

Reconstructing True Wrong Inductions

Jean-Gabriel Ganascia

■ *There have been many erroneous prescientific and commonsense inductions. We want to understand why people believe in wrong theories. My hypothesis is that mistaken inductions are due not only to the lack of facts, but also to the poor description of existing facts and to implicit knowledge that is transmitted socially. This article presents several experiments the aim of which is to validate this hypothesis by using machine-learning and data-mining techniques to simulate the way people build erroneous theories from observations.*

Why True Wrong Inductions?

Previous attempts to clarify why certain events went wrong, for instance why nuclear plants have burst or why airplanes have crashed, typically included person-based or system-based explanations (Reason 1990). Person-based approaches frequently incriminated “aberrant” mental processes of an individual due to inattention, forgetfulness, negligence, carelessness, or recklessness. System-based approaches implicated social constraints that made people exhausted or led to miscommunication. In a way, all those explanations tend to assume no deliberation; they put the blame either on persons in extraordinary circumstances or on societies. However, past experience has proven that people may make erroneous decisions even when they have goodwill, when they make all the necessary efforts, and when there is no stress, no time pressure, or no social constraints. In some situations, humans appear to be blind to what they see or know; the facts are there, and they just do not take advantage of the empirical evidence. In logical words, inductions, that is, reasoning from particulars, may be wrong not only for psychological, physiological, or sociological reasons but also because implicit knowledge biases our common sense. I postulate here that what went wrong may be the way people induce knowledge from facts, and the causes of those errors are due not only to mistakes or lapses but also to implicit cul-

tural background that might bias inductions. My aim here is to use machine-learning techniques to regenerate erroneous inductions and to highlight the possible causes of wrongness.

This study focuses on the reconstruction of various old inductive theories that have, at least at some point in the past, been recognized as possibly true. Many theories that were based on empirical evidence and that today are recognized as being wrong, such as the theory of “caloric” (heat considered as a substantial fluid) or the theory of “ether” (the medium filling the empty space through which heat and light were supposed to propagate) in ancient physics, seemed very convincing in the past. Clever scholars and scientists have sincerely believed in those theories. One could equally well imagine that most of our present scientific knowledge might be considered erroneous in the future; many currently accepted conceptions may or will be proven false.

The origin of errors in induction is partly due to the lack of information; when a fact is unknown, the theoretical consequences of such a fact cannot be perceived. In addition, the state of the art may render observations difficult. For instance, thanks to the development of optics in the 17th century, Galileo was able to make certain observations in astronomy that were not accessible before. However, even though it is possible to derive a correct theory from a set of empirical evidence, it may happen that only erroneous theories are accepted as

true. This article will try to understand and explain this strange phenomenon, using examples drawn from medicine and commonsense reasoning.

The first reason for such a study is to observe and understand how people actually derive general theories from facts, and not only to consider how they should do it. In the future, developments in cognitive psychology could be used to test the validity of my model. For the moment, I have chosen to deal with prescientific knowledge to try to explain why some misconceptions dominated the world for centuries, even though the available data could have led to more efficient theories than those that were accepted. My work is therefore of epistemological interest. I am also interested in the way people in general, and not only scientists, speculate from facts. This simulation of inexact reasoning could have many applications in the social sciences, where it could help to understand social representations, how they evolve and the way they spread. Finally, this research may also help explain some of the rhetorical strategies used by politicians who, in order to convince, prefer to give well-chosen examples rather than demonstrate their point.

To simulate the way people think and build wrong theories from facts, I have used artificial intelligence techniques such as machine-learning and data-mining tools to automatically reconstruct the pathway leading from the data to the formation of the erroneous theory. The key concept is the notion of *explanatory power* with which all conflicting theories will be compared: this explanatory power evaluates the number of observations that could be explained by a given theory, so each of the different theories generated by an inductive engine will be ranked with respect to this index. However, implicit information related to example description and background knowledge greatly influences the explanatory power. This article investigates the way it leads to misleading conclusions. More precisely, it explores how changing the description language, by adding new features, and modifying the background knowledge, by introducing new inference rules, modifies the explanatory power and, consequently, the ranks of different conflicting theories.

The first part of the article describes the general framework. It introduces the first model based on the use of supervised learning techniques. The second part, titled “Discovering the Cause of Scurvy,” provides an example of rational reconstruction of wrong medical theories using the first model. This is followed by an application to the social sciences, here to model the political beliefs in France at the end of the 19th century, a few months before the Dreyfus affair¹ broke. The model is then extended with a new induction engine using nonsupervised learning techniques. The last part of the article,

titled “Stereotype Extraction,” examines this new model. In the conclusion, I summarize the lessons of experiments presented in the article.

General Framework

Since my interest is focused on the rational reconstruction of inductive reasoning, that is, the derivation of general knowledge from facts, I shall apply inductive machine-learning techniques, that is, those that build general knowledge from facts. Both supervised and nonsupervised learning can be used, each of which has advantages and disadvantages. Although supervised learning procedures are more efficient and easier to program, they require the user to associate a label to each example, which is not always possible as we shall see in the following. In the first and second parts of the article, the scope is restricted to supervised techniques; in the third part, nonsupervised learning techniques will be included.

Sources of Induction

Whatever technique is applied, a description language is always needed; sometimes, additional background knowledge is also necessary. Therefore, the generated theory depends on all this additional knowledge, which biases the learning procedure. In other words, there is no pure induction because the way facts are given to an inductive machine considerably influences the induced theory.

Moreover, many empirical correlations may be observed, which lead to many different possible theories. Since the aim of most machine-learning programs is to build efficient and complete recognition procedures, that is, that recognize all the examples, they tend to preclude most of the possible theories, by using general criteria to prune and eliminate them. For instance, in case of top-down induction of decision trees (TDIDT), information entropy is a very efficient heuristic that makes the generated decision tree quite small and decreases the number of leaves. The goal here is totally different: I want to generate all possible theories and extract explanatory patterns from the results.

More precisely, a set of examples is extracted from historical records (for example, cases of diseases or news item). These examples are formalized with an artificial language to define a training set; the latter is used by association rule-extraction techniques to induce different theories. Then, those theories are ranked with respect to their relative explanation power. This procedure is repeated with new features that enrich the description language and with additional inference rules corresponding to historical implicit knowledge. It appears that the new features and the added inference rules affect both the induced theories and

their respective ranks. We are then looking for feature and inference rules that make the induced theories recreate historical interpretations.

Explanatory Power

As already stated, the key concept here is the notion of explanatory power drawn from Thagard and Nowak (1990): it corresponds to the ratio of the learning set explained by a theory. The inductive engine generates many conflicting theories that can be compared with respect to their explanatory power.

In the case of supervised learning, an example E is said to be *covered* or *explained* by a theory T if and only if the label associated to the example, that is, $\text{class}(E)$, is automatically generated by the theory, which means $T(E) = \text{class}(E)$. Then, $E_p(T)$, the explanatory power of the theory T , is the number of examples belonging to the learning set that are covered by the theory T :

$$E_p(T) = \sum_{E \in \text{Learning Set}} \delta(T(E) = \text{class}(E))$$

where $\delta(\text{true}) = 1$ and $\delta(\text{false}) = 0$.

Association Rules

My experiments make use of the so-called association rule-extraction techniques (Ganascia 1987; Agrawal, Imielinski, and Swami 1993), the goal of which is to detect frequent and useful patterns in databases and then to generate production rules expressing correlations between descriptors. One important point is that, using association rule-extraction techniques, training examples may be covered by many extracted patterns, while it is rarely the case using classical machine-learning techniques. The result is that almost all conflicting hypotheses are extracted, which would not be the case with other inductive techniques.

Discovering the Cause of Scurvy

My first experiment concerns the historic attempt to discover the cause of scurvy and to understand why it took so long to realize that fresh fruit and vegetables could cure the disease. Remember that hundreds of thousands of sailors contracted scurvy and perished in the past. Explanations at the time included a “physical explanation” where the disease was thought to be related to a cold temperature or to humidity, a “physiological explanation” making the lack of food responsible, or even a “psychological explanation” explaining the disease as the result of abstinence and the lack of family life. It was only at the beginning of the 20th century with the discovery of the role of vitamin C that physicians knew how to cure the disease (Carpenter 1986).

I have tried to understand why it was not possi-

ble to induce the correct theory. My starting point was the *Dictionnaire Encyclopédique des Sciences Médicales* (Mahé 1880), which contains relatively precise descriptions of 25 cases of scurvy, and I introduced these descriptions in the inductive engine (Corruble and Ganascia 1997). To be precise, I used a small description language derived from the natural language expressions employed in the medical encyclopedia to describe the 25 cases. This language contained the 10 following attributes: year, location, temperature, humidity, food-quantity, diet-variety, hygiene, type-of-location, fresh-fruit or vegetables, affection-severity, each of them taking one or more values according to its type. In my experiment, I restricted the induction engine so it would generate only those rules ending with the attribute “affection-severity,” which quantifies the evolution of the disease. Note that the same examples may be simultaneously covered by multiple association rules, which render possible the coexistence of different explanatory systems.

Once those rules had been induced, it was possible to distribute them into small subsets according to the attributes present in their premises. Each of these subsets corresponded to a possible explanation of the disease, since it was the set of rules ending with the severity of the disease that contained a given attribute. For instance, the “diet-variety” set corresponded to the theory that explained the evolution of the disease in terms of “diet-variety.” Figure 1 shows the rules generated from the 25 examples of the encyclopedia, classified according to the attributes they contain in their premises. The results showed (see figure 1) that the “best theory,” that is, the theory with the highest explanatory power, was the set of rules containing the attribute “fresh fruit and vegetable” in its premise, since it is the set II of rules that collectively have the highest coverage. Note that the set coverage may be lower than the sum of the rule coverage because it corresponds to the sum of covered examples and double counting is excluded.

The automatically generated explanatory patterns all correspond to some explanation given in the encyclopedia (Mahé 1880). What is more, the explanatory power ranks these five explanatory patterns in the same order of preference expressed by the authors of the medical encyclopedia, the first being the presence of fresh fruit and vegetables in the diet, which is correct considering the present state of our knowledge. But the theory considered as the most plausible explanation of scurvy at the time, that is, the theory of humidity, did not appear once in this list.

This first result supported the role of artificial intelligence: a machine was able to induce the correct theory while people with the same information were not. However, it did not explain why, in

Set I: *diet variety*. [15]
 R3: IF diet-variety \geq high THEN disease-severity \leq 0. [5]
 R4: IF diet-variety \leq average THEN disease-severity \geq 3. [4]
 R8: IF diet-variety \geq average THEN disease-severity \leq 2. [11]
 Set II: *presence (or absence) of fresh fruit and vegetables in the diet*. [18]
 R7: IF fresh_fruit/vegetables = no THEN disease-severity \geq 2. [5]
 R10: IF fresh_fruit/vegetables = yes THEN disease-severity \leq 2. [13]
 Set III: *quantity of food available*. [4]
 R2: IF food-quantity \geq ok THEN disease-severity \leq 0. [4]
 Set IV: *level of hygiene*. [8]
 R5: IF hygiene \leq bad THEN disease-severity \geq 3. [3]
 R6: IF hygiene \leq average THEN disease-severity \geq 2. [4]
 R9: IF hygiene \geq average THEN disease-severity \leq 2. [7]
 R12: IF hygiene \geq good THEN disease-severity \leq 1. [6]
 Set V: *temperature*. [9]
 R1: IF location = land, temperature \geq hot THEN disease-severity \leq 0. [4]
 R11: IF temperature \leq severe-cold THEN disease-severity \geq 1. [5]

Figure 1. Rules Generated without Background Knowledge.

the past, people adopted the theory of humidity to explain scurvy. Because the goal is to model wrong reasoning and the way people reason, I considered this first result insufficient and therefore tried to understand what biased their inductive ability. This entailed looking for some implicit medical theory that could influence induction. I found as a candidate the “blocked perspiration theory” that had been prevalent in medical schools for centuries. This idea was based on the old theory of fluids introduced by Galen (131–201) in the second century and further developed by Santorio Sanctorius (1561–1636) in the early 1600s. According to this hypothesis, without excretions and perspiration the internal body amasses bad humors, which result in fluid corruption and cause diseases. Since humidity and bad hygiene tend to block up the pores of the skin, it makes perspiration difficult and consequently leads to accumulation of bad humors. Furthermore, the lack of fresh fruit and vegetables thickens internal humors, which makes their excretion more difficult.

I translated this theory by introducing production rules that inferred two new features, fluids and perspiration, from existing attributes (see figure 2). Those rules stipulate that the degree of perspira-

tion decreases with humidity and hygiene while the fluids tend to become corrupted when the perspiration turns to be heavy or very heavy. As a result, the inductive engine induced five more rules (see figure 3), in addition to the rules generated previously. Taking these rules into account, it appeared that the rules containing the attribute “humidity” constituted one of the possible explanatory patterns whose explanatory power was higher than that of the other theories.

We have seen here that adding some implicit knowledge during the inductive process may change the results: the theory that appears to prevail without the background knowledge is dominated by another explanation that seems more satisfactory in that it explains more examples than the first one.

This induction bias was caused both by the way the rules were induced, that is, by the induction engine used, which was based on the notion of association rules, and by the lack of information. More precisely, it was mainly due to the incomplete description of the examples. For instance, diet and the presence of fresh fruit and vegetables were not always mentioned. The reason for this was that people only spoke about the facts that

IF humidity = high THEN perspiration \geq hard
 IF hygiene \geq good, humidity \leq high THEN perspiration \leq hard
 IF humidity \geq very-high THEN perspiration \geq blocked
 IF perspiration \leq hard THEN fluids \leq healthy
 IF fresh_fruit/vegetables = yes THEN fluids \leq healthy
 IF fresh_fruit/vegetables $\langle \rangle$ yes, perspiration \geq blocked THEN fluids \geq corrupted
 IF hygiene \leq average, location = sea THEN humidity \geq very-high
 IF hygiene \geq good THEN humidity \leq high

Figure 2. Rules Translating the “Blocked Perspiration” Theory.

Set VI: *fluid theory*. [23]
 IF humidity \geq high, fresh_fruit/vegetables = unknown THEN disease-severity \geq 2. [4]
 IF humidity \leq high, hygiene \geq average THEN disease-severity \leq 1. [6]
 IF perspiration \leq hard THEN disease-severity \leq 1. [6]
 IF fluids \geq corrupted THEN disease-severity \geq 2. [9]
 IF fluids \leq healthy THEN disease-severity \leq 2. [14]

Figure 3. New Rules Produced with Domain Knowledge.

seemed relevant. It would therefore be of interest to compare the way examples are given to some implicit theories, and to see if some example sets are more adequate for a particular theory. My later experiments investigate such a comparison.

Application to the Social Sciences

In order to confront different inductions with different example sets, I have tried to model the way people reason and how preconceived ideas bias political judgements and the interpretation of news items. In this sense, it is an application of artificial intelligence techniques to the social sciences and could help to understand the way people react to

specific cases. In the past, many mathematical and computer science models have been used in sociology, but they have mainly been based on statistical analysis. My perspective is totally different, and the aim is to model the way individuals think and interpret facts with respect to the implicit theories they use. In other words, the aim is to model social representations, that is, social biases.

Additionally, this application is an opportunity to compare the theories that are induced from different data sets and to show how data presentation influences the induced knowledge.

The example taken here is xenophobia in France at the end of the 19th century. I chose the first 10 days of September 1893, a few months before the

Learning from Noise

Martin Eric Mueller

A graduate student with no AI or machine-learning education was given the problem of training a classifier that predicts a radio-frequency identification (RFID) location using field strength data from several antennas. Because the data consisted of long records of real values, the student was advised to use artificial neural networks. After several weeks of producing random classifiers, the student showed up at my office and asked whether I could help. It always seems a good idea to analyze the data first, so we constructed a primitive visualization: signal strength of four antennae over time. The graphs looked like we'd glued a pen on a dog's tail while showing him a juicy T-bone steak. I suggested we add a few functions, such as pair-wise difference, mean, deviation, and so on—just to get a feel for the data.

The image did not change in general; it was like having the dog run over the same picture several times. So I suggested we sort the data points by the actual target location and then see whether the plot changed at all. It did not. Same dog, same steak—this time the dog was just jumping back and forth instead of walking from left to right. In other words, the data simply didn't include any information that could be used to extract knowledge.

It turns out the data had been collected in a building with steel girders whose reflections had reduced it to white noise. The experimenters hired the student because they were unable to learn a Bayesian classifier. People often show a behavior known as confirmation bias: if something doesn't work as expected, try again—just a little bit harder. In this case, it led to a totally unnecessary attempt to extract information from nothing.

Martin Eric Mueller is assistant professor in the Department of Computer Science at the University of Augsburg. Since 1997, he has been researching in machine learning and its applications to adaptive web search and user modeling (for which he earned his Ph.D. in 2001 from the University of Osnabrück) and, since then, to human-computer interaction, context-aware systems, and cognitive robotics.

Dreyfus affair broke. Three daily newspapers, a conservative one, *Le Matin*,² an antisemitic right-wing one, *La Libre Parole*,³ and a Catholic one, *La Croix*,⁴ also conservative, were scanned (Ganascia and Velcin 2004). I collated articles concerning the dysfunctioning of society, including political scandals, corruption, bankruptcies, robberies, and murders. Each article was viewed as an instance and was described using a small representation language similar to that used in the scurvy experiment. This language contains 30 attributes corresponding to the political commitment of the protagonists (socialist, radical, or conservative), their religion, national origin, ethnic origin, whether they are internationally connected, and so on. Sets of articles from each daily newspaper (*Le Matin*, *La Libre Parole*, and *La Croix*) were represented in the same way, using the same description language, but they were considered separately,

each of them constituting a separate learning set.

Note that the target is the class variable *society_dysfunction*, which covers attributes such as political scandals, corruption, incompetence, acts of violence, and so on. As an illustration, figure 4 depicts an induced rule. However, my goal was not only to induce rules and theories, using each of those learning sets, but also to examine the role of four different implicit theories that, according to historians (Taguieff 1998, Bredin 1983), were considered at the time to explain social disorder. These four theories have been drawn from historical studies and correspond to:

- T1—the deterioration of society by an international conspiracy of Jews and Freemasons
- T2—the loss of national traditions and qualities
- T3—the incompetence and inability of politicians
- T4—corruption

IF respect_legislation = no and connection_with_jews = yes
 and connection_with_affairs = yes and political_scandal = yes
 THEN society_dysfunction = yes.

Figure 4. An Induced Rule.

I simplified and translated them into a set of production rules and then looked to see how well each learning set, that is, each set of examples, corresponded to each theory.

My aim here was not only to study the effect of background knowledge on the explanatory power but also to investigate the implicit knowledge underpinning the examples. This is the reason I needed different data sets, which correspond here to different sets of articles from different daily newspapers.

Note that I did not investigate the explanatory patterns by themselves, but the hidden implicit theory or theories underpinning them. People frequently read newspapers with preconceived ideas, and my goal was to identify these ideas that made the paper easier to read. More precisely, for the data sets S1, S2, and S3 corresponding to the three newspapers *Le Matin*, *La Libre Parole*, and *La Croix*, I induced explanatory patterns, first without theory (WT), then with each of the four theories T1, T2, T3, and T4 inserted as background knowledge. I thus obtained 3 X 5 theories $\{S_i, T_j\}$, which were induced from one data set S_i among S1, S2, and S3 with one initial theory T_j , among WT, T1, T2, T3, and T4. For each induced theory $\{S_i, T_j\}$, I computed the explanatory power of all the explanatory patterns and determined the highest value among them.

The results show (see table 1) that the value of the optimal explanatory power reflects the political sympathies of the corresponding newspaper. For instance, corruption and the conspiracy theory have a very high relative explanatory power for *La Libre Parole*, an antisemitic far-right newspaper. On the contrary, the explanatory power of corruption is relatively low for *Le Matin* and *La Croix*, two traditional and conservative newspapers. It means that corruption and the conspiracy theory are implicit for most of the readers of *La Libre Parole*, while neither is implicit for the other two.

Theory/Newspaper	WT	T1	T2	T3	T4
<i>La Croix</i>	25	42	44	55	30
<i>La Libre Parole</i>	38	68	61	38	73
<i>Le Matin</i>	42	55	47	62	40

Table 1. Variation of the Optimal Explanation Power with Different Background Theories.

Incompetence, that is, T3, which had a low value for *La Libre Parole*, seems to explain many examples drawn from *Le Matin* and *La Croix*, even if it is less significant for *La Croix*. Morality, that is, T2, appears to be more explanatory than the conspiracy theory, that is, T1, for *La Croix* while it is the contrary for *Le Matin*. Since *La Croix* is a Roman Catholic newspaper and *Le Matin* just a conservative one, this difference could be easily understandable. For more details concerning this study see Ganascia (2005) and Ganascia and Velcin (2004).

Since it became apparent, when simulating my model on different data sets with different implicit theories, that some data sets can more easily be understood with one implicit theory than with the others, I concluded that different data sets incline to different interpretations. Since those implicit theories were directly related to the political sympathies of the daily newspapers from which the examples were taken, it validates my model. In other words, it explains how examples induce misrepresentations. Even if none of the examples is false, the way they are represented, the lack of description, and the presence of implicit knowledge may influence the induction considerably.

Since this phenomenon appeared to be crucial in commonsense induction, that is, in the way people derive knowledge from personal experience, I tried to model and to generalize it in a logical framework. The next section presents this logical framework.

Stereotype Extraction

The notion of stereotype was introduced by Walter Lippmann in his famous book *Public Opinion* (1922) to characterize the way partial information is crystallized in our mind. Lippmann says that each of us builds stereotype folders from partial information we gather through family discussions, school, newspapers, TV, rumors, and so on. These stereotypes then filter information and help to form opinions concerning public events about which we have in general no precise knowledge.

According to Lippmann's hypothesis, stereotypes are constructed from poorly described data, the descriptions of which are mainly implicit. Therefore, stereotype learning could be seen as a case of unsupervised learning from sparsely described data.

To formalize this idea we have developed an algorithm that learns from very sparsely described data (Velcin and Ganascia 2005). The idea is that each piece of information, a news item for instance, corresponds to a fragment of a stereotype that a learning algorithm would be able to rebuild. This algorithm finds a set of full descriptions that minimizes a cost function, which corresponds to the sum of the distances between learning set examples and their nearest stereotype. In other words the cost function h may be defined as follows:

$$h(E, S = (s_1, s_2, \dots, s_n)) = \sum_{e \in E} D_s(e, C_s(e))$$

where E is the learning set, S is the set of stereotypes, $C_s(e)$ the stereotype of S that is the closest to e and $D_s(e, e')$ the distance between e and e' . Note that the learning set examples e are supposed to be sparsely described while stereotypes S have to be full descriptions, which prohibits a data overfitting.

Newspaper Stereotypes

My last experiment involves extracting sets of stereotypes from news items taken from each of the three newspapers mentioned earlier and interpreting them with respect to the political sympathies of the readers. Depending on the newspaper, the results are quite different. For instance, the news from *La Libre Parole*, which is a far-right newspaper, generated two stereotypes, one of which covers 90 percent of the initial examples. Moreover, it appears that only 4 percent of the

examples are not covered by any of the constructed stereotypes. The news taken from *Le Matin*, a moderate conservative newspaper, generated three stereotypes that are far more balanced, while 16 percent of the examples are not covered by any of the stereotypes. In contrast to *Le Matin*, *La Libre Parole* appears far more dogmatic.

Let us now consider the descriptions of the generated stereotypes. The main stereotype of *La Libre Parole* corresponds to a man who is socialist, internationalist, antipatriotic, has connections with Jews and Protestants, is corrupt, anticlerical, involved in freemasonry, and is immoral. The second stereotype, which covers only 6 percent of the examples, corresponds to a Catholic who is involved in freemasonry. Of the three stereotypes generated from *Le Matin*, the first corresponds to a socialist who is involved in freemasonry, is anticlerical, a traitor to the nation, all of which corresponds to the dominant stereotype of *La Libre Parole*. However, the second and third stereotypes are quite different: the second corresponds to an opportunistic politician who is republican and incompetent, while the third evokes health problems that affected the French president at this time.

Briefly speaking, we built an analysis tool that takes as input a set of news items and that outputs the implicit stereotypes conveyed by those news items. More precisely, to be understandable, news items refer to stereotypes shared by the readers while, simultaneously, the way the information is given reinforces the stereotypes that readers have in mind. I claim that the stereotype extraction process may help to make the stereotypes conveyed by the newspapers explicit.

Conclusion

The aim of this article was to try to understand why we adopt wrong theories even when they are contradicted by empirical evidence. Machine-learning and data-mining inductions based on various data sets can be used to identify different causes of wrongness. The first is related to the language used to describe examples, that is, to the set of categories in which we classify and describe factual evidence. The second concerns the background knowledge, and corresponds to the hidden implicit theories that underpin possible conceptualization. The third is the incomplete description of facts. This was the case in the experiments presented in this article: the cause of scurvy and xenophobia in France at the end of the 19th century. In all cases, it appears that example descriptions were very sparse, which made different interpretations possible. For instance, in the scurvy example, diet was not always explicitly mentioned in the description of all the case studies. This is why, giv-

en the prevailing blocked perspiration theory, the explanatory power of the humidity attribute passes above the explanatory power of attributes relative to the presence of fruit and vegetables in the diet. Lastly, the news items that are published in newspapers may influence the reader and contribute to the building of certain specific stereotypes.

More generally, the article endeavors to elucidate, with the use of AI techniques, one particular cause of wrongness, that is, erroneous induction. Other works elicit what makes some people successful at a given time while others, or the same people at different times, fail. For instance Dörner's *Logic of Failure* (Dörner 1996) observes behaviors of individuals confronted with complex tasks, for example, playing SimCity, and extracts psychological presages of success or failure. According to Dörner, it appears that self-confidence in a priori theories is responsible for many failures. My purpose here was to show why a priori theories, that is, preconceptions and mental stereotypes due to education or culture, are not only misleading because they are erroneous; they also make people unable to interpret new contradictory facts; in other words, they make people blind to the outside world. This is not only of theoretical interest; it might help prevent errors and wrongness.

In conclusion, I can offer a word of warning about induction in daily life. We must bear in mind the influence of culture on the categories we use to understand examples and the impact of education on our background knowledge. Furthermore, our mental stereotypes bias our perception of reality. Lastly, our personal experience of life determines the very types of examples that we consider.

Note

1. The Dreyfus affair was a political scandal that divided France at the end of the 19th century, when a young Jewish officer was accused of treason because of his ethnic origins.
2. *Le Matin*, daily newspaper from September 1, 1893, to September 10, 1893.
3. *La Libre Parole*, daily newspaper from September 1, 1893, to September 7, 1893.
4. *La Croix*, daily newspaper from September 1, 1893, to September 10, 1893.

References

Agrawal R.; Imielinski, T.; and Swami, A. 1993. Mining Associations between Sets of Items in Massive Databases. In *Proceedings of the ACM-SIGMOD 1993 International Conference on Management of Data*, Washington DC, 207–216. New York: Association for Computing Machinery.

Bredin, J.-D. 1983. *L'affaire*. Paris: Julliard.

Carpenter, K. J. 1986. *The History of Scurvy and Vitamin C*. Cambridge: Cambridge University Press, 1986.

Corruble, V., and Ganascia, J.-G. 1997. Induction and the Discovery of the Causes of Scurvy: A Computational Reconstruction. *Artificial Intelligence Journal* (91)2: 205–223.

Dörner, D. 1996. *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*, translated by Rita and Robert Kimber. New York: Metropolitan Books.

Ganascia, J.-G., 2005. Rational Reconstruction of Wrong Theories. In *Logic, Methodology, and Philosophy of Science*, ed. Petr Hajek, Luis Valdes-Villanueva, and Dag Westerstaahl. London: College Publications.

Ganascia, J.-G. 1987. CHARADE: A Rule System Learning System. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, Milan, Italy. San Francisco: Morgan Kaufmann Publishers.

Ganascia J-G, and Velcin J. 2004, Clustering of Conceptual Graphs with Sparse Data, in *Proceedings of the Twelfth International Conference on Conceptual Structures*, 156-169, LNAI 3127. Berlin: Springer.

Lippmann W., 1922, *Public Opinion*, Free Press; Reissue edition June 1997. First published 1922

Mahé, J. 1880. "Le Scorbut" (in French). In *Dictionnaire Encyclopédique des Sciences Médicales*, Série 3, Tome 8, 35–257. Paris: Masson.

Reason, J. 1990. *Human Error*. New York: Cambridge University Press.

Taguieff, P.-A., 1998, *La couleur et le sang. Doctrines racistes à la française*, Éditions mille et une nuits, janvier 1998, « Essai ».

Thagard, P., and Nowak, G. 1990. The Conceptual Structure of the Geological Revolution. In *Computational Models of Scientific Discovery and Theory Formation*, 27–72, ed. J. Shrager and P. Langley. San Francisco: Morgan Kaufmann, Publishers.

Velcin, J., and Ganascia, J.-G. 2005. Stereotype Extraction with Default Clustering. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, 883–888. Denver, CO: Professional Book Center.

Jean-Gabriel Ganascia first studied mathematics and physics to be an engineer. Then he graduated in physics from Orsay University and got a DEA in acoustics (Paris VI University). In parallel, he studied philosophy (obtaining a licence de Philosophie from Université Paris I [Sorbonne]) and computer science (with a DEA from Paris VI University). He obtained a grant to prepare a doctorate on knowledge-based systems applied to geology. He earned his Doctorat d'ingénieur in 1983. After that, he pursued research on machine learning from both a theoretical view and a practical one until he obtained his Thèse d'état in 1987. Ganascia was successively named assistant professor at Orsay University (Paris XI) (1982), Maître de conférence at Orsay University (1987), and professor at Paris VI University (1988). He was also program leader in the CNRS executive from November 1988 to April 1992 before moving to direct the Cognitive Science Coordinated Research Program and head the Cognition Sciences Scientific Interest Group from January 1993 until 2000. He is presently a professor of computer science at Paris VI University where he leads the ACASA team in the LIP6 laboratory (Laboratoire d'Informatique de Paris VI) .



Journal of Artificial Intelligence Research

Volume 30, 2007

Toby Walsh
Editor-in-Chief
Adnan Darwiche
Associate Editor-in-Chief

Subscriptions for the
2007 and 2008
volumes of *JAIR*
are now available!

For details, see
www.aaai.org/Press/Journals/