ELSEVIER

Artificial Intelligence 107 (1999) 335–346

Artificial Intelligence

Research Note

Principles of human–computer collaboration for knowledge discovery in science

Raúl E. Valdés-Pérez [1]

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 9 June 1998; received in revised form 8 December 1998

Abstract

An important problem in computational scientific discovery is to identify, among the diversity of discovery programs written in various sciences, a commonality that will take a next step beyond the acknowledged general—but weak—framework of heuristic search.

We characterize discovery in science as the generation of novel, interesting, plausible, and intelligible knowledge about the objects of study. We then analyze four current machine discovery programs in chemistry, medicine, mathematics, and linguistics according to how their design, or the circumstances of their application, heighten the chances of finding knowledge that has all four properties. Some general patterns emerge, although some strategies seem idiosyncratic.

Our candidate for a commonality, which focuses on human factors, can be used pragmatically to evaluate and compare the designs of discovery programs that are intended to be used as collaborators by scientists.

Keywords: Scientific discovery; Human–computer collaboration; Knowledge discovery

1. Introduction

Early work on machine scientific discovery such as Logic Theorist [25], DENDRAL [23], and AM [22], and later work on the cognitive modelling of historical discovery (e.g., [21]), have shown that heuristic search in combinatorial spaces is a useful, general framework for automating and explaining discovery. However, it has been unclear what further generality could be found among programs that accomplish diverse tasks in different sciences. Absent general principles, each discovery task must be tackled from scratch, but it is better to transfer knowledge from one design experience to the next.

[1] Email: valdes@cs.cmu.edu; web: www.cs.cmu.edu/~sci-disc.

0004-3702/99/$ – see front matter
PII: S0004-3702(98)00116-7

We propose human–computer collaboration as a source of generality. We will first characterize discovery in science (the activity of scientists, not the AI subfield) in terms of four dimensions that reflect the elementary goals of science. Then we will analyze four current programs that have made contributions to the scientific literature, with respect to these four dimensions. Some patterns will emerge, and also qualitative methods to evaluate, report, and compare collaborative discovery programs.

2. Knowledge discovery and discovery in science

There has been great recent interest in knowledge discovery in databases (KDD) for reasons that include the increasing automatic capture of data. The various definitions of KDD are similar, e.g.,

> "Knowledge discovery in databases is the nontrivial process of identifying valid, novel, potentially useful, and ultimately *understandable* patterns in data." [11]
>
> "*Knowledge discovery* is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data." [13]

These are a good starting point for characterizing discovery in science, but several changes in emphasis are needed due to the nature of the domain of science.

First, we prefer to remove the reference to *nontrivial processes* because, in our view, the importance of a discovery is separate from the processes that generated it; simpler, even trivial, processes are better than complicated ones if they are enough for the job of discovery [32].

Second, we prefer the term *plausible* over the term *valid* because the latter connotes *certain* inference in fields such as formal logic. The many failed attempts to solve Hume's problem of induction [29] in philosophy of science show that the best one can expect is to induce plausible knowledge; only deductive, non-ampliative inference yields "valid" (certain) results, but these do not go beyond the limited content represented by the premises or the data.

Third, the major traditional value in science is not *usefulness* but *interestingness*, which is often linked to more specific values like generality or simplicity.

Fourth, and most important, we liberalize the exclusive focus on *data* to encompass the *objects of study*, of which plain data are a special case. There are many knowledge-driven (also known as theory-driven) tasks in science; discounting them will lead to a highly incomplete understanding of discovery.

We preserve the reference to *novel*, but there is more to add. For example, replications can contain (second-class) novelty if they make previous discoveries more credible. Also, meta-statements that are idiosyncratic to machine discovery programs also provide novelty, e.g., the meta-statement that there exist no simpler solutions within a given space of hypotheses.

Thus, we characterize discovery in science as follows:

> Discovery in science is the generation of novel, interesting, plausible, and intelligible knowledge about the objects of study.

This description[2] has the flexibility to cover the many types of contribution that are publishable knowledge in science, e.g., models, conjectures, theorems, patterns, causes, rules, typologies, taxonomies, descriptions, refutations, confirmations, and so on.

These four dimensions can vary independently, as is seen by forming their combinations, e.g., the number of blades of grass viewable from one's office window can be novel, plausible, and intelligible but also dull. However, their relative independence does not mean that the dimensions are wholly unrelated. For example, both interestingness and intelligibility can be enhanced by simplicity or undermined by complexity. The point is that scientists can regard some statements as intelligible but dull, and others as interesting but obscure, i.e., hard to relate to anything that scientists know or could try to find out.

Clearly, the criteria for *interestingness* (or intelligibility) vary greatly, even within one science over time. So, to achieve some generality, it is best to consider (pragmatically) that results are interesting in science X if practitioners of X (or users of the programs) believe them to be interesting. One then seeks patterns among the factors that enable these domain-dependent judgments.

The next sections project four successful discovery programs along the dimensions of novelty, interestingness, plausibility, and intelligibility. Notably, none of the cited articles uses these dimensions overtly to describe the programs. Our analysis will suggest a working hypothesis about successful applications of machine discovery. Finally, we will advocate that any discovery program should be qualitatively evaluated along these dimensions.

3. Why some programs do science well

Our characterization of discovery in science suggests the following claim:

> Any program that with excessive frequency leads to knowledge that is familiar, dull, wrong, or obscure will not long be tolerated by its users and hence will fail as a collaborator.

As a corollary, a key question about any program is:

> How does the design of the program, or the circumstances of its application, heighten the chances that its use will lead to knowledge that is novel, interesting, plausible, and intelligible?

There is a small but appreciable number of contemporary programs that do science well. We have chosen four that have enabled published discoveries in their respective literatures:

- MECHEM—hypothesizes reaction mechanisms in chemistry based on the available experimental evidence [2,41,42,51,52].
- ARROWSMITH—notices connections between drugs or dietary factors and diseases in medicine [35].

[2] Note that we use the term *discovery in science* to characterize not the AI subfield of *scientific discovery*, but instead the activity of research scientists. Also, discovery that deals with *new* objects of study are not excluded if *objects* is construed broadly, e.g., the Earth is the object of geological study.

(1)	Given starting materials CH_3–CH_3 and CD_3–CD_3, the reaction does not yield HD
(2)	Given a starting material CD_3–CH_3, the reaction does not yield HD
(3)	Reactants involve jointly at most two carbon atoms
(4)	Every conjectured species must appear on left- and right-hand sides of some (different) steps in the pathway
(5)	Reject mechanisms whose only intermediate species is CH_3–CH
(6)	Reject the step: $H_2 + CH_3$–$CH \longrightarrow CH_4 + CH_2$

Fig. 1. Ethane photolysis constraints.

- GRAFFITI—makes conjectures in graph theory and similar fields [9,10].
- MPD/KINSHIP—profiles the classes within a classification, e.g., in linguistics [26–28,46].

All four programs can best be described as carrying out a heuristic search over a large combinatorial space. Also, all the programs accomplish tasks that predate the involvement of modern computers, which contrasts them from data-intensive problems in, say, molecular biology in which the sovereignty of programs may be conceded from the start. Other programs have done science well (e.g., [24,30,33,44,47,48]), but this list suffices for our purposes. There are also exploratory programs that have illuminated aspects of scientific reasoning without yet serving as collaborators (e.g., [1,4,5,7,18,21,22,36,53]); these are outside the scope of our analysis.

Each program handles the four dimensions in multiple ways, but we will usually mention only one way for each program/dimension combination. Our analysis will be *a posteriori*: we examine systems that are successful as determined by publication in the scientific literature, and then project them along the dimensions. Alternative, *a priori* approaches are, e.g., studies of interestingness based on statistics [17,19] or on general domain-free notions like unexpectedness and actionability [31].

3.1. MECHEM

MECHEM finds explanatory hypotheses (reaction mechanisms) in chemistry. That is, given the starting materials of a chemical reaction, any observed products and intermediates, and prior background knowledge expressed as constraints, the program finds all simplest mechanistic hypotheses that explain how the products are formed while respecting the constraints. For example [50], given that the starting material CH_3–CH_3 (ethane) forms H_2, CH_2=CH_2 (ethylene), and CH_4 (methane), and subject to the constraints in Fig. 1, MECHEM can generate from scratch hypotheses like:

(1) CH_3–$CH_3 \longrightarrow H_2 + CH_3$–$CH$
(2) CH_3–$CH_3 \longrightarrow CH_4 + CH_2$
(3) $2(CH_2) \longrightarrow CH_2$=$CH_2$
(4) CH_3–$CH \longrightarrow 2(CH_2)$

The reader is not expected to understand the chemistry; the aim is only to give some notion of the types of prior knowledge and of the form that hypotheses take. MECHEM has a graphical interface that allows the chemist to express over 100 types of prior knowledge, and this list grows with the demands posed by new applications.

MECHEM's output tends to contain *novelty* because the pieces (elementary reactions and chemical substances) that make up a hypothesis (mechanism) are not drawn from a catalogue of common reactions; rather, they are generated from scratch using algorithms [37–39] that are minimally slanted toward particular solutions. Thus, in a sense MECHEM embodies a new representation (i.e., explores a new problem space) [43].

A second reason for novelty is that, as shown by extensive experience (e.g., [41,42, 52]) and by methodological arguments [50,51], a comprehensive search for all simplest hypotheses regularly reveals possibilities that tend to be overlooked by human scientists, who are not equipped to do massive searches of a combinatorial space. In a recent application [2], MECHEM generated 41 mechanisms consistent with prior knowledge which were experimentally tested and thus reduced to a handful which could not currently be discriminated. Finally, MECHEM enables novel *meta-statements* such as "all the simplest possible hypotheses are these" which are not feasible without comprehensive computerized searches.

The output mechanisms tend to be *interesting* because they are simple: the search is carried out in stages of simplicity [40], which minimize the number of individual reaction steps and the number of hypothesized (i.e., not given as input) chemical substances, such as reaction intermediates.

Plausibility is ensured because after inspecting an output, the user can articulate his objections, based on background knowledge, in the form of constraints and re-run the program. This interaction can continue until no more objections remain, at which time the surviving hypotheses can be considered plausible.

Finally, the hypotheses are *intelligible* because by design the program searches the conventional hypothesis space in mechanistic chemistry which was established around a century ago. The mechanism above is of this conventional form.

3.2. ARROWSMITH

ARROWSMITH makes conjectures about possible treatments or causes of medical diseases using the MEDLINE literature as a knowledge source (see [35] and http://kiwi.uchicago.edu). Given a disease or other physiological state A, the program searches for two associations AB and BC where C is typically a dietary factor, drug, or other possible intervention.

For example, the user may pose A = *migraine*, and the program may come up with B = *spreading depression* and C = *magnesium* (a light metal which is essential to the human diet). After subsequent human examination of the complementary but disjoint subliteratures, which report that "magnesium can inhibit *spreading depression* in the cortex, and *spreading depression* may be implicated in migraine attacks" [35, p. 185] there is the plausible suggestion that magnesium could be a treatment for migraine.

Swanson et al. have reported eight successful matchings of complementary but disjoint literatures, four of these in collaboration with Neil R. Smalheiser, a neurobiologist. The

best confirmed connection is between magnesium deficiency and migraine headaches [34]. After that publication, more than 12 laboratories have reported confirmatory clinical or laboratory tests [35].

ARROWSMITH's conjectures tend to be *novel* because the methodology involves a citation analysis to verify that no (or few) MEDLINE articles cite both subliteratures responsible for the associations AB and BC, so that there is no evidence from MEDLINE that anyone has noticed these connections. Swanson's colorful term for the program's output is "undiscovered public knowledge".

The *interestingness* of the conjectures are enhanced by heuristics that filter out C factors that are too broad to be significant. For example, Swanson cites *hormone, pressure, lipid,* and *membrane* as overly broad single words, which are removed by means of a stoplist unless they are in an interesting phrase.

The conjectures tend to be *plausible* because many of the associations are causal, and causal relations tend to be transitive. Even if the associations are based on similarity and not causality, the shortness of the path length $C \rightsquigarrow B \rightsquigarrow A$ lends plausibility to the conjecture $C \rightsquigarrow A$.

Finally, ARROWSMITH's outputs are *intelligible* because they tend to be straightforward conjectures like *C may be a treatment for, or a cause of, A* which can be clinically tested, at least in principle.

3.3. GRAFFITI

GRAFFITI [9,10] makes mathematical conjectures in domains such as graph theory or geometry. In graph theory, the program generates members within a space of conjectures $\sum x_i \geqslant \sum y_j$, where the terms x_i and y_j are numerical features of a graph (called invariants in graph theory) such as its diameter, its largest eigenvalue, and so on. The terms can also be quotients or products of the elementary features.

GRAFFITI has motivated many graph theoreticians, including its designer, to try to refute or prove its conjectures which are broadcast on an email list. Various of the conjectures have been proven (by mathematicians) and published as regular mathematical contributions, for example, the conjecture that the independence number of every connected graph $\geqslant$ the average distance between its vertices [3].

The program[3] keeps a database of previous conjectures so that the program does not repeat itself and instead will produce *unseen* conjectures. However, GRAFFITI has no easy way to make direct use of the mathematical literature, unlike the ARROWSMITH program which is used in conjunction with citation analyses of MEDLINE.

The program's "Echo heuristic" [10] rejects less *interesting* conjectures by testing whether a conjecture is seemingly implied by a previous conjecture that has not been refuted. For example, a conjecture $x > y$ is stronger, and hence more interesting, than a conjecture $x \geqslant y$, because the former implies the latter.

Every conjecture is tested against a file of qualitatively different graphs and survives this *plausibility* test only if no counterexample is present. If a surviving conjecture is later

[3] There have been multiple versions of GRAFFITI which use different heuristics. We have selected from among these heuristics without requiring that they all be found in the most current version of the program.

refuted by a new graph counterexample, the graph can be added to the file so that over time the conjectures' plausibility should grow.

Finally, GRAFFITI's conjectures are *intelligible* because they are expressions of the form *a short sum of graph properties is* $\leqslant$ *another short sum of graph properties*. It seems possible, as an illustrative contrast, that conjectures that made arbitrary use of more complicated functional forms (e.g., $x_i^{y_j}$) would be less understandable by human mathematicians and less likely to merit consideration.

3.4. MPD/KINSHIP

MPD finds concise profiles of all the classes within a classification, given at least one example of each class and features (numeric, symbolic, or mixed) that describe the examples. Thus, the program's input is the same as the input to multiclass, supervised concept learning programs [8].

The output of MPD is not a decision tree nor set of classification rules, but rather (1) a guaranteed-minimal list L of features that enable all N-choose-2 pairwise contrasts among the N classes, and (2) individual profiles for each class C, where a profile is a subset of L together with a quantitative statement of which other classes are contrasted from C by which feature in the subset. Partial contrasts are allowed if the overlap between feature values falls below a user-set maximum.

MPD is a data-driven program with general scope [45,46], but so far it has been applied mostly to linguistics [28]. Thus, we refer specifically to the KINSHIP program [27], which is just MPD with a front-end that computes linguistic features of the examples from the raw data. The KINSHIP program automates the task of componential analysis [15] in linguistic anthropology, which seeks concise descriptions of the pairwise contrasts among the members of a set of lexical terms.

Part of KINSHIP's input is a list of the kinship terms in a language. For example, Yankee English has 35 terms [27] such as *son, daughter, uncle, cousin, grandfather, mother, wife, step-father, brother-in-law*, etc. The rest of the input is a listing of the alternative sequences of genealogical/matrimonial relations that correspond to a kinship term. For example, in English an *uncle* can be a mother's brother, a mother's sister's husband, and so on.

More formally, the program's input is a set of kinship terms (i.e., classes), a set of examples for each class, and a description of each example in terms of a sequence of matrimonial (e.g., wife) or blood relations (e.g., son). Over 20 linguistic features are then computed from these sequences. Then MPD outputs a minimal overall list of features plus a profile for each kinship term, such as this profile for "brother": *generation* = 0, *sex* = *male*, *generation-of-last-link* = 0, and *affinity* = *consanguineal*, which distinguishes "brother" from the other 34 English kinship terms.

KINSHIP found a *novel* simplest analysis of Yankee English and of an unanalyzed language (Bulgarian). Like MECHEM, it searches a large combinatorial space which is potentially dense with solutions that are easy to overlook without a comprehensive search. Also like MECHEM, it enables meta-statements of the form "and no other simpler solutions exist".

The program's outputs are *interesting* because they are maximally concise, i.e., the program guarantees a minimal use of overall features, as well as minimal individual profiles of the kinship terms.

Plausibility can be interpreted in two ways. One view is that KINSHIP (and MPD) merely find concise descriptions of the available data, so that plausibility is irrelevant. However, the methods can in some cases be used inductively, and like all data-driven programs, plausibility is heightened by an abundance of input data, as long as some elementary statistical checks are applied [14].

KINSHIP's output is *intelligible* because, as with MECHEM, the solution space of kinship descriptions is a conventional one taken from the anthropological linguistics literature, although it is not the only one, since alternative description spaces have been proposed (noted in [27]), for example, one that takes into account the effects of morphemes like "grand" in "grandfather" and "in-law" (e.g., sister-in-law) in English.

3.5. General patterns

MECHEM and KINSHIP both address *novelty* by relying on comprehensive searches to turn up possibilities that scientists would likely overlook. Also, comprehensive searches enable novel meta-statements like "these are all the simplest solutions" which scientists find valuable. ARROWSMITH treats novelty by the unique device of citation analysis. GRAFFITI, which is a conjecture generator like ARROWSMITH, could in principle do likewise, but there is no practical way to automatically decide whether a conjecture is without precedent in the mathematical literature. One GRAFFITI heuristic for addressing novelty is simply to remember (and not repeat) all of its previous conjectures.

Each program except GRAFFITI derives for free some degree of *interestingness* from the circumstance that a user poses the initial problem which presumably is of interest. Both MECHEM and KINSHIP rely on the simplicity of their outputs to ensure interestingness. GRAFFITI prefers stronger or more general statements, i.e., it prefers the stronger conjecture A over the weaker conjecture B in the implication $A \rightarrow B$, e.g., $x > y \rightarrow x \geqslant y$. Conversely, ARROWSMITH prefers *less general* conjectures, since it rejects broad terms like "hormone" in favor of more specific physiological factors. One reason for the discrepancy may be that acceptable medical conjectures should correspond to direct experimental tests, and it is easier to test conjectures about specific hormones than about hormones in general.

MECHEM, which is a highly knowledge-driven program, handles *plausibility* by means of a multi-year knowledge engineering effort to let users input constraints that express prior knowledge about a chemical reaction. ARROWSMITH relies on the transitivity of associations based on causality or simple similarity. GRAFFITI tests its conjectures against a database of graphs which are potential counterexamples, whereas KINSHIP relies on its data-driven nature, i.e., abundant data leads to plausible descriptions or inductions. Thus, GRAFFITI and KINSHIP both use empirical methods, although GRAFFITI uses hypothesize-and-test whereas KINSHIP directly uses the data to derive its output.

Finally, all four programs address *intelligibility* similarly: their outputs all lie within different solution spaces that are conventional (but not necessarily unique) for their respective domains.

So what general patterns emerge? Some strategies, such as ARROWSMITH's use of citation analysis, are powerful but idiosyncratic. However, we can discern these general patterns:

(1) People rarely carry out comprehensive combinatorial searches, so programs that do so will often turn up novelty in spaces that are dense with solutions. This has been demonstrated convincingly in chess play, but it is no less true for discovery in science.
(2) Comprehensive searches of a problem space also enable novel meta-statements like "these are all the simplest solutions" which scientists find valuable.
(3) Simplicity is a frequent guarantor of interestingness, especially in tasks of model building [49], where it corresponds to minimizing the number of components, processes, and other pieces that compose a model.
(4) Generality enhances interestingness in sciences that have a formal character, whereas in experimental sciences like medicine, specificity—although in principle less interesting—may enhance acceptability because of the greater ease of experimental test.
(5) When the user poses the problem, some interestingness is inherited for free.
(6) The use of abundant data ensures some plausibility if elementary statistical checks are used.
(7) Knowledge-driven programs should include a (likely substantial) knowledge engineering effort so that users can express their prior knowledge, which will vary if the program has any generality.
(8) Programs that generate solutions from a conventional space for the scientific task are guaranteed to be intelligible. Often these spaces are not good matches to the spaces targeted by domain-independent programs not written specifically for discovery in science [20].
(9) Intelligibility and novelty can be conflicting goals whenever a program involves a new representation of a task, i.e., when it searches a novel problem space.

These design patterns complement the sociotechnical guidelines in [6], which constructs a futuristic scenario of human–computer collaboration and reason backwards about what is needed to enable such scenarios. Our analysis is based not on social systems but on an individual scientist/program collaboration.

3.6. Other programs

It is instructive to consider the two classic scientific discovery programs DENDRAL and AM. DENDRAL has been described as a scientific success [12] and a "failure" in terms of its adoption by external chemist users [24]. AM was an exploratory project that never seriously sought mathematician users. Both programs were successful early demonstrations of the usefulness of heuristic search for explaining or automating discovery in science.

Our close reading of the DENDRAL literature [16,23] reveals no clear failure to handle our four dimensions. DENDRAL and MECHEM address these dimensions similarly; the main difference is that in MECHEM simplicity is crucial for preferring hypotheses that have fewer substances and steps. DENDRAL's failure to be adopted by working structural chemists has been much speculated upon; one factor pointed out in both [16] and [24] is that the task did not arise often enough to justify the learning effort. Clearly, programs can fail to be adopted for reasons other than those contained in our four dimensions.

AM [22] never received a serious application effort, so its analysis must be more speculative. We surmise that one obstacle to its adoption would have been how to ensure occasional novelty, since it could not make use of the comprehensive-search-by-simplicity tactic common to MECHEM and KINSHIP, nor of the citation analysis used by ARROWSMITH. Also, users would have needed better assurances that AM's conjectures were plausible, by more extensive testing or by connecting the program to an automated theorem prover. Finally, Lenat himself pointed out the obscurity of AM's internal representation of concepts and conjectures, which harmed intelligibility.

4. Conclusion

It has long been known that heuristic search in combinatorial spaces provides a plausible framework for automating some problems of discovery in science [24] and mathematics [22]. Finding other sources of generality has been a problem; it has not been evident what other commonalities there were among programs that do diverse tasks in different fields, although at least one attempt [49] did address model building as a subtype of discovery and built one system based on this modest generalization [48]. This article proposes a generalization based on a program/scientist collaboration.

We have characterized discovery in science as the generation of novel, interesting, plausible, and intelligible knowledge, which builds on, but diverges from, prior descriptions of knowledge discovery. We then analyzed four successful programs in chemistry, medicine, mathematics, and linguistics by asking how their design, or the circumstances of their application, heighten the chances of finding knowledge having all four properties. Some general patterns are discernible.

We propose that future work on collaborative discovery programs, especially in science but also elsewhere in knowledge discovery, would do well to design, evaluate, and report their programs along these four dimensions. This practice would improve evaluation in this field, it would ease the comparison of different programs that do different tasks, and it could lead to further qualitative principles.

Acknowledgements

This work was supported by grants #IRI-9421656 and #DBI-9727699 from the (USA) National Science Foundation, by the NSF Division of International Programs, and by the Center for Light Microscope Imaging and Biotechnology at Carnegie Mellon. Thanks to Siemion Fajtlowicz and Don Swanson for comments on the descriptions of the GRAFFITI and ARROWSMITH programs; any remaining inaccuracies are not their responsibility. Vladimir Pericliev and Andrew Zeigarnik have been, respectively, collaborators on the KINSHIP/MPD and the MECHEM projects. This article is based on a presentation at the International Congress on Discovery and Creativity held May, 1998 in Ghent, Belgium.

References

[1] E. Alberdi, D. Sleeman, ReTAX: A step in the automation of taxonomic revision, Artificial Intelligence 91 (2) (1997) 257–279.

[2] L.G. Bruk, S.N. Gorodskii, A.V. Zeigarnik, R.E. Valdés-Pérez, O.N. Temkin, Oxidative carbonylation of phenylacetylene catalyzed by Pd (II) and Cu (I): Experimental tests of forty-one computer-generated mechanistic hypotheses, J. Molec. Catal. A: Chemical 130 (1–2) (1998) 29–40.
[3] F. Chung, The average distance and the independence number, J. Graph Theory 12 (2) (1988) 229–235.
[4] V. Corruble, J.-G. Ganascia, Induction and the discovery of the causes of scurvy: A computational reconstruction, Artificial Intelligence 91 (2) (1997) 205–223.
[5] L. Darden, M. Cook, Reasoning strategies in molecular biology: Abstractions, scans, and anomalies, in: R. Burian (Ed.), Proceedings of the Philosophy of Science Association, 1994.
[6] H. de Jong, A. Rip, The computer revolution in science: Steps toward the realization of computer-supported discovery environments, Artificial Intelligence 91 (2) (1997) 225–256.
[7] L. de Ledesma, A. Pérez, D. Borrajo, L.M. Laita, A computational approach to George Boole's discovery of mathematical logic, Artificial Intelligence 91 (2) (1997) 281–307.
[8] T.G. Dietterich, C. Bakiri, Solving multiclass learning problems via error-correcting output codes, J. Artificial Intelligence Res. 2 (1995) 263–286.
[9] S. Fajtlowicz, On conjectures of Graffiti, Discrete Math. 72 (1988) 113–118.
[10] S. Fajtlowicz, On conjectures of Graffiti V, in: Proceedings 7th International Quadrennial Conference on Graph Theory, Combinatorics and Applications, Vol. 1, 1995, pp. 367–376.
[11] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, The KDD process for extracting useful knowledge from volumes of data, Comm. ACM 39 (11) (1996) 27–34.
[12] E. Feigenbaum, B. Buchanan, DENDRAL and Meta-DENDRAL: Roots of knowledge systems and expert system applications, Artificial Intelligence 59 (1–2) (1993) 233–240.
[13] W.J. Frawley, G. Piatetsky-Shapiro, C.J. Matheus, Knowledge discovery in databases: An overview, in: G. Piatetsky-Shapiro and W.J. Frawley (Eds.), Knowledge Discovery in Databases, AAAI Press, Menlo Park, CA, 1991, pp. 1–27.
[14] C. Glymour, D. Madigan, D. Pregibon, P. Smyth, Statistical themes and lessons for data mining, Data Mining and Knowledge Discovery 1 (1) (1997) 11–28.
[15] W.H. Goodenough, Componential analysis, Science 156 (1967) 1203–1209.
[16] N. Gray, Dendral and Meta-Dendral—The myth and the reality, Chemometrics and Intelligent Laboratory Systems 5 (1988) 11–32.
[17] M. Kamber, R. Shinghal, Proposed interestingness measure for characteristic rules, in: Proceedings AAAI-96, Portland, OR, MIT Press, Cambridge, MA, 1996.
[18] P.D. Karp, Design methods for scientific hypothesis formation and their application to molecular biology, Machine Learning (1993).
[19] W. Klosgen, Efficient discovery of interesting statements in databases, J. Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies 4 (1) (1995) 53–69.
[20] P. Langley, The computer-aided discovery of scientific knowledge, in: Proceedings 1st International Conference on Discovery Science, Springer, Berlin, 1998.
[21] P. Langley, H. Simon, G. Bradshaw, J. Zytkow, Scientific Discovery: Computational Explorations of the Creative Processes, MIT Press, Cambridge, MA, 1987.
[22] D.B. Lenat, AM: Discovery in mathematics as heuristic search, in: R. Davis, D.B. Lenat (Eds.), Knowledge-Based Systems in Artificial Intelligence, McGraw Hill, New York, 1982.
[23] R. Lindsay, B. Buchanan, E. Feigenbaum, J. Lederberg, Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project, McGraw Hill, New York, 1980.
[24] R. Lindsay, B. Buchanan, E. Feigenbaum, J. Lederberg, DENDRAL: A case study of the first expert system for scientific hypothesis formation, Artificial Intelligence 61 (2) (1993) 209–261.
[25] A. Newell, J. Shaw, H. Simon, Programming the logic theory machine, in: Proceedings Western Joint Computer Conference, 1957, pp. 230–240.
[26] V. Pericliev, R.E. Valdés-Pérez, A discovery system for componential analysis of kinship terminologies, in: Proceedings 16th International Congress of Linguists, 1997.
[27] V. Pericliev, R.E. Valdés-Pérez, Automatic componential analysis of kinship semantics with a proposed structural solution to the problem of multiple models, Anthropological Linguistics 40 (2) (1998) 272–317.
[28] V. Pericliev, R.E. Valdés-Pérez, A procedure for multi-class discrimination and some linguistic applications, in: Proceedings COLING-ACL: 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics, 1998, pp. 1034–1040.

[29] W. Salmon, The Foundations of Scientific Inference, University of Pittsburgh Press, Pittsburgh, PA, 1966.
[30] S. Salzberg, A. Delcher, S. Kasif, O. White, Microbial gene identification using interpolated Markov models, Nucleic Acids Research 26 (2) (1998) 544–548.
[31] A. Silberschatz, A. Tuzhilin, What makes patterns interesting in knowledge discovery systems. IEEE Trans. Knowledge and Data Engineering 8 (6) (1996) 970–974.
[32] H.A. Simon, R.E. Valdés-Pérez, D.H. Sleeman, Scientific discovery and simplicity of method (Editorial), Artificial Intelligence 91 (2) (1997) 177–181. (Special issue on Scientific Discovery.)
[33] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, Lecture Notes in Statistics, Springer, New York, 1993.
[34] D.R. Swanson, Migraine and magnesium: Eleven neglected connections, Perspectives in Biology and Medicine 31 (4) (1988) 526–557.
[35] D.R. Swanson, N.R. Smalheiser, An interactive system for finding complementary literatures: A stimulus to scientific discovery, Artificial Intelligence 91 (2) (1997) 183–203.
[36] P. Thagard, Explanatory coherence, Behavioral and Brain Sciences 12 (3) (1989) 435–502.
[37] R.E. Valdés-Pérez, Algorithm to generate reaction pathways for computer-assisted elucidation, J. Comput. Chem. 13 (9) (1992) 1079–1088.
[38] R.E. Valdés-Pérez, Algorithm to test the structural plausibility of a proposed elementary reaction, J. Comput. Chem. 14 (12) (1993) 1454–1459.
[39] R.E. Valdés-Pérez, Algorithm to infer the structures of molecular formulas within a reaction pathway, J. Comput. Chem. 15 (11) (1994) 1266–1277.
[40] R.E. Valdés-Pérez, Conjecturing hidden entities via simplicity and conservation laws: Machine discovery in chemistry, Artificial Intelligence 65 (2) (1994) 247–280.
[41] R.E. Valdés-Pérez, Human/computer interactive elucidation of reaction mechanisms: Application to catalyzed hydrogenolysis of ethane, Catal. Lett. 28 (1) (1994) 79–87.
[42] R.E. Valdés-Pérez, Machine discovery in chemistry: New results, Artificial Intelligence 74 (1) (1995) 191–201.
[43] R.E. Valdés-Pérez, Some recent human/computer discoveries in science and what accounts for them, AI Magazine 16 (3) (1995) 37–44.
[44] R.E. Valdés-Pérez, A new theorem in particle physics enabled by machine discovery, Artificial Intelligence 82 (1–2) (1996) 331–339.
[45] R.E. Valdés-Pérez, V. Pericliev, Concise, intelligible, and approximate profiling of numerous classes, submitted for publication.
[46] R.E. Valdés-Pérez, V. Pericliev, Maximally parsimonious discrimination: A generic task from linguistic discovery, in: Proceedings AAAI-97, Providence, RI, AAAI Press, Menlo Park, CA, 1997, pp. 515–520.
[47] R.E. Valdés-Pérez, C.A. Stone, Systematic detection of subtle spatio-temporal patterns in time-lapse imaging. II. Particle migrations, Bioimaging 6 (2) (1998) 71–78.
[48] R.E. Valdés-Pérez, J.M. Zytkow, Systematic generation of constituent models of particle families, Physical Review E 54 (2) (1996) 2102–2110.
[49] R.E. Valdés-Pérez, J.M. Zytkow, H.A. Simon, Scientific model-building as search in matrix spaces, in: Proceedings AAAI-93, Washington, DC, AAAI Press, Menlo Park, CA, 1993, pp. 472–478.
[50] A.V. Zeigarnik, R.E. Valdés-Pérez, A proposed methodological improvement in the elucidation of chemical reaction mechanisms based on chemist/computer interaction, J. Chemical Education, submitted (revised) for publication.
[51] A.V. Zeigarnik, R.E. Valdés-Pérez, O.N. Temkin, Metal-catalyzed ethylene hydrogenation: The method of interactive search for multiple working hypotheses, Langmuir 14 (16) (1998) 4510–4516.
[52] A.V. Zeigarnik, R.E. Valdés-Pérez, O.N. Temkin, L.G. Bruk, S.I. Shalgunov, Computer-aided mechanism elucidation of acetylene hydrocarboxylation to acrylic acid based on a novel union of empirical and formal methods, Organometallics 16 (14) (1997) 3114–3127.
[53] J. Zytkow (Ed.), Machine Discovery, Kluwer, Dordrecht, 1996.