15

The Dendral Project

After ED FEIGENBAUM MOVED FROM UC BERKELEY TO STANFORD IN 1965, HE became interested in "creating models of the thinking processes of scientists, especially the processes of empirical induction by which hypotheses and theories were inferred from data." As he put it, "What I needed was a specific task environment in which to study these issues concretely."[1] Feigenbaum recalls attending a Behavioral Sciences workshop at Stanford and hearing a talk by Joshua Lederberg (1925–2008; Fig. 15.1), a Nobel Prize–winning geneticist and founder of the Stanford Department of Genetics. Lederberg talked about the problem of discerning the structure of a chemical compound from knowledge of its atomic constituents and from its mass spectrogram. This sounded like the kind of problem Feigenbaum was looking for, and he and Lederberg soon agreed to collaborate on it.[2]

Chemical molecules are described by formulas that give their atomic constituents. For example, the formula for propane is C_3H_8, indicating that it consists of three carbon atoms and eight hydrogen atoms. But there is more to know about a compound than what atoms it is made of. The atoms composing a molecule are arranged in a geometric structure, and chemists want to know what that structure is. The three carbon atoms in propane, for example, are attached together in a chain. The two carbon atoms at the ends of the chain each have three hydrogen atoms attached to them, and the single carbon atom in the middle of the chain has two hydrogen atoms attached to it. Chemists represent this structure by the diagram shown in Fig. 15.2.

Chemists have found that it is not too difficult to discern the structure of simple compounds like propane. However, it is more difficult for more complex compounds, such as 2-methyl-hexan-3-one, a ketone with chemical formula C_7H_{140}. One method that chemists have used to infer the structure of a compound is to bombard it with high-energy electrons in a mass spectrometer. The electron beam of a mass spectrometer breaks the compound into fragments, and the resulting fragments are sorted according to their masses by a magnetic field within the spectrometer. A sample mass spectrogram is shown in Fig. 15.3.

The fragments produced by the mass spectrometer tend to be composed of robust substructures of the compound, and the masses of these reveal hints about the main structure. An experienced chemist uses "accumulated knowledge" (to use Berliner's phrase) about how compounds tend to break up in the mass spectrometer to make good guesses about a compound's structure.

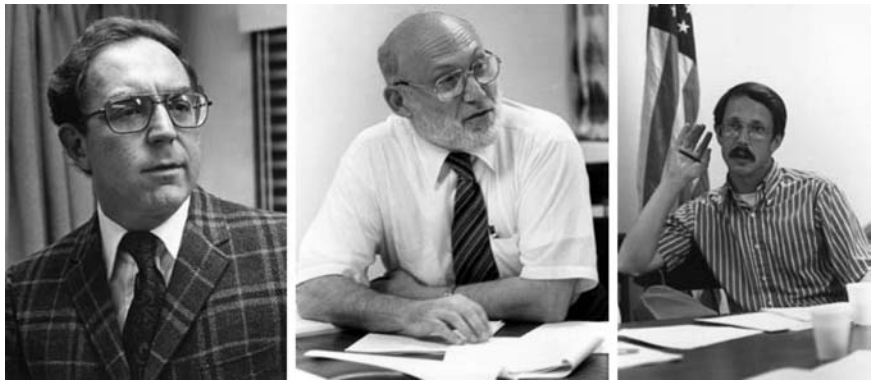

Figure 15.1. Edward Feigenbaum (left), Joshua Lederberg (middle), and Bruce Buchanan (right). (Photographs courtesy of Edward Feigenbaum.)

Feigenbaum and Lederberg, together with their colleague Bruce Buchanan (1940–), who had joined Stanford in 1966 after obtaining a Ph.D. in Philosophy at the University of Michigan, set about attempting to construct computer programs that could use mass spectrogram data, together with the chemical formula of a compound, to "elucidate" (as they put it) the structure of the compound.

Lederberg had already developed a computer procedure called Dendral (an acronym for Dendritic Algorithm) that could generate all topologically possible acyclic structures given the chemical formula and other basic chemical information about how atoms attach to other atoms. (An acyclic structure is one that does not contain any rings. You might recall, for example, that benzene contains six carbon atoms arranged in a hexagon, which chemists call a ring. Each of the carbon atoms has a hydrogen atom attached to it.) Lederberg's algorithm proceeded incrementally by generating partial structures from the main formula, then generating more articulated partial structures from these and so on in a treelike fashion. The tips or leaves of the tree would contain the final, fully articulated topologically possible structures. Finding the actual structure of a compound (or at least the most plausible actual structures) can be likened to a search down the tree to the appropriate tip or tips.

Feigenbaum and colleagues proposed using the knowledge that skilled chemists used when interpreting mass-spectral data. The chemists knew that certain features of the spectrograms implied that the molecule under study would contain certain substructures and would not contain other ones. This knowledge could be used to limit the possible structures generated by Lederberg's Dendral algorithm.

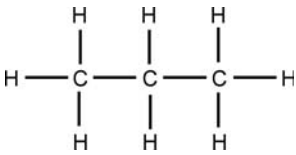

Figure 15.2. The structure of the propane molecule.

https://doi.org/10.1017/CBO9780511819346.018 Published online by Cambridge University Press

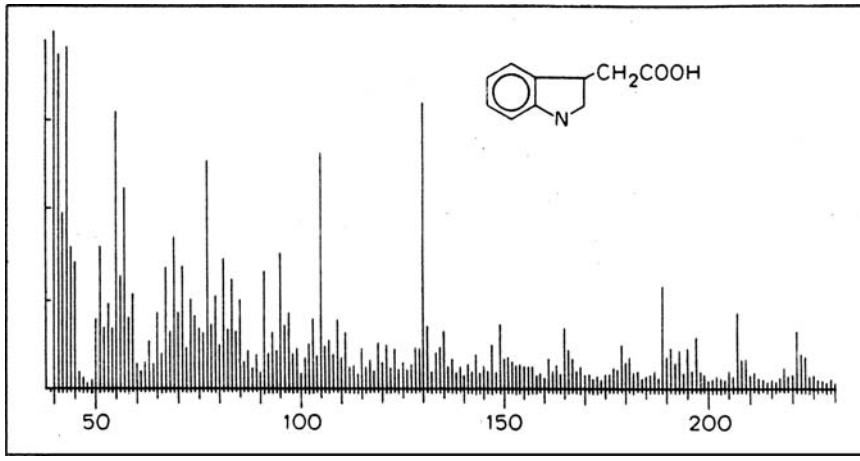

Figure 15.3. A mass spectrogram. (Illustration used with permission of Edward Feigenbaum.)

Knowledge of this sort was represented as "rules." Here is one example of a Dendral rule:

```
Rule 74:
    IF   The spectrum for the molecule has two peaks
         at masses X1 and X2 such that:
             X1 + X2 = M + 28
                 and
             X1 - 28 is a high peak
                 and
             X2 - 28 is a high peak
                 and
             at least one of X1 or X2 is high
    THEN The molecule contains a ketone group
```

The first program to employ this kind of knowledge was called HEURISTIC DENDRAL. (The adjective "heuristic" was used because knowledge from the chemists was used to control search down the Dendral tree.) It used as input the chemical formula and mass-spectrometer data (and sometimes nuclear-magnetic-resonance data) and produced as output an ordered set of chemical structure descriptions hypothesized to explain the data. Early work with HEURISTIC DENDRAL was limited to elucidating the structure of acyclic compounds because these were the only ones that Lederberg's algorithm could handle. These included saturated acyclic ethers, alcohols, thioethers, thiols, and amines. Here is one example of the power of their early program: There are 14,715,813 possible structures of *N*,*N*-dimethyl-1-octadecyl amine. Using the mass spectrum of that compound, HEURISTIC DENDRAL reduced the number to 1,284,792. Using the mass spectrum and nuclear-magnetic-resonance data, just one structure survived.[3]

https://doi.org/10.1017/CBO9780511819346.018 Published online by Cambridge University Press

The name "DENDRAL" came to describe a whole collection of programs for structure elucidation developed during the Dendral project, which continued to the end of the 1970s. Many of these programs are used by chemists today. Computer scientists and chemists working on the project were able to extend Lederberg's algorithm to handle cyclic compounds. After Lederberg persuaded Stanford chemist Carl Djerassi to join the project, performance was expanded greatly in both breadth and depth.[4]

An important innovation made during the Dendral project was a simulation of how a chemical structure would break up in a mass spectrometer. After HEURISTIC DENDRAL produced some candidate structures for a particular compound, these structures were subjected to analysis in the simulated mass spectrometer. The outputs were then compared with the actual mass spectrometer output. That structure whose simulated spectrogram was closest to the actual spectrogram was likely to be the actual structure of the compound. This process of "analysis by synthesis" came to be widely used in artificial intelligence, especially in computer vision.

From his experience during the DENDRAL years, Feigenbaum went on to champion the importance of specific knowledge about the problem domain in AI applications (as opposed to the use of general inference methods). He proposed what he called the "knowledge-is-power" hypothesis, which he later called the "knowledge principle."[5] Here is how he later described it:[6]

> We must hypothesize from our experience to date that the problem solving power exhibited in an intelligent agent's performance is primarily a consequence of the specialist's knowledge employed by the agent, and only very secondarily related to the generality and power of the inference method employed. Our agents must be knowledge-rich, even if they are methods-poor.

Embedding the knowledge of experts in AI programs led to the development of many "expert systems," as we shall see later. It also led to increased concentration on specific and highly constrained problems and away from focusing on the general mechanisms of intelligence, whatever they might be.

Notes

1. The quotation taken from "Comments by Edward A. Feigenbaum" in Edward H. Shortliffe and Thomas C. Rindfleisch, "Presentation of the Morris F. Collen Award to Joshua Lederberg," *Journal of the American Medical Informatics Association*, Vol. 7, No. 3, pp. 326–332, May–June 2000. Available online at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61437. [197]
2. For an interesting account of the history of their collaboration, see "How DENDRAL Was Conceived and Born," by Joshua Lederberg, a paper presented at the Association for Computing Machinery (ACM) Symposium on the History of Medical Informatics at the National Library of Medicine on November 5, 1987. Later published in Bruce I. Blum and Karen Duncan (eds.), *A History of Medical Informatics*, pp. 14–44, New York: Association for Computing Machinery Press, 1990. Typescript available online at http://profiles.nlm.nih.gov/BB/A/L/Y/P/_/bbalyp.pdf. [197]

3. Robert K. Lindsay, Bruce G. Buchanan, Edward A. Feigenbaum, and Joshua Lederberg, *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project*, p. 70, New York: McGraw-Hill Book Co., 1980. [199]
4. For a thorough account of achievements of the Dendral project, see *ibid.* [200]
5. The hypothesis seems to have been implicit in Edward A. Feigenbaum, "Artificial Intelligence: Themes in the Second Decade," *Supplement to Proceedings of the IFIP 68 International Congress*, Edinburgh, August 1968. Published in A. J. H. Morrell (ed.), *Information Processing 68*, Vol. II, pp. 1008–1022, Amsterdam: North-Holland, 1969. [200]
6. Edward A. Feigenbaum, "The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering," *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pp. 1014–1029, 1977. See also Edward A. Feigenbaum, "The Art of Artificial Intelligence: I. Themes and Case Studies of Knowledge Engineering," Stanford Heuristic Programming Project Memo HPP-77-25, August 1977, which is available online at http://infolab.stanford.edu/pub/cstr/reports/cs/tr/77/621/CS-TR-77-621.pdf. [200]