument by pointing out that with the uninterrupted circulation of the body's fluids, the friction and mutual interaction with the solid tissues resulted in sweet and healthful components being "abraded and degenerated" into "various degrees of acrimony and corruption." Just as food had to be ingested to replace these components, so had the end products to be excreted. This seems a reasonable statement for someone to make in a period dominated by the success of physical (i.e. mechanical) theories in explaining different natural phenomena. He went on to say that minerals and acid salts we mostly excreted in the urine, but that a greater part of the total excretion was through the skin. He was impressed (we would say overly impressed) by the quantitative experiments that were supposed to have proved this beyond doubt. They involved a subject weighing his food and drink and also his excreta over a period, and also himself at the beginning and end. If the intake weighed more than the urine and feces, together with any gain in body weight, it was said that the excess was lost by perspiration. If the subject had not visibly sweated, this loss was entirely "invisible perspiration." It was probably the quantitative aspect that particularly appealed to Lind, but of course, there was no measure of the carbon dioxide gas and water vapor lost in the air expired from the lungs. [...]»

Fig. 10. The blocked perspiration theory described in [4].

```
IF  (humidity = high)                      THEN (perspiration ≥ hard)
IF  (hygiene ≥ good) (humidity ≤ high)     THEN (perspiration ≤ hard)
IF  (humidity ≥ very-high)                 THEN (perspiration ≥ blocked)
IF  (perspiration ≤ hard)                  THEN (fluids ≤ healthy)
IF  (fresh_fruits/vegetables = yes)        THEN (fluids ≤ healthy)
IF  (fresh_fruits/vegetables <> yes)
    (perspiration ≥ blocked)               THEN (fluids ≥ corrupted)
IF  (hygiene ≤ average)(location = sea)    THEN (humidity ≥ very-high)
IF  (hygiene ≥ good)                       THEN (humidity ≤ high)
```

Fig. 11. Axiom set describing the "blocked perspiration" theory.

Attribute	**Type**	**Domain**
perspiration	Ordered set	normal < hard < blocked
fluids	Ordered set	healthy < corrupted

Fig. 12. Definitions of the two new concepts.

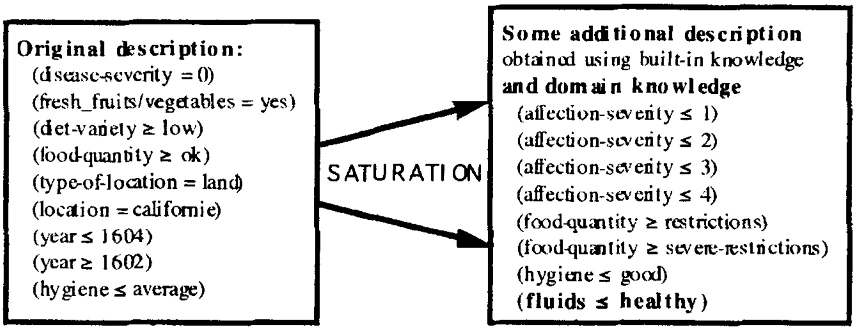

Fig. 13. Saturation on one example with some domain knowledge.

We repeated the previous experiment giving this set of axioms to CHARADE. The aim was to observe the behavior of the system within the conceptual framework of the blocked perspiration theory given as domain knowledge to the system. To illustrate how CHARADE practically uses the domain knowledge, Fig. 13 shows how the saturation is done in this experiment on example #6. It can be compared to Fig. 2 when only general built-in knowledge on the types of attributes was used.

The results confirm our hypothesis about the importance of implicit knowledge. The rules produced (cf. Fig. 14) correspond very well to the explanations given by 18th and 19th century physicians: the humidity appears (in conjunction with other factors) as an important predictor. Moreover, considering the number of examples covered, the rules using the *fluids quality* in their premises override the simpler (and true) rule found in the previous experiment condemning the lack of fresh fruits and vegetables. They indeed cover $14+9=23$ examples out of 25 instead of $13+5=18$ for the rules putting forward fresh fruits and vegetables. This is illustrated by a new plot comparing the cover of the competing hypotheses in Fig. 15. In this experiment, since the new factor (the quality of fluids) has a potential cover of 23, the *relative* cover of the new hypothesis is 100%, which places it at the same level as the one based on diet variety on the relative scale.

This phenomenon has been analyzed within a machine learning framework. It can also be explained in epistemological terms. In other words, the induction bias constituted by the domain knowledge can be seen as a hypothesis (implicit in 19th century writings) about the effect of the environment on some internal functions of the human body. The induction itself is then an attempt to generate some rules from examples using this

```
IF humidity ≥ high
   fresh_fruits/vegetables = unknown,  THEN disease-severity ≥ 2.   [4]
IF humidity ≤ high, hygiene ≥ average  THEN disease-severity ≤ 1.   [6]
IF perspiration ≤ hard                 THEN disease-severity ≤ 1.   [6]
IF fluids ≥ corrupted                  THEN disease-severity ≥ 2.   [9]
IF fluids ≤ healthy                    THEN disease-severity ≤ 2.   [14]
```

Fig. 14. New rules produced when the domain knowledge is given to the system.

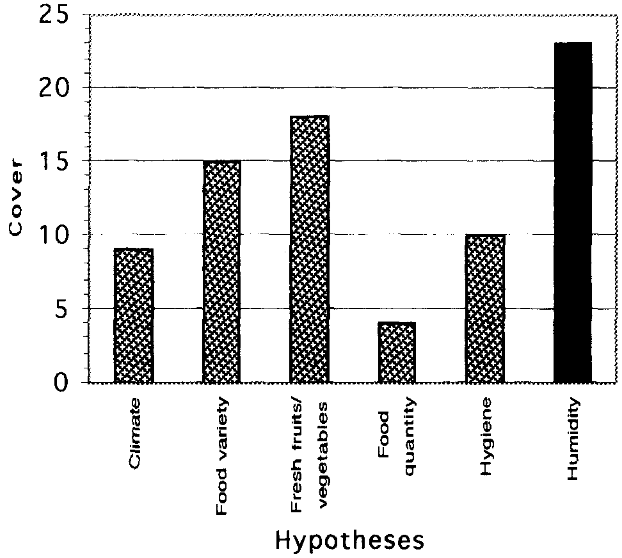

Fig. 15. Absolute cover of causal hypotheses, with a priori knowledge.

hypothesis. Therefore, the rules produced and their evaluation can be seen as a test of the explanatory or descriptive power of the hypothesis. In our case, the fact that the rules which use the abstract concepts cover a larger number of examples than the ones that we found in our first experiment is a sign of their greater explanatory power. We think that this might provide a good explanation for the importance given to humidity over the lack of fresh fruits and vegetables until the end of the nineteenth century. This analysis meets the point of view expressed in [26] about the role of the explanatory power of a theory to explain its acceptance by the scientific community at a given time, given a particular conceptual framework.

4. Conclusion and perspectives

There are two axes to the work just presented. The first one is the use of symbolic induction for discovery. In the case study that we have treated here, pure induction on 25 scurvy cases gave surprisingly good results in bringing forward the real cause of scurvy. However, this pure induction did not produce the theory that was the most widely accepted in the 18th and 19th century. The idea of reproducing these explanations expresses what we pursue in a second, descriptive axis of research: the reconstruction of reasoning as it took place in history. A second experiment in which some domain knowledge was given to the system eventually produced the missing explanation. This shows the importance of taking into account the conceptual framework of the scientists —even in a science which has reputedly relied only on observation since the beginning of the 19th century—to understand their reasoning.

The analysis of experimental results has used two exploratory criteria for the prior-assessment of causal hypotheses. The first one is a straight-forward count of examples covered by consistent subsets of rules. It was useful in bringing forward the real cause of scurvy, and then the cause put forward historically given appropriate *a priori* knowledge. This criterion however showed a bias in favor of well-described attributes. A second relative criterion, less biased by the quality of the data, uses a ratio of the previous criterion over the potential cover of the factor. It is even more efficient in bringing forward the real cause of scurvy, but does not put forward the cause favored in history as effectively, even with the appropriate background knowledge. Therefore, even though this second relative criterion seems more appropriate from a normative standpoint, it appears less useful from a descriptive standpoint: in our experiment, it is the absolute criterion, which does not take into account the quality of the data, that leads to the selection of the same hypothesis as the one selected by physicians in the 18th and 19th century.

In these experiments, the inductive bias seems to have a harmful effect, since it hides the real cause of scurvy. However, it is obviously not our goal to convince the reader that the inductive bias is harmful by nature. We know (especially from machine learning research [18]) that in most cases of induction, even more importantly in cases of human induction, a bias is necessary to constrain the search space. However, a lesson from our research is that it is critical to take into account one's induction bias to understand one's reasoning. In order to do so, it seems necessary to render this bias explicit since it is often the result of implicit knowledge, especially in medicine. It is our belief that computer simulation can be very helpful in this task.

From our experiments on scurvy, it seems that, in one case of a major medical discovery, artificial induction can be useful to reproduce or assist the reasoning of a physician, both normatively, and descriptively. Even though many discoveries in medicine are not the result of such an inductive process (e.g. the ones resulting from a unique observation with a microscope), more experimentation has been carried out on other historical medical discoveries in etiology. We obtained similar results in the simulation of the discovery of the causes of leprosy as an inductive process, though this required using a new inductive algorithm (PASTEUR) designed to model exceptions explicitly [7].

Similar ideas are currently being applied in joint projects with physicians to aid contemporary medical research. The first project involves a collaboration with hematologists working on leukemia research. From patient data collected in a number of US hospitals, we use the same type of computational induction presented here to come up with new hypotheses on the reasons why only some of the patients affected with myelodysplasia (often named pre-leukemia) develop acute leukemia. In this project, we hope that, as for the discovery of the causes of scurvy and the discovery of the causes of leprosy, the limitations in our understanding of the disease are more the results of erroneous reasoning than of a lack of necessary data. If this hypothesis proves true, then computational modeling and simulation of experts' inductive reasoning on leukemia should prove useful to elucidate the dogmas which limit our ability to reason on this particular problem [6]. A look at the recent history of leukemia research [3], or at other areas of cancer research [11] suggests that these dogmas are still numerous. Another project applies this approach to psychiatric research on human depression.

The main issue in moving from historical reconstruction to contemporary research is, apart from the scaling up in the complexity of the problems tackled, the absence of an important resource available in historical work: the analyses performed over time in the field of history of medicine. An exciting research direction is therefore to use the type of modeling and simulation presented in this article as a tool to aid this elicitation process.

Acknowledgements

The early stages of the work reported here owe enormously to a collaboration with the late Professor Marcel Bessis who shared his knowledge and questions about history of medicine and current issues for medical research, and suggested the discovery of the causes of scurvy as candidate for an initial case study. This work benefited from a grant of the Fondation de France. This article was greatly improved thanks to the judicious comments made by the reviewers of its earlier drafts.

References

[1] J.A. Barker, Lind and limeys, *J. Biol. Educ.* **26** (1), (2) (1992) 45–53, 123–129.
[2] C. Bernard, *Introduction à l'Étude de la Médecine Expérimentale* (Champs Flammarion, Paris, 1865, 1984).
[3] J. Bernard, History of concept and dogmas, *Blood Cells* **19** (1994) 549–553.
[4] K.J. Carpenter, *The History of Scurvy and Vitamin C* (Cambridge University Press, Cambridge, 1986).
[5] V. Corruble, Une approche inductive de la découverte en médecine: les cas du scorbut et de la lèpre, PhD Thesis, Université Pierre et Marie Curie, Paris (1996).
[6] V. Corruble and J.G. Ganascia, Aid to discovery in medicine using formal induction techniques, *Blood Cells* **19** (1994) 649–659.
[7] V. Corruble and J.G. Ganascia, Discovery of the causes of leprosy: a computational simulation, in: *Proceedings AAAI-96*, Portland, OR, (1996) 731–736.
[8] M. Foucault, *Naissance de la Clinique* (Presses Universitaires de France, Paris, 1963).
[9] J.G. Ganascia, AGAPE et CHARADE: deux techniques d'apprentissage symbolique appliquées à la construction de base de connaissances, PhD Thesis (Doctorat d'Etat), Université Paris-Sud, Centre d'Orsay (1987).
[10] J.G. Ganascia, Deriving the learning bias from rule properties, in: J.E. Hayes, D. Michie and E. Tyugu, eds., *Machine Intelligence* **12** (Oxford University Press, Oxford, 1991) 151–168.
[11] D. Hellman, Dogma and inquisition in medicine: breast cancer as a case study, *Cancer* **71** (7) (1993).
[12] T.S. Kuhn, *The Essential Tension* (The University of Chicago Press, Chicago, IL, 1977).
[13] P. Langley, G.L. Bradshaw and H.A. Simon, Rediscovering chemistry with the bacon system, in: R.S. Michalski, J.G. Carbonell and T.M. Mitchell, eds., *Machine Learning: An Artificial Intelligence Approach* **1** (Morgan Kaufmann, Los Altos, CA, 1983) 307–330.
[14] P. Langley, H.A. Simon, G.L. Bradshaw and J.M. Zytkow, *Scientific Discovery: Computational Explorations of the Creative Processes* (MIT Press, Cambridge, MA, 1987).
[15] P. Langley and J. Zytkow, Data-driven approaches to empirical discovery, *Artif. Intell.* **40** (1989) 283–312.
[16] J. Lind, *A Treatise of the Scurvy* (Millar, Edinburgh, 1753).
[17] J. Mahé, Le scorbut, in: *Dictionnaire Encyclopédique des Sciences Médicales, Série 3, Tome 8* (Masson, Paris, 1880) 35–257.
[18] T. Mitchell, Generalization as search, *Artif. Intell.* **18** (1982) 203–226.
[19] K. Popper, *The Logic of Scientific Discovery* (Basic Books, New York, 1959).

[20] L.E. Roddis, *James Lind, Founder of Naval Medicine* (William Heinemann, London, 1951).

[21] S. Russel and B. Grosof, Declarative bias: an overview, in: P. Benjamin, ed., *Change of Representation and Inductive Bias* (1990).

[22] K.F. Schaffner, *Discovery and Explanation in Biology and Medicine* (University of Chicago Press, Chicago, IL, 1993).

[23] P. Spirtes, C. Glymour and R. Scheines, *Causation, Prediction, and Search* (Springer, Berlin, 1993).

[24] P. Thagard, *Computational Philosophy of Science* (MIT Press, Cambridge, MA, 1988).

[25] P. Thagard, Ulcers and bacteria I: Discovery and acceptance, *Studies in History and Philosophy of Science* (forthcoming).

[26] P. Thagard and G. Nowak, The conceptual structure of the geological revolution, in: J. Shrager and P. Langley, eds., *Computational Models of Scientific Discovery and Theory Formation* (Morgan Kaufmann, Los Altos, CA, 1990) 27–72.

[27] P. Utgoff, Shift of bias for inductive concept learning, in: R.S. Michalski, J.G. Carbonell and T.M. Mitchell, eds., *Machine Learning: An Artificial Intelligence Approach* **2** (Morgan Kaufmann, Los Altos, CA, 1986) 107–149.