SESSION VIII VAX USERS SYMPOSIUM

Alan M. Lesgold, *Presider*

COGNET: A national network for cognitive science

RICHARD B. MILLWARD

Center for Cognitive Science, Brown University, Providence, Rhode Island 02912

A planning committee funded by the Alfred P. Sloan Foundation has agreed in principle on a draft proposal for the establishment of a national network of computers for research in cognitive science. The main ideas of the proposal are discussed in this paper.

A number of scientific and technical developments in the past 10 years have led to the idea of a national network of computers for research in cognitive science. The more obvious scientific factors are the influence of artificial intelligence on psychology and linguistics, the importance of special-purpose software for psychological research, and the growth of information processing and cognitive models in psychology, linguistics, and artificial intelligence. The technical developments relate to the decrease in the cost of hardware, the better understanding of how to build networks, and the success of networks like the ARPANET, TYMNET, and TELENET. Thus, the idea of a network is not outlandish or futuristic but a natural culmination of a steadily growing science and technology.

Our proposal has been preceded by a proposal from a number of computer science departments for a CSNET.¹ The CSNET proposal emphasizes software and hardware developments that will make networks easier to implement. The CSNET will provide technical solutions to existing problems that are certainly not within our competence to solve. Once it is possible to have networks at a reasonable cost, there will be many academic groups setting up organizations much like the one in our proposal.

At this time, there is a planning committee funded by the Alfred P. Sloan Foundation to write a proposal for a cognitive science network. Appendix B contains the names and addresses of the members of this committee, along with their network addresses. The committee has met once and agreed in principle on a draft proposal that will be submitted to at least the following organizations: the National Science Foundation, the Sloan Foundation, the Office of Naval Research, the National Institute of Mental Health, and the National Institute of Education. Various relevant industries will also be

contacted both for support and to insure that cognitive research in industry is included in the network.

This paper will be restricted to a discussion of the main ideas in the proposal. There are a number of seemingly unrelated aspects to a proposal such as the one to be described. What problems will a network solve? What are its general goals? What is our model of the network when it is finished? What are the costs of such a network? Do the benefits justify the costs? Here, we try to provide a general answer to these questions. This paper will also provide a questionnaire for all to use to send information to the planning committee about individual needs and interests in a cognitive science network. The questionnaire is presented in Appendix C. Our intentions are to provide as broad a network as possible for communication among cognitive scientists, and the questionnaire will help us attain this goal.

When we say "network," we use the term to emphasize the interconnection of a number of computers at various cognitive science centers. We are not trying to develop a network but simply use one. The physical network, the details of how to make terminals and computers talk to each other over common carriers is not of central concern to us. However, since the main idea rests on that technology, it had to be considered to some extent. The technical aspects of networking will be developed in a later paper.

The most important focus of the proposal is to develop needed software for cognitive science, distribute that software widely to a number of centers, and make it available to a large number of researchers who do not now have access to it. A secondary purpose is to train researchers in the use of computers as tools for cognitive science. A third goal is to enhance communication among cognitive scientists in all departments across the

country. To motivate these goals, we first look at some problems in detail and then discuss the proposed solution, COGNET.

SOME PROBLEMS

Communication among scientists in a given discipline has always been a problem. In an emerging field, such as cognitive science, problems of communications are magnified because the traditional means of contact are highly constrained by disciplinary boundaries. Before the printing press, letters and visits were the only way for two scientists to communicate. Today we take journals, books, and even the telephone for granted and rarely consider what might happen if these were no longer possible. Yet, as important and adequate as they are (compared with the alternative of not having them), they do have some failings. The time lag for the distribution of the latest ideas through journals and books is very long and overly formal. We often claim for our journals and books that they "archive" the development of the science. Although an archival system is necessary, it is not all that is desired by an active researcher.

The traditional solution arrived at to overcome the lag in publication is the yearly convention. These serve a different function from publications because the ideas are not so well developed, the presentations are less formal, and the participants can interact personally in order to argue and exchange ideas. But conventions usually only occur once a year, and they do not allow enough time for serious interaction. They tend to be limited to a single discipline, which is not very helpful when disciplines interact.

A third obvious means of communication is the telephone, but it also has serious problems, such as no visual means of communication. The problem, then, is a general one: How do we improve the way scientists communicate with each other? A network represents another type of communication that will have significant import on the interactions of scientists.

Another general problem is the relationship between scientific research and an educational institution. Most of us believe that keeping research active in our universities is extremely important not only because such research is less tainted by economic goals, but also because that is where we train future generations of scientists. Yet, today's research often requires expensive equipment and teams of scientists, each of whom represents some special focus on the problem. Universities are concerned with representing knowledge as broadly as possible and not with hiring just the right team to solve one of nature's problems. Generally, then, scientists in universities work as individuals. This has often been satisfactory, but it is becoming less and less adequate. There is no simple solution to this problem, but it has caused some concern among many scientists because so many problems are really too

difficult for a single person to solve. The network we are proposing will help to solve this problem by bringing isolated scientists into closer contact with those working on similar problems, regardless of where they are.

Another problem is training scientists. Although this problem is not independent of communication, it is slightly different. We know that in most cases training requires a longer period of time with intensive study of a subject. It often requires specific work in a laboratory. It certainly requires reading and doing exercises related to the subject. These conditions for training scientists are created in universities and cannot be easily distributed. However, some of this kind of interaction can be done remotely if the subject matter is right. When the methodology depends heavily on computers, there are ways to help train scientists over a computer network.

Perhaps the most serious problem facing cognitive science is the lack of proper software and facilities to use such software. Too many researchers in this emerging field have not had adequate access to software. The exception has been a few artificial intelligence laboratories with large government grants. At most universities the computer facilities provided have emphasized programming languages for mathematical or business applications, not for on-line control of experiments, speech research, or cognitive simulation. Fortunately, as long ago as 1967, psychologists were buying their own minicomputers to run experiments, and more recently, they have been able to set up speech laboratories with medium-sized machines. However, there has been no easy way for psychologists to work with simulation languages such as LISP. There is little prospect that university computing centers will begin to adequately support research using languages like LISP. Fortunately, however, machines are becoming available that not only will support cognitive science research but also are within reach of departmental budgets. However, individual departments are not going to be able to develop software on their own. It is imperative that more coordination be imposed on the software development. A network is an ideal way to bring about such coordination.

A SOLUTION

There are any number of solutions to the above problems. By focusing on one aspect of the problem or on one particular subgroup of scientists, programs could be introduced that would help to alleviate part of the problem. The solution that we are proposing tries to encompass the set of problems as a whole and solve them in a way that fits in with the future developments of the field. The solution is to create a network for cognitive science: COGNET.

As the above emphasis should have made clear, the

nature of the physical network is not our concern. We believe that the technical questions have been solved and plan to present a later paper addressing them. The emphasis is on creating a communication link that will help those in this emerging field talk to each other, despite the fact that they are in different departments and at different universities. We want a network that will provide research tools for the community at a reasonable cost and in a reasonable time. The network will be a service facility for all interested in cognition regardless of their field, their location, or their resources. It will provide one way to train scholars about research involving cognitive simulation through access to special software.

There is a model of research development that is the source of many of our ideas: the SUMEX-AIM project. The SUMEX-AIM project serves as both a source of ideas and a point of contrast. SUMEX is a "nationally shared computing resource devoted entirely to designing AI [artificial intelligence] applications for the biomedical sciences" (Freiherr, 1980). It provides computing power and special high-level software for studies in complex mental processes to many researchers at Stanford University and across the nation. Three points important for this proposal have their basis partially in the SUMEX model. One is that the facility provides software tools to researchers that they could never develop on their own. Second, the software tools were made available to researchers all over the country by the use of networks such as TELNET, TYMNET, and the ARPANET. Third, the facility served as a training site for many researchers, who then returned to their own university to continue the research.

We are not trying to duplicate the SUMEX organization, but we acknowledge its importance to our thinking. The idea of a center at which special software is developed is critical to producing that software expeditiously. Second, remote access to the specially developed software is essential because of the problems discussed above. Third, the idea of training researchers fits easily into our discussion of the needs of cognitive science.

However, we are going to contrast the SUMEX solution to medical problems to our proposed solution to cognitive science problems. By starting 10 years later, we have advantages not available to SUMEX. Networks are more highly developed, and many of the cognitive science host nodes have large computers already established. Thus, the solution of having everyone work on a single machine seems much too constraining. The students of research faculty also need computing resources. There are simply too many researchers for all to work on one or two machines. There is too much diversity for one center to handle all the different kinds of projects that we can expect to develop in the next 10 years. Essentially, then, we are using many of the good ideas from SUMEX but applying them in a new way to take into account the revolution in computing

facilities. It is now possible to have many machines, widely dispersed, yet working together. It is also necessary to accommodate the large cognitive science community.

Our conception of the network is as follows. First, it will consist of a set of host nodes representing universities with some critical mass in cognitive science. At the present time, we have a list of 20 active participants that is presented in Appendix A. Although these nodes will vary in the size of the machine(s) they have connected to the net, each will be considered a site at which some special aspect of cognitive science is of importance. The larger sites will have timeshared computers large enough to run high-level languages like INTERLISP. Generally, the machine of choice is a VAX-11/780, and by the time the network is running, there will be a very satisfactory library of programs available for that machine.

There will be a "center" that will be primarily responsible for developing software. We have in mind four full-time programmers for 3 years and two for the last 2 years of the 5-year period of the grant. There will also be two technical assistants who will oversee the distribution of software and help the users on the nodes make use of the new software. The center will have a dedicated machine for developing software and for others to use in learning about the software. The mail system will be placed on the center's machine. A series of summer workshop sessions will be held to train users in the use of the special software. Some of these workshops may be held at the center and others at various host nodes.

Two important ideas will further clarify the network concept. One is that researchers from remote locations without existing facilities can hook into the network via some common carrier and work on any one of the host machines. We do not intend for these isolated researchers to have to work on the center's machine, since it will make more sense if they work on the machine at the node doing research of most interest to them. The second point is that each node will develop its own expertise and will become a resource for the whole community. We do not believe that cognitive science is so easily defined that one center can be all things to all people. Thus, the advantages of the present concept is that it centralizes the development of software that is of general use, such as an excellent INTERLISP-like language, but it allows independence of local facilities to work with these tools and develop their own content material. Eventually, as the host nodes grow in size and knowledge, different nodes will become "centers" for special purposes. The more isolated researchers, or even a researcher at a center of some size, may want to connect to another node because that is where the appropriate software exists.

While we emphasize the ability to work remotely, it is also important to realize that, by making the nodes as comparable as possible, the software developed at one

node should be easily made to run at any other node. That is one of the purposes of the center: to start the community off with a common set of general-purpose programming languages. In this way, we gain the advantage of developing the software at one location where there are experts to do the job right. At the same time, each host node can use the software as it wishes and can develop its own special systems, if required, using the general-purpose languages available.

We plan to make the network perform three major functions: (1) It will serve as a message system for communication. (2) It will allow files to be transferred from one machine to another quickly. (3) It will allow anyone anywhere to run programs on one of the host machines—with permission, of course.

The main system will provide for a kind of communication that cannot be duplicated with letters, journals, phones, or conventions. Since anyone anywhere can be connected to the network, this mail system will connect everyone, regardless of department or university.

The fast-file transfer facility will be important for program development and maintenance. New versions of the special languages can be sent to the node machines over the network during the evening hours, thus providing software updates to everyone at the same time. The problem with magnetic tape is that the various nodes will not put the software up at the same time. By allowing the center to control the software, it is not necessary for each node to have an expert on the general-purpose languages; or, at least, the fast-file transfer facility makes it possible to have fewer experts. Also, large data bases can be shipped to other nodes if they are wanted. Documents and papers can be sent to all who are interested in them. Of particular importance is the fact that papers can be written much more easily by teams not located at the same place, since all the authors will have simultaneous access to the developing paper.

The idea behind remote program access is to let someone run a program without going to all the trouble of bringing the program up on a local machine. If a researcher simply wants to see how someone's program runs, or to test it with some special cases or new data, it will not be necessary to bring the whole program to the local machine. This helps to overcome the problems of local idiosyncracies. In some cases, for example, speech synthesis, it might be that the peripheral equipment needed to run the program is not available at the distant node. Finally, even if the program is transferred, sometimes seeing it run is a great help in getting it to run locally. A colleague at the source can step the novice through the program in an interactive dialogue just as if they were both in the same room.

The last point illustrates how the network will help in the training of researchers. It will be possible to develop special computer-assisted instruction programs that will teach the student in special subjects related to cognition. Although we plan on having summer

workshops to help in this training, even a week of intensive work is not enough. When the week is over, it will be important that there be continued access to the material learned about earlier. It would be almost impossible under present conditions for the scientist to go back to his own university and begin work using the newly learned tools because he would not have the tools available. Even carrying the software back with him would only be a first step, since he would then have to have that software brought up on his local machine. That is certainly not easily done and fails with dismal regularity. The evidence is that, without a network, there will be very little chance that researchers will be able to use these new tools if they are not at a location where the tools already exist.

An important point needs to be made here. It is nearly impossible to evaluate the research being done using special software systems unless the reviewer is familiar with the way such systems work. Just as it is nearly impossible to evaluate a mathematical model if one does not understand the mathematics, so evaluating cognitive simulations requires understanding the language and techniques being used to perform the simulations. The ELIZA program well illustrates the difficulties here. Thus, if the field is to develop with proper criticism, it is imperative that further training be provided to a much broader set of researchers than those who exist at particularly well endowed universities.

COSTS AND BENEFITS

At the present time, it is difficult to make exact estimates of the cost of such a network. We do not know what kind of reductions to expect in communication charges as the volume increases and as the hardware is mass-produced. An important aspect of the network design being considered is that the cost of connecting to one of the value-added networks, such as TELENET or TYMNET, will be commensurate with the expected use to be made of the network. One of the serious difficulties with the ARPANET was that the cost of connecting a computer to the network was extremely high—over \$70,000/year. This high entry fee, along with the restrictions imposed by the Department of Defense, made it impossible for all but a few to make use of the network.

The solution for COGNET, following ideas proposed in the CSNET proposal, is to keep costs related to use. This will be done by producing gateways of various costs and speeds and by allowing single users to enter a net without an expensive gateway or even a local computer connection. The cost of connecting a stand-alone terminal to a remote computer is about \$6-\$8/h, including charges for data transmission. Although that may seem high if text is being composed on-line, with microprocessor-based terminals, one can compose the message off-line and then connect and ship the

message at low speeds, 300 or 1,200 baud, but continuously, thus reducing connection time substantially. Thus, the entry fee for being included in a simple message system is a terminal and funds to pay for the connection. There is a very wide network connected to over 350 universities that most faculty members could make use of today. This is the EDUNET subdivision of EDUCOM. They support microprocessor-based terminals such as Apples also. The publication EDUNET NEWS describes their network.

The simplest way to connect a computer to another computer is to use a telephone connection. Although this is expensive if many long connections are made, it is not unreasonable for small files sent infrequently. One of the first developments in COGNET will be a phonenet that will allow host nodes to communicate via dial-up facilities. The cost of this is about \$2,000 for the hardware at each host site and then the cost of a telephone call. At 300 baud, one can transmit about 16 characters/sec in this way. There are only a few problems remaining before phonenets can be easily used on most laboratory machines. Thus, this represents a very satisfactory development on the immediate horizon.

Once the hardware and software problems associated with connecting host computers to value-added carriers has been solved, the various host machines will be interconnected. Then fast-file transfer and interactive programming from remote machines will be possible. The cost of the slower gateways will be around \$6,000. The line rental is still fairly high, \$1,000/month for a 9,600-baud line, but since these lines are multiplexed, the cost to an individual can be fairly reasonable. It is certainly not unreasonable to expect even a small college to have a \$6,000 network connection with a \$12,000 line rental per year. The implications of these rough estimates is that every college and university in the country will be connected through a network. This means that many mini-research clusters will be formed in which a group will discuss a particular problem even though its members are not on the same campus.

The fast gateways that will be developed under the CSNET proposal will allow multiple host computers and local terminals not connected to local host computers to connect to a value-added carrier. These gateways will cost about \$25,000/month, but since the line is multiplexed, if it is needed and used, it will be highly cost effective. These larger gateways will probably become standard in the larger universities of the country in which many research computers are currently being used. In fact, there will be great economy in setting up an adequate gateway in the beginning rather than installing multiple systems.

We do know that the center will need a fairly large machine for software development. There will also be personnel charges for the programmers and assistants.

However, producing software once instead of having it duplicated again and again is certainly less expensive in the long run. Besides, much of the needed software cannot be developed at most cognitive science sites because they do not have adequate staff support. Thus, the cost for software is high, as it always is, but much less expensive than if every research group tried to develop its own (often inadequate) software.

REFERENCE

FREIHERR, G. *The seeds of artificial intelligence. SUMEX-AIM* (Public Health Service NIH Publication No. 80-2071). Washington, D.C: U.S. Department of Health, Education, and Welfare, 1980.

NOTES

1. CSNET, a computer science research network, is the product of a team of computer scientists interested in networks. Their work has resulted in a proposal to the National Science Foundation under the direction of Lawrence H. Landweber. I intend to cover in much more detail the nature of their proposal in my next paper on networking.

2. A phonenet has a computer dial-up another computer and send messages over an ordinary telephone line. There are a couple of running systems in operation. The details of the phonenets will be covered in a later article.

Appendix A The 20 University Nodes

University	Contact
Brown University	Richard Millward
Univ. of California, Berkeley	Robert Wilensky
Carnegie-Mellon University	Allen Newell
University of Chicago	Lance Rips
University of Colorado	Peter Polson
Cornell Medical School	Michael Gazzaniga
Univ. of Illinois, Urbana-Champaign	Emanuel Donchin
Univ. of California, Irvine	Leah Larkey
Univ. of Massachusetts, Amherst	Michael Arbib
Mass. Institute of Technology	Samuel Jay Keyser
University of Michigan	Gary M. Olson
Univ. of Pennsylvania	Aravind Joshi
Univ. of Pittsburgh-LRDC	Alan Lesgold
Princeton University	George Miller
Rutgers University	N. S. Sridharen
Univ. of California, Santa Barbara	John Cotton
Univ. of California, San Diego	Donald Norman
Stanford University	Edward Feigenbaum Gordon Bower
Univ. of Texas, Austin	Philip Gough
Yale University	Roger Schank

Appendix B
The Planning Committee

Name, School, and Department	Telephone/Network Address	Name, School, and Department	Telephone/Network Address
Edward Feigenbaum	(415) 497-4079	Univ. of Calif., La Jolla Psychology	NORMAN@ISI CSUVAX.SDCVAX!DAN@BERKELEY
Stanford-SUMEX-AIM	CSD.FEIGENBAUM@SCORE		
Alan Lesgold	(412) 624-4901	Peter Polson	(303) 492-6991
Univ. of Pittsburgh-LRDC	LESOLD@CMUA	Univ. of Colorado-Psychology	POLSON@SUMEX-AIM
Richard Millward, Director	(401) 863-2608	N. S. Sridharan	(201) 932-2894
Brown University-Psychology	MILLWARD@CMUA	Rutgers-Computer Science/A.I.	(201) 932-2001 SRIDHARAN@RUTGERS
Allen Newell	(412) 421-2602	Robert Wilensky	(415) 642-7034
Carnegie-Mellon-Computer Science	NEWELL@CMUA	Univ. of Calif., Berkeley Computer Science	(415) 642-1024 CSVAX!WILENSKY@BERKELEY
Donald Norman	(714) 452-4888		

Appendix C
Cognitive Science Network Questionnaire

A. INFORMATION SOURCE

1. Name and address of person filling out this questionnaire

2. What group is represented in this response? (For example, department, research group, or just an individual)

3. Composition of the group. Indicate the number of each type of worker typically associated with the group.

Faculty members _____ Postdoctoral fellows _____

Graduate students _____ Undergraduate students _____

Research associates _____ Others (specify) _____

4. What are the major research interests of the group?

5. What use is made of computers by the group? Indicate the percentage of each type of use. Estimates are expected to be rough, but they should add up to 100 percent.

On-line control _____ Data analysis _____

Numeric simulation _____ Symbolic simulation _____

Stimulus analysis & synthesis _____ Other (specify) _____

Appendix C (Continued)

6. What other groups at your university or company should be contacted? Please list contacts as specifically as possible.

B. CURRENT COMPUTER FACILITIES

1. Center facility. (If you have access to a central computing facility, answer this question.) For later tallies, call this system W.

Main frame(s) _____

Modes of use (indicate approximate percentage of use of each type):

Batch _____	Local terminals _____
Remote terminals _____	Intercomputer _____
Remote entry stations _____	Other (specify) _____

Monitor(s) used _____

2. Special computing facilities. Describe any special computer systems that your group has access to. If more than three such systems exist, use the back of this form for the others.

	System X	System Y	System Z
Main frame & manufacturer	_____	_____	_____
Memory size	_____	_____	_____
Peripheral devices (e.g., disks, tapes, A/D, D/A)	_____	_____	_____
Name of monitor	_____	_____	_____
Approximate total cost of system	_____	_____	_____

3. Using the letters W, X, Y, and Z, indicate whether the system(s) described above can perform the functions listed below.

Languages available on system:

FORTRAN _____	PL/1 _____	APL _____	ALGOL _____
BASIC _____	LISP _____	PASCAL _____	Other _____

Text processing:

Script _____	TECO _____	Other (specify) _____
--------------	------------	-----------------------

Appendix C (Continued)

Text printers:

Diablo _____ Other (specify) _____

Mail/message system (name and source of system):

Statistical packages:

BMDP _____ SSPS _____ IMSL _____ MDS _____

Other (specify) _____

Graphics facilities (specify):

_____Speech facilities (specify):

_____Real-time control (specify):

_____Network connections (specify):

C. SUPPORT LEVEL

1. Amount spent each year on computing. Include money spent by the whole group mentioned in A.3 above. Do not include cost of equipment listed above or personnel costs.

From grant support _____ Percent of total grant support of group _____

From university _____

2. Cost of personnel devoted to computer development. (Estimate, for example, the percentage of time a research associate devotes to software and hardware development. Do not include routine program writing for experiments or data analysis.)

From grant support _____ From university support _____

D. ANTICIPATED USE OF COGNET

1. Do you now make use of any national network? _____ Which ones?

ARPANET _____ EDUNET _____ TELENET _____ TYMNET _____

2. Current cost of your network use.

Original equipment purchases _____

Local line rental or phone costs _____

Transmission costs (TELENET or TYMENT charges) _____

Remote computer costs _____

Appendix C (Continued)

3. Indicate the expected use to be made of the various functions of a network as follows: "0" for no use, "1" for slight use, "2" for moderate use, and "3" for extensive use.

Message and mail system _____

Fast-file transfer-papers _____

Remote data base _____

Remote programming _____

Use of "center"-developed software on your system _____

4. Do you anticipate that any of the computer systems, W, X, Y, or Z, will be connected to the network? _____ If so, which ones? _____
5. Which type of connection do you think is best for your group?

Terminal only _____ Medium-speed microprocessor-based gateway _____

Phonenet _____ Large, shared gateway _____

6. How much money do you estimate you could find for support for a network connection and its use?

Equipment (one-time cost for connection) _____

Yearly rental costs _____

Return to

COGNET QUESTIONNAIRE
c/o Richard Millward
Cognitive Science
Brown University
Providence, R.I. 02912