Henry Ford Hospital Medical Journal

Volume 32
Number 3 *Special Issue on Medical Computing*

Article 8

9-1984

Automating the Expertise of the Neuropsychologist

Kenneth M. Adams

Gregory G. Brown

Henry Ford Hosp Med J
Vol 32, No 3, 1984

Automating the Expertise of the Neuropsychologist

Kenneth M. Adams, PhD* and Gregory G. Brown, PhD*

The wide availability of computers that quickly process numbers and symbols has had a dramatic impact on medicine and the practice of clinical psychology in medical settings. Clinical psychologists can now use computers as aids to present psychological tests, perform rehabilitation training, and interpret psychological test protocols. The aim of this paper is to describe how computer programs have been used to interpret psychological data in order to answer questions about the effects of brain disease on the behavior of individual patients. These programs are able to make such inferences either by comparing the psychological test protocol of a patient with those of ideal cases or by emulating the expertise and clinical problem-solving style of a neuropsychologist. The programs described below are examples of expert problem-solving programs, a subdiscipline within artificial intelligence.

Expert Systems

An expert system or expert problem-solving program *is a set of computer programs that can derive conclusions about problems requiring the knowledge of a specialist.* The key feature of this definition is that the computer programs represent and work with related facts within a specialized area that are not common knowledge outside that field. Knowledge is used here in its general sense to denote a familiarity with facts and principles. Expert systems develop their power by representing specific knowledge rather than from use of a powerful inferential technique. The knowledge found in particular fields, such as neuropsychology, is relatively well defined and contrasts with the kind of broad, hard-to-define knowledge that we call common sense. For this reason, expert systems are easier to build than systems that reason about common sense.

The use of the word "derive" in the above definition is also intentionally vague. Computer programs derive their conclusions by a variety of methods, including the use of propositional or predicate logic, statistical inference, and heuristics. The last of these methods is any strategy that produces a derivation that may not be the best possible but is at least a useful derivation. When using a heuristic method, automated programs may come closest to assuming human qualities (1).

Expert Programs in Neuropsychology

Characteristics of neuropsychological variables

The principal assessment method in neuropsychology is the psychological test. Many tests exist for many purposes, and each has technical characteristics that makes it more or less useful in various applications. Detailed evidence about time demand, item characteristics, reliability, validity, and normative data is required before these tests can be used. In this regard, psychology may be unique among behavioral sciences in medicine because its procedures for evaluation are objective and public rather than dependent upon the private acumen or deftness of practitioners. The qualitative and quantitative information accumulated by neuropsychological assessment and testing is the evidence which the clinician uses to make clinical predictions and judgments about patient behavior. This objective process of data inspection, evaluation, weighting, and interpretation provides an excellent opportunity for computer modeling and possible automation because it offers the advantages of reducing the amount of professional staff time involved and increasing the reliability of repetitive tasks.

Much scientific evidence (2-5) suggests that defined clinical classification tasks can be executed more reliably and rapidly by a computer than by expert clinical judges, regardless of how much experience or time they have. This is true not because computers are inherently superior to human beings but because the computer program executes its decision rules with utter reliability and without distractions, hunches, or "second guesses."

Approaches to computer classification or diagnosis in neuropsychology

Our work has focused chiefly on the construction of computer programs to help us classify the behavior of patients with cerebral dysfunction. A rich literature exists about behavioral changes and syndromes attributable to brain disease or dysfunction. How can these facts, findings, and their methods be translated into computer algorithms so that patient parameters and individual test findings can be entered into the computer and descriptive statements emerge as a result?

Three major approaches to this problem have been used: 1) a taxonomic "key," 2) a geographic/geometric

Submitted for publication: April 11, 1984
Accepted for publication: November 12, 1984
*Department of Psychiatry, Neuropsychology Division, Henry Ford Hospital
Address reprint requests to Dr. Adams, Neuropsychology Division, Henry Ford Hospital, 2799 W Grand Blvd, Detroit, MI 48202.

probability approach, and 3) methods that attempt to recreate the cognitive activity of the clinician. Each of these approaches has distinctive advantages and liabilities.

1. *Taxonomic approaches to computer classification*

In biology the "key" method to identify and name species is the taxonomic manual (6). The manual is not itself a system of classification but a means of placing an individual specimen within a classification scheme that has already been established. Classification is the grouping of species so that their characteristics are consistent with each other; that is, a class or category consists of a group in which all members have the same characteristics. The key uses these characteristics to locate the class or category to which the species belongs. If a person has an unknown specimen in hand, the key should identify the classification or name of the specimen by using characteristics of the organism.

For the student of medical computing, the applicability of the key should be obvious. Since diseases or problems, such as those of cerebral dysfunction, have discrete characteristics or interrelated symptoms, it should be possible to construct computer programs which follow their own key. This presumes that the computer program can be constructed to include objective decision rules or "characters" that allow the user to enter raw data and obtain results. Problems can occur with this approach if 1) the criteria for the key are unclear, and 2) if the decision rules concerning the disposition of a case are not clear-cut, eg, for cases in which the degree of "fit" to a syndrome ideal is imperfect.

This method and these problems are well illustrated in a neuropsychological key program developed by several neuropsychologists in 1970 (6). They based their key on experience and expectancy about the performance of patients on tests of intelligence, achievement, learning/cognition, sensation/perception, and motor skills. Moreover, they focused their key on the absolute numerical scores or level of performance on the tests — rather than on the qualitative pattern of scores, special signs, or other more elaborate methods of clinical inference.

Their key included several subprograms to 1) identify brain-damaged performances, 2) localize the cerebral problem, and 3) evaluate the momentum or process of the lesion (ie, active versus static). In initial studies, the key produced a significant degree of agreement with a clinical judge (6). Cross-validation studies have confirmed these initial findings (7,8), but the extent to which the program can be generalized to other settings may still be questionable.

2. *Geometric/geographic approaches*

Key or taxonomic programs focus on rules for identifying a specimen or case within a preexisting framework of classification or diagnosis. In contrast, the geometric/geographic approach attempts to develop a "goodness of fit" for an individual case or specimen against some mathematical or graphic standard. More specifically, this approach can be applied to brain-behavior relationships. Results from known index or ideal cases of brain damage or dysfunction can be programmed into a computer. The relative location of the array of test results to the three spatial coordinates of the brain (X,Y,Z) can be stored for reference (Fig. 1).

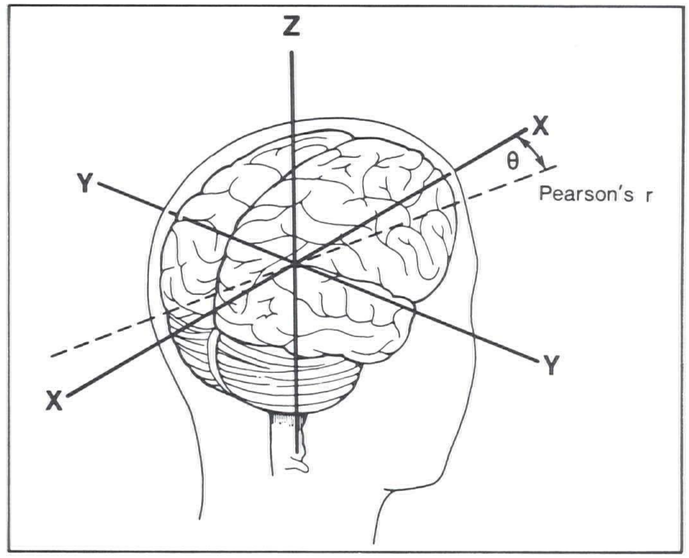

Fig. 1

Correlation coefficient plotted as an angle in a Cartesian coordinate system.

One neuropsychological program (9) used an approximation of this method to compare new cases against a three-dimensional ideal which served as a standard. In an ideal program, each case could be mathematically compared by calculating a correlation coefficient between the standard index case and the unidentified one (Fig. 1). This method can be further refined by using various statistics for "goodness of fit" so that new, unknown cases can be measured against identified standard ones whose brain-behavior meaning is clear. The method could be refined still more by using a probabilistic model for such fitting.

The disadvantage of such systems rests in their advanced mathematical conception of the nervous system. In many instances, these "smart" calculation schemes—many of which could incorporate stunning graphic displays—are "dumb" about fundamental problems and cannot recognize variability when an attempt is made to use such schemes for clinical and behavioral purposes. For example, in the cognate area of clinical electrophysiology, when very sophisticated programs were used to analyze evoked potentials or EEG spectra (10,11), incredibly powerful technology was being used to verify incredibly simple features, features that can be readily shown by simpler means that do not require complex computer programs. Other applications attempt to extract information from sinusoidal brain waves, but the relation of these brain waves to behavior is controversial (11).

3. *Attempts to simulate the cognitive activity of the clinician*

Research on the process of clinical judgment indicates that man-made reconstructions of the cognitive activity of the clinician should emulate the style and content of the clinician's cognitive process as closely as possible. In this respect, the computer program would be a "paramorphic representation" of this process, mirroring the clinician's activity. This effort is part of the wider and developing field of artificial intelligence.

Two programs have been developed which attempt to reflect the clinician's activity. One program, entitled BRAIN I, uses four types of clinically-derived rules to infer the presence, locus, and type of brain disease from neuropsychological test protocols (12). The second, Adams' Revised Program, makes interpretations about the effects of brain disease on psychological abilities (13); patterns of ability deficits are the bases for inferences about the presence of brain disease.

BRAIN I

BRAIN I takes as its basic data a patient's age and numerical scores from the Halstead-Reitan Neuropsychology Battery. This battery of tests includes procedures to examine higher cortical, sensory-cortical, and motor functioning. However, the results generated by this program are sparse. It concludes that brain disease is either present or absent; if present, the program concludes that the disease is focal, multi-focal or diffuse, and either recent or not recent. Finally, the program offers a conclusion about the type of neurological disease that may have produced the profile of neuropsychological test scores, which served as the basis for the inference.

BRAIN I simulates the decision-making of a single clinician, Dr. Ralph Reitan. It attempts to represent four types of inferential rules that focus a clinician's attention on four features of a neuropsychological protocol: 1) the absolute level of performance of a patient's test scores, 2) comparisons between the left and right sides of the body, 3) pathognomonic signs, and 4) differential patterns of test scores that may predict brain dysfunction.

1. Level of performance

A variety of neuropsychological rules of inference can compare a patient's level of performance with normative expectations. These rules characterize patients on the basis of a statistical concept of normality. Patients with brain disease often perform so poorly on psychological tests that their scores cannot readily be accounted for by normal variation of ability. For example, it is rare for an individual to have an IQ of 50. Further, it would be uncommon for a 25-year-old individual to be impaired on 60% of the tests which are sensitive to the presence of brain dysfunction on the Halstead-Reitan Battery. When a clinician uses a patient's general level of performance to infer the presence of brain dysfunction, the inference is usually sensitive to the existence of cerebral disease but not specific to it. The presence of psychosis, cultural deprivation, and very poor cooperation might each produce poor performance on neuropsychological tests without implying any brain disorder disease substrate.

2. Right and left sides of the body

Clinically useful rules of inference used to assess brain dysfunction can answer questions about asymmetries in the sensory, perceptual, and motor functions of the two sides of the body. The pitfall of these rules is that it is difficult to discriminate peripheral from central mechanisms.

3. Pathognomonic signs

Although inferences about the absolute level of performance and left-right comparisons are based on quantitative aspects of the patient's neuropsychological protocol, pathognomonic signs rely on qualitative features of the patient's performance that are very rare among people who do not have brain disease. For example, patients with Broca's aphasia have agrammatic, nonfluent, and telegraphic speech that occurs only among individuals with cerebral dysfunction. Although pathognomonic signs often have high specificity, they typically are low in sensitivity since each sign is uncommon even among patients who have brain disease.

4. Differential patterns

Rules based on differential patterns lead to inferences that derive from clusters of scores, and these scores are uncommon among patients who do not have brain disease. For instance, the presence of paresis of the right upper extremity, which is associated with Broca's aphasia, produces a distinctive pattern of findings on neuropsychological tests; in turn, that pattern conveys information about the localization of brain disease that is not evident when such findings occur individually. As with pathognomonic signs, any particular pattern is usually not commonly found even among patients with brain disease.

Although the algorithm of BRAIN I faithfully attempts to simulate a clinician's thinking, it does ignore an essential component of the clinician's thought in analyzing brain-behavior relationships. The program ignores inferences about the effects of brain disease on specific psychological abilities. Clinical neuropsychological reports usually describe a patient's verbal-intellectual ability, difficulties in solving visual and spatial problems, expressive and receptive language skills, and memory functioning. More specialized tests would evaluate other, more specific abilities. Often, it is not possible to infer a deficit of specific ability from a single test score, since an adequate score on psychological tests depends on many factors. Not only does the patient's performance depend on abilities ostensibly measured by the test, but it also depends on other abilities essential to complete the task. Even so-called simple clinical tests of memory, such as are used in mental status examinations, can involve many component skills. For example, to successfully recall the names of three objects after a few minutes of distracting conversation, the patient must have adequate auditory acuity, a certain level of arousal, relatively intact speech comprehension, the capacity to speak, and adequate motivation, in addition to sufficient verbal memory capacity. Therefore, it requires the knowledge of an expert to describe the fundamental difficulties of the patient's psychological ability as this is reflected in a particular pattern of test scores.

Adams' Revised Program (ARP)

Adams' Revised Program (ARP) simulates the clinician's ability to make decisions about which psychological abilities are impaired in specific patients with brain disease. The program uses numerical data from psychological test protocols to produce both a narrative and a tabular description of how a patient performed on 18 psychological abilities. The program also makes inferences about the presence and laterality of brain disease. By explicitly outlining the relationships between psychological test scores and psychological abilities, on the one hand, and between abilities and focal brain lesions, on the other, the program developer can identify and correct faulty behavioral assumptions about brain-behavior correlates.

Comparisons of the different automated methods

Each of these automated programs has its own unique way of organizing knowledge about brain-behavior relationships. In the key approach, knowledge is organized around taxons, or classifications which have their characteristic features organized in a hierarchical manner. In the geometric-geographic approach, knowledge is represented in an idealized, localized view of the unique cognitive and electrophysiological aspects of different brain regions. The two programs that illustrate the third approach (simulating the cognitive activity of the clinician) also differ from each other. The BRAIN I program relates knowledge of brain-behavior relationships to a theory of clinical inference that relies on four different features of a neuropsychological protocol. The Adams' Revised Program, by contrast, represents knowledge of brain-behavior relationships within a theory of human abilities.

Although these automated, interpretive programs differ in how they organize knowledge of neuropsychology, they do have several similarities. All argue from antecedent conditions to consequent conditions. None assembles data provided by the user and investigates what antecedent conditions could produce such data. This failure to look back means that, by and large, these programs do not adjust for contradictions based on inferences made early in the reasoning chain. None of the programs is adaptive; that is, none alters its decision rules as a function of the success or failure of its inferences. Another limitation of all these methods is that none provides, as routine information about individual cases, a description of the line of reasoning followed to reach a specific conclusion. Such a description would be useful not only for correcting errors in the reasoning, but also for teaching clinical inference once a program has been validated. Furthermore, none of these programs provides any significant interactive capability with the user; only one (Adams') has an explicit way of handling missing data, and this is primitive. All are fairly inflexible about the questions they can answer and the data they need in order to answer questions, yet this inflexibility helps to greatly simplify the programs. Fortunately, most applications of neuropsychological testing involve the process of answering a discrete set of possible referral questions based on a fixed set of data entered into the program.

A Practical Test

Are there differences in the accuracy of the expert program approaches described above? We compared three of the above approaches, the key method, BRAIN I, and the ARP (6,12,13).

Subjects

We selected 30 older, right-handed patients (mean age: 60.7 years; standard deviation: 7.0 years) of average socio-economic status and education (mean education: 11.6 years; standard deviation: 3.2 years) from patients at Henry Ford Hospital who had transient ischemic attacks with clearly identified lesions. These patients were carefully screened to rule out predisposing factors (eg, excessive alcohol consumption, diabetic retinopathy) as well as frank, preexisting neurological disorders (eg, head injury). We obtained complete neuropsychological examinations on all patients and entered their results into each of the computer programs.

Results

We compared three computer systems (key, BRAIN I, and ARP) using the subject pool described above. Results for determining the *presence* of brain damage in brain-damaged individuals are shown in Fig. 2. For left hemisphere cases, the ARP gave 92% correct responses, followed by 67% correct responses for BRAIN I, and 58% correct responses for the key approach. For right hemisphere cases, BRAIN I identified 82% and ARP identified 73%. The key approach, which determined diffuse damage cases, was 71% accurate. Overall, ARP was correct for 82% of the cases; BRAIN I, for 75% of the cases; and the key approach, for 65% of the cases.

Figure 3 reveals that the key approach was accurate in determining *lateralization* for 14% of the left hemisphere cases, while BRAIN I identified 63% and the ARP 27%. In right hemisphere cases, the key had a 20% rate, and the ARP identified 29%; however, BRAIN I had no successes. For diffuse cases, the key identified 60%, BRAIN I 50%, and the ARP identified 57%. Overall, the key identified laterality in 29% of the cases, while BRAIN I and the ARP both accurately identified laterality in 38% of the predictions.

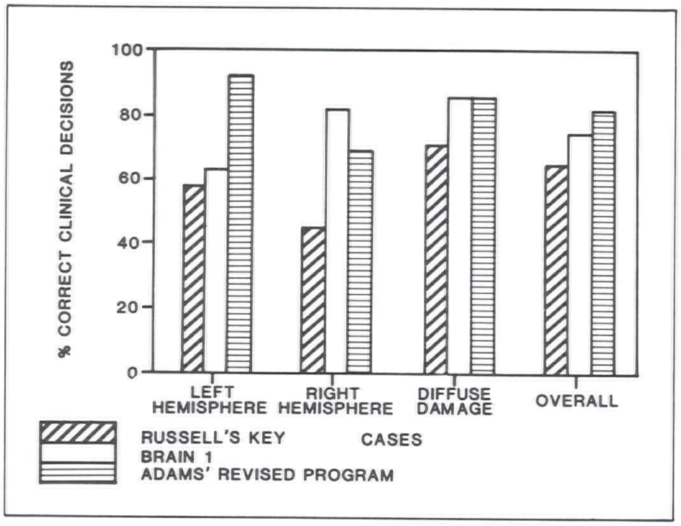

Fig. 2

Prediction of the presence of a cerebral lesion for each of three computer programs as a function of the laterality of a lesion.

Discussion

Our results indicate that these three computer programs are inadequate as comprehensive neuropsychological "experts." Although these programs accurately identify the presence of cerebral dysfunction, they are not consistently capable of making more subtle distinctions about lateralization (14). Our study did not address process or disease-type predictions. Moreover, we did not address specifically the question of *absence* of brain damage by including a control or nonpatient comparison group.

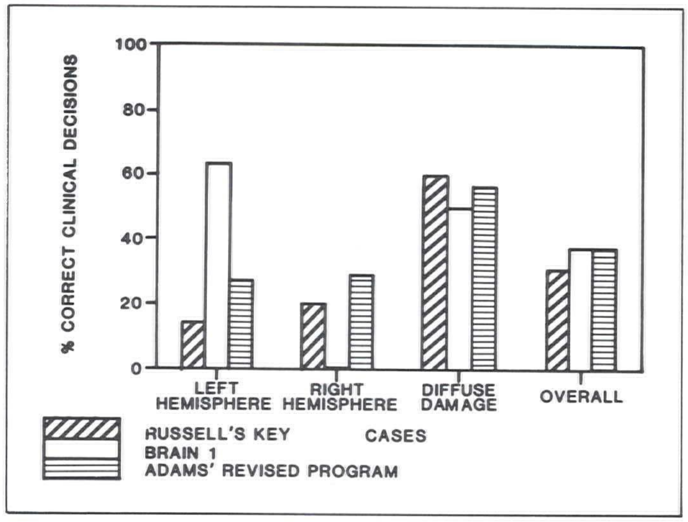

Fig. 3

Prediction of the laterality of the lesion by means of three computer programs.

The clinical value of any of these programs is limited. Each of the programs was offered by their authors as experimental endeavors needing development. It is nonetheless interesting that the three programs produce comparable results with very different theoretical approaches. The key is a quasi-taxonomic program; BRAIN I is a comprehensive, interpretative system modeling four proven modes of clinical neuropsychological inference; and ARP is an ability-based program dependent upon psychological constructs.

Various subtleties in lesion location, type, and severity might have hampered these computer programs. As Smith (15) points out, these factors will often render invalid strict rules for lateralization and localization. In the future, expert programs must take into consideration interactions among these three variables and account for differences in the time course of the various classes of neurologic diseases. While in need of development, these programs do represent attempts to make clinical decisions by set rules. Each program represents a different approach and objective; many others could be conceived from obvious models or departure points in clinical neuropsychology. While future programs are likely to be more elegant, such programs may not need to be more sophisticated to work. Rather, one may need only to eliminate unproductive rules and build in a feedback loop enabling the program to train itself by reentering data and updating the database from each case.

In this exercise, the lateralization success that did occur was not predicated upon the use of sensory-motor data from the neurological exam (8,14). Our criteria for lateralization were neuroradiological and neurosurgical.

Despite the limited success of the programs so far developed, efforts to create actuarial or automated systems for neuropsychological analysis should continue. It is not sufficient to observe that clinical interpretation in psychology is complex and therefore to conclude that it is beyond the scientist's capacity to make it an objective procedure (16). As with most other areas of medicine, such reasoning begs the methodological question and ignores the evidence for the validity and reliability of present diagnostic procedures. It is plausible that future programs will encompass a level of expertise in neuropathology, neurology, and psychology that is beyond what any one person could learn in a lifetime. Inferences made by such programs could be augmented by analysis of information obtained from large databases. Then, specific frequency distributions, calculated for various signs and symptoms, could become the basis for neuropsychological inferences.

Also, the development of clinical neuropsychological algorithms should be based on rules of interpretation aimed at the likely possibilities. To establish clinical services for the rare or deviant case might limit the usefulness of the service. Computer programs may ultimately be developed which will draw the clinician's attention to the unusual or deviant case or lack of match between expected and actual data. If this can be done in a valid and reliable fashion, more effective use of the clinician's time will be possible.

References

1. Rich E. Artificial intelligence. New York: McGraw-Hill, 1983.
2. Sawyer J. Measurement and prediction, clinical and statistical. Psychol Bull 1966;66:178-200.
3. Wiggins J. Personality and prediction: Principles of personality assessment. Reading: Addison-Wesley, 1973.
4. Meehl P. Clinical vs statistical prediction. Minneapolis: University of Minnesota Press, 1954.
5. Goldberg L. Simple models or simple processes? Some research on clinical judgments. Am Psychol 1968;23:483-96.
6. Russell E, Neuringer C, Goldstein G. Assessment of brain damage. A neuropsychological key approach. New York: Wiley-Interscience, 1970.
7. Goldstein G, Shelly C. A further attempt to cross-validate the Russell Neuringer, and Goldstein neuropsychological keys. J Consult Clin Psychol 1982;50:721-8.
8. Anthony WZ, Heaton RK, Lehman RAH. An attempt to cross-validate two actuarial systems for neuropsychological test interpretation. J Consult Clin Psychol 1980;48:317-26.
9. Swiercinsky D. Computerized SAINT: System for analysis and interpretation of neuropsychological tests. Paper presented at the annual meeting of the American Psychological Association. Toronto: 1978.
10. Duffy FH. Topographic display of evoke potentials: Clinical applications of brain electrical activity mapping (BEAM). Ann NY Acad Sci 1982;388:183-96.
11. John ER, Karmel BZ, Corning WC, et al. Neurometrics. Science 1977;196:1393-410.
12. Finkelstein JN. BRAIN: A computer program for interpretation of the Halstead-Reitan neuropsychological test battery. Doctoral dissertation, Columbia University, 1976. Dissertation Abstracts International 37, 5349B. (University Microfilms No. 77-8,8864).
13. Adams KM. Automated clinical interpretation of the neuropsychological test battery: An ability based approach. Doctoral dissertation, Wayne State University, 1974. Dissertation Abstracts International 35, 6085B. (University Microfilms No. 75-13,289), 1975.
14. Heaton RK, Grant I, Anthony WZ, Lehman RAW. A comparison of clinical and automated interpretation of the Halstead-Reitan battery. J Clin Neuropsychol 1981;3:121-41.
15. Smith A. Principles underlying human brain functions in neuropsychological sequelae of different neuropathological processes. In: Filskov S, Boll T, eds. Handbook of clinical neuropsychology. New York: Wiley, 198l.
16. Meehl PE. Psychodiagnosis: Selected papers. Minneapolis: University of Minnesota Press, 1973.