*Bradford W. Hesse, Lee S. Sproull, Sara B. Kiesler, and John P. Walsh*

# RETURNS TO SCIENCE: COMPUTER NETWORKS IN OCEANOGRAPHY

Recent developments in network communication and scientific computation have led to calls for a massive increase in computer network support for science [22]. Computer science has benefitted from networking technology since the Arpanet began in the late 1960s [16]. Other scientific disciplines began using networks in a more or less organized fashion by the early 1980s: High-energy physicists use Hepnet, and molecular biologists use Bionet. Yet network support for science has been mainly an *ad hoc* enterprise. One vision of the future is the Collaboratory: a "combination of technology, tools, and infrastructure that allows scientists to work with remote facilities (co-laboratory) and one another (collaboratory) as if they were colocated and effectively interfaced" [17].

In this article we examine the alternative theories of the returns to science from networking, present early results from one group of networked research scientists, and discuss some implications for policy and technical development.

The general case is familiar: benefits to science from network technology derive from scientists' increased access to remote scientific resources. These resources include scarce and expensive technology such as the space telescope or supercomputers, large databases such as those for DNA sequences or satellite weather images, or scientific-literature databases such as *Chemical Abstracts.* They include services such as "mail-order" fulfillment of requests for reprints, algorithms, or even experimental-chip fabrication. They also include the time and attention of other scientists, accessible through electronic mail, b-boards, and teleconferences. Increasing ease of access to these resources should lead to a concomitant increase in scientific productivity.

This argument has gone unchallenged even though, generally speaking, computer technologies have not had empirically strong effects on productivity in any industry. In his wide-ranging review of sectoral, industrial, and organizational studies, Attewell [2] shows that these studies fail to demonstrate a positive association of investments in computer technologies with productivity. For instance, P. Weill [27] divided computer investments of 33 strategic business units into "strategic systems" (e.g., an inventory system allowing sales staff to give accurate delivery time estimates), "transactional systems" (e.g., accounts payable, order entry), and information systems (e.g., email, networked databases). Transactional-system investments (23% of the total) were related to better performance as measured by improved Return on Assets (ROA) and lower nonproduction labor costs adjusted for sales. But strategic-systems investments (which had a negative effect in the short run) and information systems were unrelated to any performance measure.

Scientists who use computer networks extensively claim that those networks have already enhanced their productivity and changed the nature of their work [16, 20, 21, 24]. Despite such claims and confident predictions for the future, there has been no systematic empirical research on how computer network support for science is related to important scientific outcomes, including productivity [4, 8, 9, 11].[1] We therefore decided to do such a study and collected data from ocean scientists who use SCIENCEnet.™ Research cannot prove that network use causes productivity to increase. Experimental control through random assignment of scientists to network and no-network conditions is infeasible. Longitudinal observations track a moving target because networks and resources available for them are changing while populations of users are changing. Nonetheless we were able to document the association between network use and scientific productivity. We also documented how one scientific network operates in science today. Our research suggests technical and policy issues that are important for the future.

To investigate the benefits of network support in any science we must ask: what kind of access to what kind of resources? From the scientist's point of view, resources can be apparatus, databases, services, or colleagues. Access can vary in the degree to which it is direct, readily available, and free. Access is direct if not mediated through third parties such as service bureaus (or even graduate students). It is readily available if the scientist has access from any location. It is "free" if the costs are subsidized or not charged to the scientist. Computer scientists at research institutions who use the Internet typify one scientific user profile: usually direct, available, "free" access to colleagues and modest access to special apparatus or services (e.g., special processors or fabrication). Computer scientists make relatively little use of network-based scientific or bibliographic databases and services. Other scientific disciplines have different profiles.

We also must ask which scientists benefit. The case for a benefit to science from networking rests on two premises. The foremost is an efficiency premise. Networks can decrease the time scientists spend in unproductive activities such as traveling to special apparatus or data analysis centers, searching databases by hand, or tracking down library references. With a constant research time budget, the time saved could be

[1]Previous research has addressed what kinds of scientific information flow across different communication channels including email [4]; how electronic communication affects the formation of interpersonal ties among scientists [8]; and how electronic communication affects scientific group interaction and collaboration [9]. Hiltz [11] examined scientists' impressions of how their work was affected by participation in a scientific teleconference.

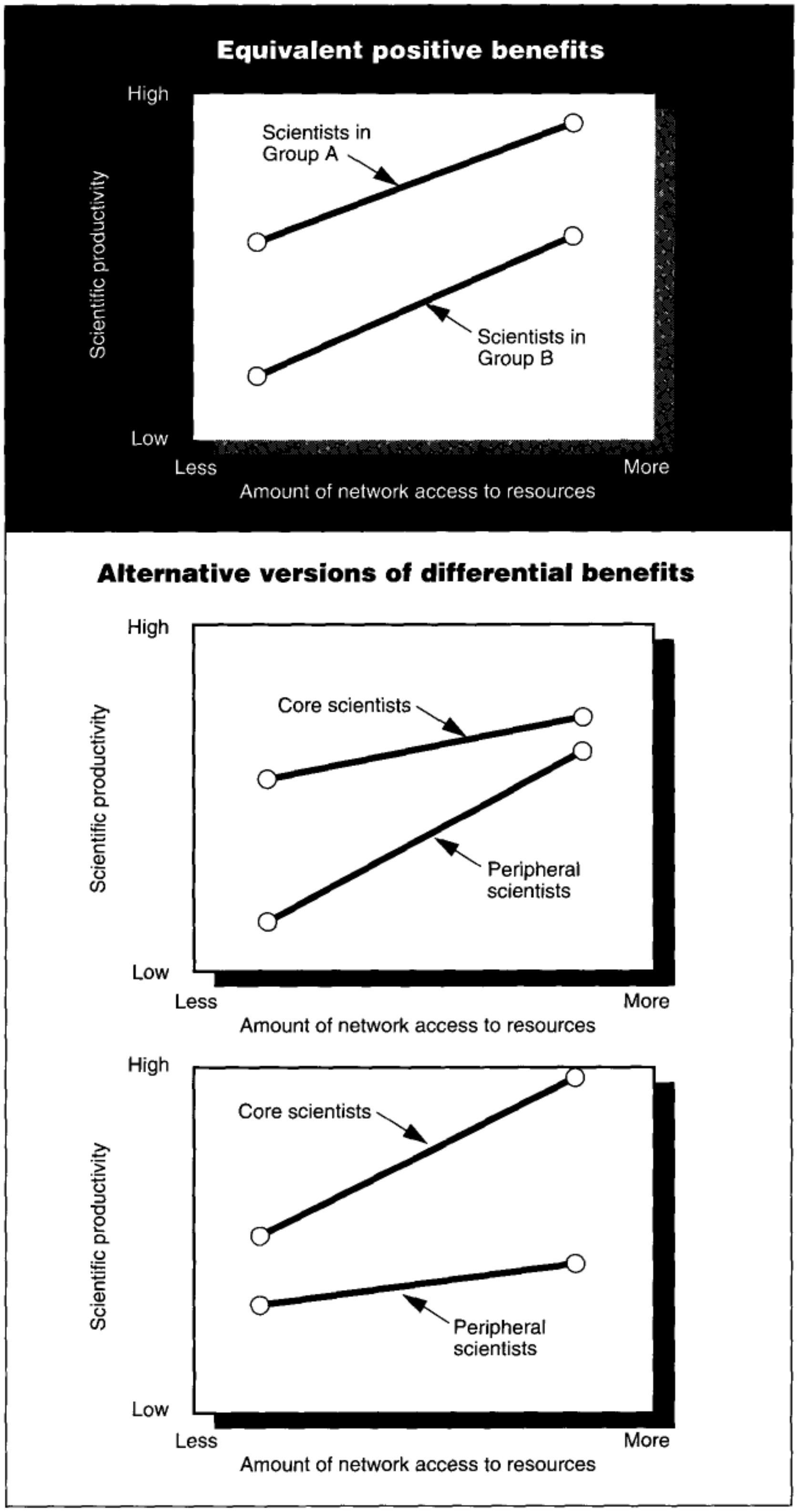


**Figure 1.** The equivalent and differential benefits arguments for returns to science from networking

put to more productive use. Access to new technology also permits scientists to use new techniques and study new problems. The efficiency premise implies that different scientists would benefit equally from network access to resources. Statistically, curves expressing the relationship between network access to resources and scientific outcomes for two different groups of scientists would have the same positive slope, though their intercepts might differ (see the top graph in Figure 1).

The second premise, a more controversial one, is that network access to resources confers differential benefits to different scientists. The scientific manpower version of this argument observes that a substantial fraction of the nation's Ph.D.'s have little or no access to good research resources because they are outside the mainstream or core of research activity. Some scientists, for instance, are located in four-year colleges and other teaching institutions. Network access to instruments, data, and research colleagues for these institutionally or geographically peripheral scientists would lead to a step function increase in their scientific productivity, an increase greater than that for core research scientists. (Presumably those in the core, particularly senior investigators at Ph.D.-granting universities and top research centers, already have ready access to instruments, data, and research colleagues.) Statistically, the curve for peripheral scientists would have a steeper slope than that for core scientists. An alternative view of differential benefits is that scientists who are already highly productive will be best able to take advantage of network-based resources: this is a "rich-get-richer" process. Statistically, the slope for core scientists would be steeper than that for peripheral scientists (see the middle and bottom graphs of Figure 1).

Here we examine the relationship between computer network support and scientific outcomes—whether generally positive and whether equivalent or differential—in one scientific community, a large group of ocean scientists. This scientific community differs somewhat from the computer science research community. For computer science, studies of other scientific groups contribute an empirically grounded "users' view," that is demand-side information on issues in connectivity, services, usability, bandwidth, applications, and science policy. Our particular examination of potentially differential benefits brings attention to questions of leverage, that is, of who will reap returns from networking.

## Oceanography and Network Support

In all scientific disciplines, the phenomena that scientists study influence how their work is organized and carried out. Physical oceanographers study fluxes in the world's oceans—for example, in temperature, salinity, and dissolved $CO^2$ over time and place. These phenomena require *in situ* measurement from research vessels, commercial ships, moored and drifting buoys, airplanes, and satellites. Many research questions require data from extremely remote locations, such as the polar ice caps and the equatorial Pacific. Many research questions require coordinated data collection across long distances and thus depend on large-scale projects involving scientists from different countries. For example, in one such study, the World Ocean Circulation Experiment, hundreds of ocean researchers from many countries are conducting a decade-long observational program to determine the general circulation of mass and heat in the world's oceans.

Oceanography is an empirical data-dense science. Large databases, for instance, on surface currents, upwelling and downwelling, prevailing winds, salinity, and biological activity are common. Terabyte-sized databases are becoming familiar. Modelers and theoreticians rely on supercomputers for data analysis, primarily those at the National Center for Atmospheric Research in Boulder, Colorado, Florida State University, and Princeton University. Many oceanographers, however, maintain small datasets and do their computing locally. A recent report on data management for oceanography noted that in one large project, despite international data centers, data coordinators, and approved data formats, "scientists have their own data on personal computers in their favorite spreadsheet database" [26].

Network support for research oceanographers is provided through Bitnet, the Internet, and largely by SCIENCEnet, which currently has about 4,000 subscribers. SCIENCEnet is operated by Omnet™, a firm established by a physical oceanographer in 1980. SCIENCEnet provides value-added services over commercial networks for access to scientific resources.

The nature of oceanography research influences the character of SCIENCEnet. Oceanography is a global science, so SCIENCEnet supports mailboxes in 45 countries and distant locations such as Kuwait and Antarctica. Mailboxes are accessible on board ship via satellite links. Oceanographers coordinate large projects, so SCIENCEnet maintains over 200 group project distribution lists. (See Figure 2 for examples.) Oceanography is data dense, so SCIENCEnet provides access to database indices and infrastructure for large datasets. Oceanographers depend on access to distant phenomena and colleagues, so the community is willing to pay for special network support tailored to these needs. Subscribers pay a monthly fee and usage charges to access SCIENCEnet resources. Most subscriptions are supported by scientists' research grants or by their departments or institutes.

SCIENCEnet provides mediated, readily available, fee-for-service access to database indices, services, and colleagues. It does not provide access to apparatus such as data sensors or supercomputers. Oceanographers use the Internet for remote access to supercomputers.

## A Survey of Network Usage and Scientific Outcomes

To describe what sorts of ocean scientists use SCIENCEnet for what purposes and to examine hypotheses about the relationships between network usage and scientific outcomes,

| | | |
|---|---|---|
| ***Apparatus*[a]** | | |
| | Alvin.Planning | Expedition plans and submersible time for the Alvin deep submersible. |
| | Ship.Sched92 | Research vessel time. |
| | Ocean | Cray time proposal reminders and deadlines. |
| ***Databases and Databoards*** | | |
| *Raw data*[b] | | |
| | OCEANIC | Ship schedules, oceanographic software, sea surface temperature maps. |
| *Metadata* | | |
| | TOGA.Subsurface | Meta-data on subsurface in-situ data. |
| | Quake.Alarm | Data reports from the National Earthquake Information Center. |
| | Melvyl | The Scripps Library catalog system. |
| *Infrastructure data* | | |
| | Seafood.Safety | Information and bibliographies on seafood contaminants. |
| ***Services*** | | |
| | Omnet.Fax | To send a fax anywhere in the world. |
| | Paper Mail | To have hard copy mailed in the US or Canada. |
| | OCE.review | To submit a proposal review to the NSF Ocean Division. |
| | Sabre | Airline schedules. |
| | Directory | Subscriber directory. |
| | OCEANIC | Gateway to data services at Ocean Network Information Center. |
| | ECS | Gateway to satellite data services at National Climatic Data Center. |
| ***Research Colleagues*** | | |
| *Email* | | |
| *Groups/Projects* | | |
| | TOGA | Tropical Ocean/Global Atmosphere program. |
| | WOCE | World Ocean Circulation Experiment. |
| | JGOFS | Joint Global Ocean Flux Experiment. |
| | CEES | Committee on Earth and Environmental Sciences. |
| | ERS-1 | Environmental Research Satellite (Joint NASA/European Space Agency). |
| *Bulletin Boards* | | |
| | Ocean | A general disciplinarybboard; also some meta-data. |
| | Jobs | Openings and availability. |
| | ENSO.Info | Discussions of the El Niño phenomenon. |
| | I.Witness | First-hand accounts from Antarctica, Leningrad. |
| | Grad.Students | Information for oceanography graduate students. |

[a]SCIENCEnet does not provide direct access to apparatus – buoys, supercomputers, etc.
[b]SCIENCEnet does not maintain raw scientific data on-line because of their size.

**Figure 2.** Examples of scientific resources for oceanographers accessed through SCIENCEnet

we collected survey data from a stratified sample of 338 SCIENCEnet users. This sample had three groups: 150 scientists randomly selected from individual subscribers whose network usage was above the median usage rate, 150 scientists randomly selected from subscribers whose network usage was below the median usage rate, and the 38 of the 50 most active users remaining after the random sample was drawn. (Our purpose in oversampling top users was to bolster our descriptive data on scientists whose usage was heaviest.) Omnet posted notices of the survey on the OCEAN b-board, and we sent the survey by email to every sample member's SCIENCEnet mailbox. Respondents either completed and returned the survey electronically, or they asked for a hard copy and returned it by postal mail. The 93-item survey asked respondents about their professional background, research activities, and network use. (See Appendix for details of the method.)

Respondents provided information about their use of email, remote databases, programs run remotely, bibliographic databases, and remote data transfers. Because these data are more fine-grained than aggregate network usage data, we used these answers to define our primary independent variable. The self-reports correlate 0.50 ($p<0.001$) with actual usage rates over the three-month period used to draw the sample. Based on our interviews, we used three other independent variables to examine whether returns from network use to core and peripheral scientists are the same or different: seniority, subfield, and geographic location. We assumed that core scientists have had more years since their Ph.D.'s than peripheral scientists. Since the oceanographic community served by SCIENCEnet consists primarily of physical oceanographers, we assumed core scientists are more likely to be physical oceanographers (the majority) than scientists from other fields such as marine biology. Given the phenomena studied and resources needed (such as ocean-going vessels), we assumed that core scientists are more likely to be employed by institutions located on a coast than inland.

Our dependent variables were three kinds of scientific outcomes: scientific publication, professional recognition, and social integration. We measured publication using the number of papers published in refereed journals in the previous year. We measured professional recognition using the sum of four questions about the past five years: if respondents had won a scientific award, served on a major professional committee, served on an editorial board, or served on a national advisory committee. We measured social integration using the number of physical oceanographers known by the respondent. (Table 1 lists other variables used to describe the sample; see the Appendix.)

## Response Rate

We received 257 completed surveys from our sample, a total response rate of 76%. Response rates by group were 84% for the top usage group (32 of 38), 84% for the above-median usage group (126 of 150), and 66% for the below-median usage group 99 of 150. The below-median group, took longer to respond and answered more often using postal mail rather than email ($X^2(2) = 9.77$, $p<0.01$).[2]

## Who Uses This Network?

SCIENCEnet users are active and productive scientists. (See Table 1.) Approximately three-quarters of the

**Table 1.** Sample characteristics

| | Usage groups: Top (n = 32) | Usage groups: Above median (n = 126) | Usage groups: Below median (n = 99) | Statistical significance (p level)[a] |
|---|---|---|---|---|
| **Personal characteristics** | | | | |
| Ph.D. | 75% | 81% | 70% | n.s. |
| Seniority (years since Ph.D.) | 16.3 | 15.5 | 13.7 | n.s. |
| Principal Investigator | 68% | 79% | 69% | n.s. |
| Researcher (< 50% administration) | 57% | 83% | 82% | <.01 |
| Gender (percent male) | 74% | 91% | 91% | <.01 |
| **Institutional characteristics** | | | | |
| Miles from coast | 151 | 76 | 92 | n.s. |
| Foreign (% non-U.S.) | 16% | 14% | 10% | n.s. |
| Prestige (1 = low to 3 = high) | 1.8 | 2.0 | 1.6 | <.01 |
| **Research characteristics** | | | | |
| Type of research (% of time) | | | | |
| Collect/analyze data | 32% | 37% | 36% | n.s. |
| Numeric modeling | 7% | 10% | 16% | <.05 |
| Theoretical work | 3% | 9% | 12% | <.05 |
| Development of tools | 10% | 15% | 13% | n.s. |
| Planning/administration | 43% | 27% | 22% | <.01 |
| **Source of data (% of data)** | | | | |
| Shipboard instruments | 35% | 38% | 37% | n.s. |
| Unattended instruments | 10% | 22% | 23% | n.s. |
| Remote sensors & satellites | 29% | 20% | 14% | <.05 |
| **Scientific outcomes** | | | | |
| Articles published last year | 1.5 | 2.7 | 1.8 | <.01 |
| Recognition (1 = low to 4 = high) | 2.0 | 1.7 | 1.1 | <.01 |
| Oceanographers known | 145 | 72 | 46 | <.01 |
| **Network history and usage** | | | | |
| Years on SCIENCEnet | 4.8 | 3.7 | 2.5 | <.01 |
| Resources used (1 = low to 4 = high) | 2.0 | 1.8 | 1.8 | n.s. |
| Messages/day SCIENCEnet | 14.3 | 4.8 | 1.9 | <.01 |
| Messages/day other networks | 1.1 | 1.0 | 2.9 | <.01 |
| **Use of email[b]** | | | | |
| Disciplinary committees | 64% | 64% | 44% | <.01 |
| Interactions with sponsors | 78% | 68% | 50% | <.01 |
| Submit articles and proposals | 43% | 43% | 40% | n.s. |
| Browsing: new ideas and people | 56% | 47% | 44% | <.05 |
| Facilitate collaboration | 94% | 90% | 83% | <.01 |
| Contact with home | 51% | 40% | 40% | <.05 |

[a]*The (probability) levels for statistical significance beyond chance (at .05) are derived from statistical tests comparing mean scale values across the three usage groups. When possible, we list percentages of respondents instead of scale values.*
[b]*We show the percentage of each group that used email for these purposes.*

respondents reported having earned a Ph.D.; more than three-quarters do research at least half-time, and more than half work at academic or non-profit research institutions. Almost three-quarters said that they were currently the principal investigator on a research project. They reported publishing an average of 2.3 papers in refereed journals within the previous year. More than half had served on a major professional committee within the past five years; more than a third had served on national advisory boards; a fourth had served on editorial boards; and almost a fifth had received one or more scientific awards.

Respondents on average reported sending and receiving 3.5 messages per day on SCIENCEnet and an additional 1.8 messages on other networks. During the previous year, 77% used networks for data transfer; 39% accessed remote scientific databases; 37% searched a remote bibliographic database; and 24% ran remote programs.

Frequent users (that is, sample respondents in the top and above-median usage groups) are the more active, productive scientists. As compared with infrequent users, they work at more prestigious institutions, have received more professional recognition, have published more, and know more oceanographers. Frequent users also show a different pattern of network use than infrequent users. They have subscribed to SCIENCEnet longer, send and receive more email messages per day on SCIENCEnet, collect more data from remote sensing satellites, and spend more time doing research administration. In contrast, infrequent users tend to use other networks more and spend more time on numeric modeling and theoretical work.

Because we sampled network users and not oceanographers generally, we must ask how network users compare with scientists from the larger oceanographic community. To make these comparisons, we used data from a 1987 sample survey of more than 5,700 ocean scientists and engineers conducted by the American Geophysical Union (AGU). Our sample contained more physical oceanographers than exist in the general population of ocean scientists (61% vs. 21%; $X^2(1) = 156$, $p<0.001$). Using weighted data, we found no difference in the ratio of males to females, but our sample contained more Ph.D.'s ($X^2(1) = 24$, $p<0.01$), more researchers ($X^2(2) = 32$, $p<0.01$), and more oceanographers at academic or nonprofit research institutions ($X^2(3) = 43$, $p<0.01$). Thus, our sample of networked oceanographers differed from the general population of ocean scientists. The network sample better represents the subfield of physical oceanography and also is more research intensive.

## Network Usage and Scientific Outcomes

We used multivariate regression techniques to examine the statistical relationships between network usage and scientific outcomes. Regression techniques evaluate the association between an observed dependent variable—in this case scientific outcomes—and one or more independent variables while controlling for the influence of extraneous variables. We controlled for personal characteristics such as number of years since a Ph.D. and institutional characteristics such as the prestige of the scientist's institution, which could influence measured scientific outcomes independent of any network use. Our equations are a tool for assessing the statistical significance associated with the slopes of the lines in Figure 1 while statistically controlling for the influence of variables unrelated to the hypotheses we want to test. To obtain a representative sample of scientists, these analyses used only data from Ph.D.'s in the above-median and below-median usage groups.[3]

We conducted the analyses as simultaneous solutions in which independent variables were evaluated as if they had been entered last in the equations. The explanatory value of each table is indicated by the $R^2$ (variance accounted for) and a positive $F$ test with a $p$ level at 0.05 or better. Within the tables, the beta coefficients show the magnitude of each independent variable's association with the dependent variable. Statistically significant coefficients for the variables in any equation can be interpreted as evidence of equivalent positive benefit (or negative effect, if there is a minus sign). A statistically significant coefficient for any characteristic and the statistical interaction between network usage can be interpreted as evidence of differential benefit. The signs of the interaction coefficients signify the direction of apparent differential benefit. In our analyses, where the sign is negative, the benefit is greater to peripheral scientists (e.g., young Ph.D.'s).

We found that network usage is positively associated with all three scientific outcomes. (See Table 2.) Scientists who use the network more also produce more papers, receive more professional recognition from their peers, and know more physical oceanographers.[4] In Table 2, eight of the nine interaction terms are negative, which is consistent with differential benefit for peripheral scientists even though only two interaction coefficients are statistically significant. One of the statistically signifi-

[2]Besides these responses from the stratified deliberate sample, we received an unexpected return of completed surveys from 104 persons who were not in the sample. We assume they saw the b-board announcement of the survey; they requested a copy and sent it to us by email. We report our analysis comparing the deliberate sample to these "volunteers" elsewhere [25].

[3]These restrictions omit two groups: non-Ph.D.'s and those in the top user group who were not in the stratified random sample. By restricting the analyses to Ph.D. scientists, we neglect mainly support staff and technicians. Support staff are crucial in the production of science, but their inclusion in our analyses would have muddied examination of equivalent and differential benefits. Furthermore, we wished to represent patterns of behavior and outcomes among those who most influence the direction of science, that is, Ph.D. scientists. By restricting our analyses to those in the stratified random sample, we obtain results that can be generalized to the (Ph.D., networks) population as a whole. Such restriction is common in sociology; the stratified random sample represents the most conservative estimate of the target population. If we had included the other 32 top users in our analyses, we would have overrepresented outlying observations relative to the random sample, thereby threatening interpretation and assumptions of linearity.

[4]This pattern held when each outcome was regressed separately and when combined in a general multivariate regression (Wilks' Lambda = 0.66, $F(10,130) = 6.71$, $p < 0.001$).

**Table 2.** Relationships between Network usage and scientific outcomes

| Independent variable | Beta | Statis. significance ($p$ level) |
|---|---|---|
| **Dependent variable: Articles published** — $R^2 = .24$; Adjusted $R^2 = .19$; $F(10.152) = 4.9$. $p<.001$ | | |
| Researcher (more than 50% time) | .07 | .001 |
| Seniority (years since Ph.D.) | .00 | n.s. |
| Subfield is physical oceanography | −.09 | n.s. |
| Prestige of institution | .07 | n.s. |
| Location of institution on coast | .10 | n.s. |
| Seniority squared (curvilinearity) | −.19 | .025 |
| SCIENCEnet usage (log) | .28 | .001 |
| Seniority X SCIENCEnet interaction | −.05 | n.s. |
| Subfield X SCIENCEnet interaction | .05 | n.s. |
| Coastal location X SCIENCEnet interaction | −.14 | .07 |
| **Dependent variable: Professional recognition** — $R^2 = .26$; Adjusted $R^2 = .21$; $F(10.151) = 5.2$. $p<.001$ | | |
| Researcher (more than 50% time) | .11 | n.s. |
| Seniority (years since Ph.D.) | .36 | .001 |
| Subfield is physical oceanography | −.05 | n.s. |
| Prestige of institution | −.02 | n.s. |
| Location of institution on coast | −.04 | n.s. |
| Seniority squared (curvilinearity) | .01 | n.s. |
| SCIENCEnet usage (log) | .27 | .001 |
| Seniority X SCIENCEnet interaction | −.14 | .07 |
| Subfield X SCIENCEnet interaction | −.05 | n.s. |
| Coastal location X SCIENCEnet interaction | −.05 | n.s. |
| **Dependent variable: Oceanographers known** — $R^2 = .35$; Adjusted $R^2 = .30$; $F(10.142) = 47.5$. $p<.001$ | | |
| Researcher (more than 50% time) | .04 | n.s. |
| Seniority (years since Ph.D.) | .02 | n.s. |
| Subfield is physical oceanography | .48 | .001 |
| Prestige of institution | .07 | n.s. |
| Location of institution on coast | .02 | n.s. |
| Seniority squared (curvilinearity) | .04 | n.s. |
| SCIENCEnet usage (log) | .31 | .001 |
| Seniority X SCIENCEnet interaction | −.03 | n.s. |
| Subfield X SCIENCEnet interaction | −.05 | n.s. |
| Coastal location X SCIENCEnet interaction | −.01 | n.s. |

*Note.—In this and the following tables, the explanatory value of the table is indicated by the $R^2$ (variance accounted for) and a positive F test with a p level at about .05 or less. Within the tables, the beta coefficients indicate the magnitude of each independent variable's association with the dependent variable. Statistical significance of a single variable or interaction in any table is tested using a t test. Statistically significant coefficients for the variables in any equation can be interpreted as evidence of equivalent positive benefit (or negative effect, if there is a minus sign). A statistically significant coefficient for any characteristic X network usage interaction term can be interpreted as evidence of differential benefit. The signs of the interaction terms signify the direction of apparent differential benefit. In our analyses, where the sign is negative, the benefit is greater to peripheral scientists (e.g., young Ph.D.'s).*

cant coefficients shows the relationship between SCIENCEnet usage and publication is stronger (i.e., has a steeper slope) for scientists at inland institutions than for scientists at institutions on a coast. Traditionally, oceanography has been a discipline dominated by institutions located on the coasts. Others [7, 10, 12] have noted that network communication can help overcome disadvantages of geographic distance. Our results suggest that this process may be particularly important for geographically peripheral ocean scientists. The other statistically significant coefficient shows that the relationship between SCIENCEnet usage and professional recognition is stronger for younger scientists than for more senior scientists, another finding consistent with differential benefit for peripheral scientists.

Other statistical effects unrelated to network use are unsurprising. Being primarily a researcher rather than administrator is associated with the number of papers published in the previous year. Seniority (years since Ph.D). has a curvilinear relationship with publication. (This finding accords with much other research showing that successful scientists often hit their peak production 5 to 15 years after the Ph.D. and then gradually slow down.) Seniority also predicts professional recognition. Being a physical oceanographer rather than in some other subfield of oceanography predicts the number of physical oceanographers known.

Because communicating with colleagues (email, b-boards, groups, projects) was the most used and reportedly useful feature of SCIENCEnet, we wanted to know more about the purposes of this communication and how these purposes might be associated with more or less positive scientific outcomes. We evaluated six categories of uses for communication: interaction with sponsors (such as negotiating an NSF grant), professional committee work in the discipline (such as work on a graduate standards committee), submitting manuscripts or proposals electronically, browsing to learn about new techniques or what people are doing, collaboration in groups and projects, and contact with home (or office) while traveling. (See Appendix for a description of how the categories were measured and constructed into six factors.) Taken together the results suggest that network communication within active work groups and work relationships (committees, projects, coauthors, sponsor-PIs) shows a positive return to the scientific outcomes we measured, whereas more individualistic networking, such as browsing b-boards and contacting home, has no or negative relationships with those scientific outcomes. (An exception: browsing b-boards is associated with knowing more oceanographers.)

## Network Usage and Scientific Work

Although most of our analyses concerned the relationship between network use and scientific outcomes, we also studied how scientific methods and techniques are changing. In oceanography, computing advances have enabled satellite data collection and have encouraged more scientists to do theoretical and modeling work. Networks also have given oceanographers access to new techniques, most notably, remote sensing of geophysical phenomena.

We found that network usage was associated with remote sensing, but networks other than SCIENCEnet play a more important role than SCIENCEnet does. (See Table 4). The Ph.D. years/SCIENCEnet interaction is statistically significant, again suggesting differential benefit. The positive sign of the interaction coefficient shows that senior scientists, instead of younger ones, are differentially benefitted by networking. Studies of scientific innovation suggest that older scientists, on average, adopt new techniques and ideas after younger ones and thus are at a disadvantage compared with younger scientists when these innovations are productive [6, 15, 19]. Therefore these results can be taken as consistent with differential benefit for peripheral scientists.

Other predictors of remote sensing are as expected. Being primarily a researcher rather than an adminis-

**Table 3.** Relationships between kinds of email use and scientific outcomes

| | Dependent variable: Articles published | |
|---|---|---|
| | $R^2$ = .22; Adjusted $R^2$ = .16; $F(10.143) = 3.9$. $p<.001$ | |
| **Independent variable** | **Beta** | **Statis. significance ($p$ level)** |
| Researcher (more than 50% time) | .22 | .01 |
| Seniority (years since Ph.D.) | −.01 | n.s. |
| Prestige of institution | .08 | n.s. |
| Seniority squared (curvilinearity) | −.13 | n.s. |
| Email for discipline committee work | .28 | .001 |
| Email for interactions with sponsors | .09 | n.s. |
| Email for document submission | −.09 | n.s. |
| Email to facilitate collaboration | .06 | n.s. |
| Email for browsing: new ideas and people | −.11 | n.s. |
| Email to contact home | −.00 | n.s. |

| | Dependent variable: Professional recognition | |
|---|---|---|
| | $R^2$ = .31; Adjusted $R^2$ = .26; $F(10.142) = 6.3$. $p<.001$ | |
| **Independent variable** | **Beta** | **Statis. significance ($p$ level)** |
| Researcher (more than 50% time) | .04 | n.s. |
| Seniority (years since Ph.D.) | .31 | .001 |
| Prestige of institution | −.06 | n.s. |
| Seniority squared (curvilinearity) | .05 | n.s. |
| Email for discipline committee work | .35 | .001 |
| Email for interactions with sponsors | .03 | n.s. |
| Email for document submission | −.12 | .09 |
| Email to facilitate collaboration | .03 | n.s. |
| Email for browsing: new ideas and people | −.16 | .05 |
| Email to contact home | −.04 | n.s. |

| | Dependent variable: Oceanographers known | |
|---|---|---|
| | $R^2$ = .14; Adjusted $R^2$ = .08; $F(10.136) = 2.3$ $p<.02$ | |
| **Independent variable** | **Beta** | **Statis. significance ($p$ level)** |
| Researcher (more than 50% time) | .02 | n.s. |
| Seniority (years since Ph.D.) | .09 | n.s. |
| Prestige of institution | .11 | n.s. |
| Seniority squared (curvilinearity) | .10 | n.s. |
| Email for discipline committee work | .16 | .08 |
| Email for interactions with sponsors | .14 | n.s. |
| Email for document submission | .07 | n.s. |
| Email to facilitate collaboration | .07 | n.s. |
| Email for browsing: new ideas and people | .18 | 0.5 |
| Email to contact home | −.05 | n.s. |

trator, being a physical oceanographer rather than in another subfield, and working at an inland geographic location (mainly in facilities of NASA and the U.S. Navy, for example) predicts the use of remote-sensing techniques.

## Discussion

This article reports the first systematic assessment of the value of computer networking for a scientific community. In contrast with the numerous failures to document empirically the benefits of computer technology for productivity in industry, we found evidence consistent with such benefit in science. Within the SCIENCEnet community of oceanographers, network usage is consistently associated with desirable scientific outcomes. The observed associations do not prove causality, but they do demonstrate that productive scientists use the network more often and for more purposes than unproductive scientists do. SCIENCEnet has become a critical part of the infrastructure of doing ocean science.

Our study also suggests that the benefits of computer networks for science are not predictable from a simple efficiency model of technology. Networks may have differential payoffs for peripheral and core scientists. "Peripheral" in our study did not mean unimportant, but removed from (or facing barriers to) those resources necessary for doing good oceanography—remote instruments, geophysical data, global projects, disciplinary committees, important research programs, and colleagues. To test for differential benefits we categorized oceanographers as to their position within the social and work structure of the SCIENCEnet community. In oceanography, to be more rather than less peripheral is to be young, to be landlocked, or to have been trained before the development of new methods such as remote sensing. We found some evidence that, at the margin, peripheral oceanographers benefit more from network usage than core oceanographers do. This lends credence to the theoretical pattern described in the middle graph of Figure 1. This argument presumes that core scientists already have relatively reliable and abundant access to scientific resources and therefore do not profit from network connections as much as peripheral scientists do.

We observed that SCIENCEnet has been developed for a subset of specialized activities—those tasks not shared with other sciences and not supported by more general-purpose networks such as the Internet. One of these stands out in our data: long-distance group management. Oceanography, like physics, is "big science." Doing oceanography demands coordination, participation of diverse researchers, and collaborative groups. Yet the movement, dispersion, and distance of researchers from one another is greater than in physics. SCIENCEnet supports a continuing infrastructure for communication within groups and projects, such as keeping directories and memberships up to date and helping group members keep in touch with their groups when traveling to remote locations.

**Table 4.** Relationship between Network usage and remote sensing

| Independent variable | Beta | Statis. significance ($p$ level) |
|---|---|---|
| | **Dependent variable: Remote sensing** | |
| | **$R^2$ = .34; Adjusted $R^2$ = .28 $F(10.152) = 4.9$. $p<.001$** | |
| Researcher (more than 50% time) | −.24 | .01 |
| Seniority (years since Ph.D.) | −.08 | n.s. |
| Subfield is physical oceanography | .25 | .01 |
| Prestige of institution | .10 | n.s. |
| Location of institution on coast | −.21 | .01 |
| SCIENCEnet usage (log) | .15 | .09 |
| Usage on other networks | .37 | .001 |
| Seniority X SCIENCEnet interaction | .21 | 0.5 |
| Subfield X SCIENCEnet interaction | .05 | n.s. |
| Coastal location X SCIENCEnet interaction | −.09 | n.s. |

SCIENCEnet also supports data retrieval. Oceanography, as physics, evaluates large datasets generated from large streams of noisy, unpredictable observations and technical uncertainty surrounding their acquisition. To evaluate such data requires considerable infrastructure—raw-data repositories, software development and maintenance, services for putting datasets on tape or on-line, programs to process the data. SCIENCEnet's on-line services for aiding data retrieval range from subsciber b-boards to indices, processing algorithms, schedules, and special metadatabases.

SCIENCEnet does not store the raw data though. In other disciplines, efforts are underway to provide network access to nationally shared databases. Such is the model for GenBank, the national repository for nucleic acid sequence data in biology [5]. Biologists are required to send sequences to GenBank when they publish the sequences in journals. This procedure allows other biologists to examine and copy the sequences and also lets journals print summaries instead of entire sequences. The situation in oceanography is very different because of the difference in the size of the databases. Ten gigabits (Gb) is not uncommon in a dataset, so oceanographers will look interactively only at summaries or metadata. Even metadatasets (such as velocities) typically are too large to retrieve efficiently through the network. To compound the problem, many scientists in oceanography would prefer pictures. Picture transmission, although likely to be in popular demand, will not become a common feature of most networks until problems of bandwidth and standards are solved.

Finally, what SCIENCEnet offers beyond computer network technology is the social and human factors element of usability. Omnet's philosophy is that oceanographers are busy doing oceanography and have no interest in computing itself. Therefore SCIENCEnet must offer usable facilities for doing science. With b-board posts, a newsletter, and subscriber survey, Omnet draws heavily on subscriber participation to improve SCIENCEnet over time. When a new program is introduced, Omnet gives subscribers personal as well as on-line and printed help. Omnet encourages subscribers to ask for and try new applications, but never pushes them to spend time learning programs, applications, or commands.

To improve usability, SCIENCEnet has many interactive help files on-line that also are easy to use (e.g, find out how to send a fax and then actually do it). Nevertheless, many scientists would rather send messages to Omnet.Service and wait for a person's reply than use instant on-line help. This is true even for directory information. The on-line directory of subscribers is easy to use and well advertised, but subscribers often send email to Omnet. The heavy traffic between subscribers and Omnet staff at first look seems inefficient, but there are some positive side effects. First, many subscribers do not know what their problem is, as indicated by the following response to our survey:

> They think they have a binary file transfer protocol problem, and will waste their time on the intricacies of Kermit. Actually they just need to capture their session as ASCII text. A human can listen and help them solve their problem, not just stream out text at them.

One of the subscribers' most frequent requests is to have bulletin board prompts set by Omnet staff. Subscribers could do this themselves, but Omnet knows more:

> For instance, a subscriber asks, "Please prompt me for JGOFS.Data." Our answer is, "OK but JGOFS.Data has been kind of inactive lately. It looks like a lot of the discussion is happening on JGOFS.News. Tell you what, I'll give you both."

Finally, Omnet can learn about its subscribers:

> Please prompt me for ENSO.Info.
>
> Will do. Are you working on ENSO?
>
> Yes, several of us are in a subgroup for ENSO instrumentation. But our discussions get kind of acrimonious. Just how private is the board?
>
> Not private at all. Would a mailing list be best for your subgroup?

All these findings should be interpreted very cautiously. For instance, our results do not show whether or not network resources would best be used for peripheral members of the greater scientific community. We studied just one discipline and within that, a cohesive community with a "critical mass" of people with like interests and aims. Our analyses of the AGU sample survey, which included many nonresearchers, technicians, and college teachers, suggests that the SCIENCEnet population is more elite and more narrowly research-oriented than the AGU-Ocean Section membership. The exclusivity of our sample suggests that the "peripherality" of its "peripheral scientists" is relative; in comparison with the average member of AGU, the entire SCIENCEnet membership would be considered members of oceanography's core. The productive use these scientists make of SCIENCEnet is likely to depend on their deep prior commitment to doing science. A peripheral member of the larger oceanographer community such as a college or high school teacher of ocean science might not have much use or time for b-boards that provide research vessel schedules and algorithms for interpreting satellite data.

Our results do suggest strongly that a generic scientist/user model is too simple to describe network needs and success. If a more differentiated view of scientists is necessary, then more attention should be paid to how scientists in the disciplines actually work—what their social structure looks like and what kinds of access they need to access what kinds of resources. Choices surrounding the development and support of networking resources should be based on grounded knowledge of scientific disciplines and scientists.

# Appendix

### Sample Frame and Sample

With the help of Omnet's administrators, we collected network usage data for the three-month period preceding our survey on the 1,356 active personal mailboxes (excluding group and institutional mailboxes) within Ocean, the largest division of SCIENCEnet (which also includes divisions for meteorologists, seismologists, and so forth). We selected random samples of 150 mailboxes each from the above-median usage and below-median usage strata. We also selected the 50 most frequently used mailboxes except for 12, which already had been included in the random sample. The result was a stratified random sample of 300 mailboxes, with an additional group of 38 top users.

### The Survey Instrument and Its Administration

On September 15, 1988, we sent a 93-item electronic questionnaire to the 338 sample network mailboxes. Recipients were given the choice of answering the survey on-line, downloading it for completion at a later time, or requesting that a hard copy be sent to them by postal mail. Meanwhile, a general announcement of the survey was broadcast publicly over the network on the Ocean general-interest electronic bulletin board. Email and b-board announcements described its purpose and noted that respondents would be entered in a $250 lottery. After two weeks, a prompt was mailed to sample nonrespondents by conventional mail, for those who had not checked their network accounts, or by email. Also, an announcement was posted again to the Ocean bulletin board. A second and final prompt to nonrespondents, following the same procedure, was sent one month later. Data collection terminated during the early months of 1989.

### Variables

The measures of network usage and scientific outcomes are described in the body of the article. We also wished to measure various aspects of "peripherality" or centrality in the production of science. Using demographic information in the survey and working with a member of the oceanographic community, we coded respondents' seniority, research fields (physical oceanographer or other), gender, and job types (researcher or administrator). Respondents who estimated they spent more than 50% of their time on research activities were coded as researchers; those who spent 50% or more of their time on administrative activities were coded as administrators.

For each respondent's institution or employer we recorded the number of miles from the nearest coast, whether the institution was contiguous or noncontiguous U.S., whether it was U.S. or international, and its type—academic, nonprofit, government, commercial, other. We rated each institution's prestige as low, medium, or high, based on expert-informant information. These variables were intended to measure various aspects of peripherality or centrality of scientists' institutions.

A factor analysis of 17 questions related to the use of email produced a coherent six-factor solution. Three factors described different kinds of audiences for electronic communication: (1) sponsors and government agencies, (2) professional committees in the discipline, (3) editorial and agency offices to which scientists sent manuscripts and proposals. Three additional factors described the utility of email for (4) browsing and encountering new research information and people, (5) conducting joint work, and (6) staying in touch with home (and office) while traveling. Sum scores for components, rather than factor scores, were used in analyses [14].

### Analysis of Response Bias

Although the overall response rate of 76% was quite good, the rate differed across sample subgroups. The above-median group had a higher representation than the below-median group (84% for above-median group; 66% for below-median group). With differential responding, inferences from the sample to the general network community may need to be adjusted to avoid bias. Weighting techniques work well in most situations but assume that nonrespondents across groups are similar to each other on critical dimensions [3]. Our nonrespondents did not meet that assumption. For example, nonrespondents in the above-median group were at the top of the above-median network usage distribution (typically the most active electronically), but nonrespondents in the below-median group were at the bottom of the below-median network usage distribution (the least active electronically), $F(1,298) = 4.88$, $p < 0.05$. Therefore, conventional weighting techniques are suspect.

We used existing information on nonrespondents to impute missing information through multiple-regression techniques [18]. Imputation is especially useful for portraying mean population values and underlying distributions, but the increase it brings in power is often purchased at the expense of Type I error rates [13]. Based on imputation analysis, the differential response rate decreased the representativeness of the sample dataset very little, changing the value of a variable more than 10% in only 3 out of 129 cases (3 groups × 43 variables). Our sample dataset underrepresents oceanographers employed by business and commercial organizations (actual respondents = 8%; imputed respondents = 31%) and overrepresents oceanographers employed by the government, either in labs or funding agencies (actual respondents = 29%; imputed respondents = 10%). Our dataset also underrepresents oceanographers who use networks to run remote programs (actual respondent = 24%; imputed respondents = 47%).

We used multiple-regression analyses to investigate the relationships between network usage and scientific outcomes. Control variables (covariates) and peripherality measures were included in the analyses based on previous research and on their *a priori* potential for covariance with outcomes. Number of years since Ph.D. was included because it is known to predict scientific outcomes (thus, should be used as a covariate) and because it can be used as an indicator of peripherality. Other control variables included geographic location, institutional prestige, job type, and subfield of research. The small numbers of female and international respondents (3% and 8% respectively) precluded using these categories in analyses.

Regression analyses are suitable for continuous variables. To use categorical data in the regressions, we created dichotomous dummy variables [23]. Geographic peripherality was defined as working at an institution that was located more than five miles inland (the median value for miles from the coast) or that was located outside of the contiguous U.S. Subfield peripherality was defined as any field or subfield other than physical oceanography. Log transformations were applied to skewed variables, that is, number of papers published in previous year and number of oceanographers known. Following Allison's [1] suggestions, all models were specified to include controls and single-order effects, and quadratic terms were included when interactions were statistically significant in order to test for curvilinearity. ■

**References**


1. Allison, P. D. Testing for interaction with multiple regression. *Am. J. Soc. 83* (1977), 144–153.
2. Attewell, P. Information technology and the productivity paradox. In the *National Research Council, Panel on Linkages Between Individual and Organizational Productivity, Committee on Human Factors.* Washington, D.C. To be published.
3. Aziz, F. and Scheuren, F. *Imputation and Editing of Faulty or Missing Survey Data.* U.S. Department of Commerce, Washington, D.C., 1978.
4. Carley, K. and Wendt, K. Electronic mail and scientific communication: The study of SOAR and its dominant users. *Knowledge 12* (1991), 406–440.
5. Cinkosky, M. J., Ficket, J. W., Gilna, P. and Burks, C. Electronic data publishing and GenBank. *Science 31* (1991), 1273–1277.

6. Coleman J. et al. *Medical Innovation: A Diffusion Study.* Bobbs-Merrill, New York, 1966.
7. Eveland, J. D. and Bikson, T. K. Work group structures and computer support: A field experiment. *ACM Trans. Off. Inf. Sys. 6*, 4 (1988), 354–379.
8. Freeman, L. The impact of computer based communication on the social structure of an emerging scientific specialty. *Soc. Networks 6* (1984), 201–221.
9. Harasim, L. M. and Winkelmans, T. Computer-mediated scholarly collaboration. *Knowledge 11* (1990), 382–409.
10. Hesse, B. W., Werner, C. M. and Altman, I. Temporal qualities of computer-mediated communication. *Comput. Hum. Behav. 4* (1988), 1–19.
11. Hiltz. S. R. *Online Communities: A Case Study of the Office of the Future.* Ablex, Norwood, N.J., 1984.
12. Huff, C., Sproull, L. and Kiesler, S. Computer communication and organizational commitment: Tracing the relationship in a city government. *J. Appl. Soc. Psychol. 19*, 6 (1989), 1371–1391.
13. Jinn, J. H. and Sedransk, J. Effect on secondary data analysis of common imputation methods. *Soc. Methods 19* (1989), 213–241.
14. Kim, J. and Mueller, C. W. *Factor Analysis: Statistical Methods and Practical Issues.* Sage University Paper Series on Quantitative Applications in the Social Sciences, vol. 07-0124. Sage University, Beverly Hills, Calif., 1978.
15. Kuhn, T. S. *The Structure of Scientific Revolutions.* University of Chicago Press, Chicago, 1970.
16. Lederberg, J. Digital communications and the conduct of science: The new literacy. *IEEE Proceedings 66*, 11 (1978), 1314–1319.
17. Lederberg, J. and Uncapher, K. Toward a national collaboratory: *Report of an invitational workshop.* Rockefeller University, New York, 1989.
18. McLaughlin. D. and Wyse, L. Guidebook for imputation of missing data, contract No. 300-78-150. The National Center for Education Statistics, Washington, D.C., 1980.
19. Messeri, P. *Social Change and the Life Course.* Vol. 1. Sage, Newbury Park, Calif., 1988.
20. National Research Council, Panel on Information Technology and the Conduct of Research, Committee on Science, Engineering, and Public Policy. *Information Technology and the Conduct of Research: The User's View.* National Academy Press, Washington, D.C., 1989.
21. Newell, A. and Sproull, R. Computer networks: Prospects for scientists. *Science 215* (1982), 843–852.
22. Palco, J. Getting together bit by bit. *Science 248* (1990), 160–162.
23. Pedhazur, E. *Multiple Regression in Behavioral Research: Explanation and Prediction.* Holt, Rinehart, and Winston, New York, 1982.
24. Steele, G. *COMMONLisp The language.* Digital Press, Bedford, Mass., 1983.
25. Walsh, J. P., Kiesler, S., Sproull, L. and Hesse, B. W. Self-selected and randomly-selected respondents in a computer network survey. *Pub. Opin. Q. 56* (1992), 241–244.
26. Webster, F. Data management for global ocean observing systems. Univ. of Delaware, Newark, Del., 1990.
27. Weill, P. Ph.D. thesis, Stern School of Business Administration, New York Univ., 1988.

**CR Categories and Subject Descriptors:** H.4.3 [**Information Systems Applications**]: Communications Applications—*bulletin boards, electronic mail*; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*evaluation/methodology, organizational design*; K.4.3 [**Computers and Society**]: Organizational Impacts

**General Terms:** Economics, Human Factors, Management

**Additional Key Words and Phrases:** Productivity, science and engineering networks, scientific computing, scientists as users

**About the Authors:**

**BRADFORD W. HESSE** is a research scientist at the American Institutes for Research in Palo Alto, Calif. He codirects AIR's Center for Research on Technology, where his current research includes investigations into the personal and organizational impact of electronically distributed work communities, evaluations of computerized information systems within complex organizations, assessments of usability needs (i.e., human factors), and the development of applications in artificial intelligence. **Author's Present Address:** American Institutes for Research, P.O. Box 1113, 1791 Arastradero Rd., Palo Alto, CA 94302; email: bhesse@forsythe.stanford.edu

**LEE S. SPROULL** is professor of management at Boston University and head of the Information Systems department. She has done extensive research in information behavior in organizations, social aspects of computing, and computer networks in organizations. **Author's Present Address:** Room 402, School of Management, Boston, Univ., 704 Commonwealth Ave., Boston MA 02215; email: lsproull@bu-pub.bu.edu

**SARA B. KIESLER** is professor of social sciences and social psychology at Carnegie Mellon University. She has conducted numerous experimental behavioral studies of computer-based communication and has participated in national policy committees on the social implications of technology. **Author's Present Address:** Department of Social and Decision Sciences Carnegie Mellon University Pittsburgh, PA 15213; email: sara.kiesler @andrew.cmu.edu

**JOHN P. WALSH** is an assistant professor of sociology at the University of Illinois at Chicago. His current research involves studying the uses of computer-mediated communication technology by scientists in several disciplines. **Author's Present Address:** Department of Sociology (M/C 312) University of Illinois at Chicago, Chicago, IL 60680; email: u33695@uicvm.bitnet