SOCIAL SCIENCE COMPUTER REVIEW

SOCIAL SCIENCE COMPUTER REVIEW

Volume 19, Number 4 / Winter 2001

SOCIAL SCIENCE COMPUTER REVIEW

Volume 19, Number 4 / Winter 2001

CONTENTS

Understanding Internet Usage

A Social-Cognitive Approach to Uses and Gratifications

ROBERT LAROSE
Michigan State University

DANA MASTRO
Boston College

MATTHEW S. EASTIN
Ohio State University

Several studies have applied uses and gratifications to explain Internet usage. Like Bandura's social-cognitive theory, the uses and gratifications framework explains media use in terms of expected positive outcomes, or gratifications. However, previous uses and gratifications research accounted for little variance in Internet behavior, although there were conflicting results. This research identifies new variables from social-cognitive theory that might further explain Internet usage and resolve inconsistencies in prior research. Measures of self-efficacy and self-disparagement were developed for the domain of Internet behavior. Internet addiction was interpreted as a deficient self-regulation within the social-cognitive framework. Finally, the negative outcomes of online behavior were analyzed for their impact on Internet usage. In a survey of 171 college students, the social-cognitive model explained 60% of the available variance in Internet usage using multiple regression analysis, a significant improvement over prior uses and gratifications research.

Keywords: Internet use, Internet behavior, social-cognitive theory, self-efficacy, self-disparagement, gratifications, multiple regression models

An understanding of Internet usage assumes considerable importance as society encounters problematic forms of online behavior. These include both instances of excessive use, in the form of so-called Internet addictions (Chou & Hsiao, 2000; Young, 1999), and underutilization by disadvantaged groups, known as the Digital Divide (Hindman, 2000; Hoffman & Novak, 1998; National Telecommunications and Information Administration, 2000). The new medium brings with it the opportunity to re-examine conventional models of media behavior.

THE USES AND GRATIFICATIONS PARADIGM

Uses and gratifications is perhaps the dominant paradigm for explaining media exposure in the field of communication studies. It has been applied to a wide range of conventional

AUTHORS' NOTE: The authors would like to acknowledge the assistance of the following students in the first author's research methods class at Michigan State University in the spring of 1999: Ho Il Kang, Yen-Chih Lue, Sung Bae Moon, Hau Wei Wang, and Chih Jen Yang.

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 395-413

mass media (Palmgreen, Wenner, & Rosengren, 1985) as well as to interpersonal communication (Rubin, Perse, & Barbato, 1988) and now to the Internet (Charney & Greenberg, 2001; Dimmick, Kline, & Stafford, 2000; Eighmey & McCord, 1998; Ferguson & Perse, 2000; Flanagin & Metzger, 2001; Kaye, 1998; Korgaonkar & Wolin, 1999; Lin, 1999; Papacharissi & Rubin, 2000; Parker & Plank, 2000; Perse & Greenberg-Dunn, 1998). Uses and gratifications researchers typically start with descriptions of common media uses, obtain ratings of the frequency or importance of those uses, and factor analyze the results to obtain gratification factors that are then correlated with media use. However, uses and gratifications typically only account for between 1% and 15% of the variance in use of conventional media (Palmgreen et al., 1985). With two notable exceptions (Charney & Greenberg, 2001; Lin, 1999), uses and gratifications have disappointed when applied to the Internet as well. The purpose of this article is to critically examine uses and gratifications theory as it applies to the Internet and to propose a more comprehensive theory of Internet usage based on social-cognitive theory (Bandura, 1986, 1997).

Uses and gratifications theory assumes that audiences actively seek out media in a goal-directed way that provides them with the means of gratifying a wide variety of needs (Katz, Blumler, Gurevitch, 1974; Palmgreen et al., 1985). In recent years, the theory has been reformulated to stress comparisons between the gratifications sought from a medium with gratifications obtained (GO). The underlying process is now conceived as an iterative one in which initial expectations about the outcomes of media exposure (the gratifications sought) are continually modified through observation of the gratifications actually obtained from the media, feeding back into the gratifications sought through future media exposure (Palmgreen et al., 1985). Gratifications sought do not in themselves predict media behavior very well; they have far greater explanatory power when compared with the gratifications obtained (Palmgreen, Wenner, & Rayburn, 1981).

USES AND GRATIFICATIONS OF THE INTERNET

With its emphasis on active media use and its ability to span both mass and interpersonal communication, uses and gratifications was initially regarded as a natural paradigm for understanding the Internet (Morris & Ogan, 1996). However, studies of the uses and gratifications of the Internet have tended to repeat the pattern of weak predictions of media behavior common to this body of research.

Several studies have applied conventional mass media gratifications to computer media. Kaye (1998) found low correlations between weekly Web usage and entertainment ($r = .30$), social interaction ($r = .23$), and escape ($r = .17$) gratifications. Perse and Greenberg-Dunn (1998) compared respondents who chose computers as the source of each of nine typical mass media gratifications with those who did not in an analysis of variance. Perceptions that computers were a habit or were good for keeping busy were significantly related to computer use, with an eta squared of .04 in each case. Ferguson and Perse (2000) obtained a significant regression result that explained 9% of the variation in Web usage, mostly from a measure of "expertise" (based on prior experience with computers and the Web). None of the four gratifications factors tested was a significant predictor, although significant but low zero-order correlations (ranging between .20 and .27) with usage were observed. For Parker and Plank (2000), a relaxation and escape factor predicted Internet usage with a standardized beta of .302, indicating that approximately 9% of the variance was explained (no overall *R*-squared was reported). Lin (1999) found three classic mass media motivations (surveillance; escape, companionship, and identity; and entertainment) predicted 47% of the variance in a multi-item measure of the likelihood to adopt online services.

Other researchers have explored potentially unique gratifications of the Internet. Korgaonkar and Wolin (1999) found that factors of escapism, information control, interactive control (relating to the ability to control the presentation of information), socialization, and economic motivation differentiated light (less than an hour per day) and heavy Web users. The discriminant analysis did not yield variance explained; however, the improvement over chance for these five factors (the respondents' levels of education and household incomes were also part of the prediction equation) was modest (45%) although statistically significant ($p < .05$). Charney and Greenberg (2001) established eight gratification factors for the Internet (keep informed, diversion and entertainment, peer identity, good feelings, communication, sights and sounds, career, and "coolness"), and two of these (keep informed and communication) explained 36% of the variance in weekly time spent on the net. Papacharissi and Rubin (2000) added a measure of interpersonal communication motives to conventional mass media gratifications and found the former to be the only significant predictor in a regression analysis that accounted for 7% of the overall variance in Internet exposure. Eighmey and McCord (1998) identified personal involvement, entertainment, and personal relevance gratification factors. Finally, Flanagin and Metzger (2001) explored a variety of needs gratifications not found in the conventional mass communication literature, including persuading others, problem solving, relationship maintenance, status seeking and achieving personal insight. However, neither of the latter two studies related the gratifications to Internet usage.

Thus, the challenge remains to provide a more comprehensive explanation of media behavior than is forthcoming from the majority of uses and gratifications research. But also, we must examine why two studies of Internet uses and gratifications (Charney & Greenberg, 2001; Lin, 1999) were more successful than others, indeed to a degree unprecedented in the annals of uses and gratifications research.

A SOCIAL-COGNITIVE THEORY OF INTERNET USAGE

The gratifications sought–gratifications obtained formulation is seemingly indistinguishable from an important mechanism in social-cognitive theory, enactive learning. Enactive learning describes how humans learn from experience. In the social-cognitive view, interactions with the environment (the media environment, in this case) influence media exposure by continually reforming expectations about the likely outcomes of future media consumption behavior (after Bandura, 1986). Seemingly, this represents the same process that describes the relationship among gratifications sought, media behavior, and gratifications obtained (Palmgreen et al., 1985). Recognizing this parallel, we next examine social-cognitive theory as a source of further insight into Internet usage.

Social-cognitive theory explains behavior in terms of reciprocal causation among individuals, their environments, and their behaviors (Bandura, 1986, 1989, 1999). The triadic causal mechanism is mediated by symbolizing capabilities that transform sensory experiences into cognitive models that guide actions. The human capacity for vicarious learning allows individuals to acquire rules for conduct without physically enacting the behavior but rather, by observing others. Direct experience with enacting behavior also affects these perceptions and that is called *enactive learning*. Individuals use their capacity for forethought to plan actions, set goals, and anticipate potential behavioral consequences. Through evaluations of personal experiences and self-assessments of their thought processes, they employ a self-reflective capability that helps them better understand themselves, their environments, and variations in situational demands.

Outcome expectations, defined as judgments of the likely consequences of a behavior (Bandura, 1997), provide incentives for enacting behavior while expectations of aversive outcomes provide disincentives (Bandura, 1986). Food, drink, and physical contact are primary incentives that motivate human behavior from infancy, but adults respond to symbolic incentives as well. These include monetary incentives, social incentives (such as obtaining approval from others), and status incentives. Sensory incentives involve exposure to pleasing or novel sensations. Preferences for enjoyable activities are the basis for activity incentives. There are also internal, self-reactive incentives resulting from comparisons of personal actions with standards for behavior (see the following sections). The objective magnitude of incentives does not matter as much as perceptions of how incentives are contingent on a particular course of action.

Humans also possess a self-regulatory capability that provides the basis for purposive action through the subfunctions of self-monitoring, judgmental process, and self-reaction (Bandura, 1986, 1991b). Self-monitoring is the observation of one's own actions to provide diagnostic information about the impact of behavior on the self, others, and the environment (Bandura, 1991b). The judgmental process compares self-observations of behavior to personal standards, personal or social norms, and the valuation of the activity, particularly when the locus of control for the behavior resides in the individual. The self-reactive function supplies the behavioral incentive through the satisfaction derived from accomplishing an activity that meets desired standards. Dysfunctional forms of self-regulation may also affect behavior. Addictions mark the failure of self-regulatory functions (Bandura, 1999), and deficient self-regulation has been conceptualized as the mechanism of so-called Internet addictions (LaRose, 2001). Self-slighting of one's accomplishments is another form of dysfunctional self-monitoring that reduces the self-reactive incentive to persist, and self-disparagement of one's capabilities can also inhibit performance (Bandura, 1986).

Another important determinant of behavior is self-efficacy, or belief in one's capability to organize and execute a particular course of action (Bandura, 1997). Those who perceive themselves to be highly efficacious with reference to a particular task will invest sufficient levels of effort to achieve successful outcomes, whereas those with low levels of self-efficacy will not persist.

Applying social-cognitive theory to Internet usage, expectations about the positive outcomes of Internet use, such as encountering informative Web pages or making valued social contacts, should increase usage. Each type of incentive (i.e., sensory, monetary, social, status, activity, and self-reactive) may make unique contributions. Expected negative outcomes, such as having one's computer freeze up while surfing the Web, should discourage use. Internet self-efficacy, or individuals' beliefs about their capability in using the Internet to accomplish useful tasks (Eastin & LaRose, 2000), should also determine exposure to a medium that many users find troublesome (Graphic, Visualization and Usability Center, 2000).

Self-regulatory mechanisms are also important in a medium that invites intense self-reflection (Turkle, 1995). Self-disparagement of one's abilities to use the Internet may negate the self-reactive incentive to persist in the face of failure or aversive outcomes. Others may engage in self-slighting of their Internet skills, depriving themselves of the satisfaction of successful performance. Self-disparagement and self-slighting may afflict even those with high levels of Internet self-efficacy if they compare their abilities to unrealistic standards set by the most accomplished Web users and by constantly changing Internet technology.

Whether excessive use of the Internet is truly an addiction in clinical terms is a controversial issue (Mitchell, 2000; Shaffer, Hall, & Vander Bilt, 2000; Walther & Reid, 2000) that we

will not attempt to resolve here. From a social-cognitive perspective, the so-called addictions are another form of deficient self-regulation; users are aware that the time they spend online is excessive and disruptive but suspend their comparisons to desirable standards of conduct. However, deficient self-regulation is not limited to extreme "addictive" cases and might affect Internet usage even at moderate levels (LaRose, 2001). In the absence of self-regulation, Internet use may continue to mount, unabated.

Compeau and Higgins (1995) provided empirical support for this sociocognitive conceptualization in the domain of general personal computer usage. They found that self-efficacy and professional outcome expectations (e.g., improvements in productivity) predicted the amount of computer use, explaining 34% of the variance in usage. Longitudinal research provided evidence of the direction of causality (Compeau, Higgins, & Huff, 1999). However, the subset of outcome expectations that the researchers termed *personal outcomes* had an unexpected negative relationship to usage, which they attributed to the unrealistic nature of the personal outcomes presented to their respondents (e.g., "I will get a raise or a promotion as a result of computer use"). Moreover, the outcome expectation items used in these studies did not reflect the full range of incentives that motivate behavior under social-cognitive theory.

A SOCIAL-COGNITIVE CRITIQUE OF INTERNET USES AND GRATIFICATIONS

In sociocognitive terms then, gratifications may be viewed as outcome expectations. Attempts by uses and gratifications researchers (Babrow & Swanson, 1988) to distinguish gratifications from formulations involving outcome expectations were of no avail and failed to produce more robust explanations of media exposure, suggesting that they may be related constructs. The outcome expectation construct parsimoniously bridges the gulf between gratifications sought and gratifications obtained in uses and gratifications research. Outcome expectations reflect current beliefs about the outcomes of prospective future behavior but are predicated on comparisons between incentives expected and incentives attained in the past.

Uses and gratifications may yield weak predictions of media exposure because they ignore important incentive categories that motivate behavior. Table 1 shows how Internet gratification dimensions map onto Bandura's (1986) incentive categories. This analysis was performed by applying the incentive definitions to individual gratification items and then characterizing gratification factors on the basis of their predominant incentive category. Perse and Greenberg-Dunn (1998) and Compeau et al. (1999) were included for the sake of completeness, although these studies did not examine Internet use per se but rather general personal computer use. Likewise, Eighmey and McCord (1998), Dimmick et al. (2000), and Flanagin and Metzger (2001) were included for completeness, but those studies did not investigate relationships between gratification factors and usage and so are not included in the analysis that follows.

The most commonly assessed incentive categories are activity (i.e., fun, entertaining, exciting, or boredom-relieving activities), social (e.g., social interaction or communication), novel sensory (i.e., information seeking), and self-reactive (i.e., to relax or escape). Each of these has been included as the major component of a gratification factor and found to be significantly related to usage in at least two of the studies analyzed. Pleasing (as opposed to novel) sensory incentives (e.g., interesting or enjoyable graphics or sounds) were included by Charney and Greenberg (2001) in a sights and sounds dimension, but that factor was not a predictor of Internet use. Status incentives (e.g., to win promotions or to seem "cool" or

TABLE 1
Incentive Components of Internet Gratification Factors

	Incentive Categories						
Study	*Sensory and/or Pleasing*	*Sensory and/or Novel*	*Social*	*Status*	*Monetary*	*Activity*	*Self-Reflective*
Kaye (1998)		XX	**X**x**x**			**X**X	**X**x**x**
Eighmey and McCord (1998)	X					X	X
Korgaonkar and Wolin (1999)		**x**X**xx**	**X**		Xx**X**	**xx**	**X**x**x**x
Perse and Greenberg-Dunn (1998)[a]		X	X			**X**	X
Compeau, Higgins, and Huff (1999)			x	x	x		**X**x
Lin (1999)		**xXx**	**x**			**Xx**	**xx**
Charney and Greenberg (2001)	X	**X**x	X**x**xxx	xxx	x	Xxx	xxxx
Parker and Plank (2000)		X	Xx			**X**	**X**
Papacharissi and Rubin (2000)		**x**X	**X**x	**x**	xX	**x**XX	**x**xx
Ferguson and Perse (2000)		xxX	Xxx			XX	XX
Dimmick, Kline, and Stafford (2000)		x	Xx		X	x	
Flanagin and Metzger (2001)		X	X	XX		X	XXXX

NOTE: X = incentive was a major component of a gratification factor; x = incentive was a minor component of a gratification factor. Bold indicates that the component was a significant predictor of usage. Multiple entries indicate that the incentive was represented in multiple gratification factors. Factors with no clear dominant component were listed as minor components. Thus, the entry in the Social row under the Kaye (1998) column indicates that there was one factor in which social incentives were a major component that was significantly related to usage. Social incentive items were also minor components of two other factors, one of which was significantly related to usage and one that was not.
a. Single-item measures were used in this study, so all components are listed as major components.

important) have never been a major component of Internet gratification factors, although they were minor components of those used by Compeau et al. (1999) as well as by Charney and Greenberg (2001) and Papacharissi and Rubin (2000). Monetary incentives (e.g., to find bargains online) also have received slight attention, although they were significant predictors of usage for Korgaonkar and Wolin (1999) and a major component of a gratifications factor that did not predict usage for Papacharissi and Rubin (2000). The tendency to overlook pleasing sensory, status, and monetary incentives may be due to a reliance on gratification items developed for television research (e.g., drawn from Rubin, 1983, and Greenberg, 1974) that may differ from those relevant to the Internet. The practice of asking about uses (as opposed to consequences) when eliciting new items may also create a bias in favor of certain types of incentives, notably activity incentives.

Negative as well as positive outcomes may shape behavior, a possibility generally neglected by uses and gratifications (although with some exceptions, Becker, 1979; Levy, 1977). The Internet regularly frustrates its users (Graphic, Visualization and Usability Cen-

ter, 2000), so this is a potentially significant oversight in Internet-related studies. Charney and Greenberg (2001) included a Net Frustrations measure but did not conceptualize this as a gratification in their terms or as an outcome expectancy in social-cognitive terms.

Although gratifications and outcome expectations are conceptually similar, the operational definitions found in uses and gratifications studies do not to distinguish between use and the consequences of use. This is a particularly important issue with respect to novel sensory incentives related to information seeking. Consider the gratification item, "I use the Web when I want to find specific information" (Kaye, 1998). Because much of the information on the Internet is unreliable at best and erroneous or fraudulent at worst, the consequence of information seeking could be negative or neutral. A better construction would be "I use the Internet to get information I can trust" (as in Charney & Greenberg, 2001). This issue may not be as relevant in conventional mass media studies because professional journalistic standards apply to assure the quality of information in the mass media, but such standards are generally lacking on the Internet.

The response task is a related issue. In uses and gratification studies, respondents are typically asked to indicate whether they use the Internet for a particular purpose (e.g., Eighmey & McCord, 1998; Ferguson & Perse, 2000; Kaye, 1998; Korgaonkar & Wolin, 1999; Papacharissi & Rubin, 2000; Parker & Plank, 2000) but not the likelihood of achieving the goal that is implicit in the specified use. We may agree that we use the Internet to obtain information, but this says nothing about the perceived likelihood of obtaining useful information in the future. Indeed, negative experiences with bogus information could lead us to conclude that it is unlikely we will obtain useful data in the future. In contrast, the response task in Charney and Greenberg (2001) assessed agreement about whether specific stated outcomes were likely to result from Internet use, and Lin (1999) presented an outcome likelihood scale as the response task. At this point, we can perhaps understand why Charney and Greenberg and Lin achieved superior and unprecedented results compared to the other studies. They measured outcome expectations very much as they should be measured from the sociocognitive view, linking the media exposure behavior in question to expected outcomes.

The gratification opportunities construct, or consumers' beliefs that a medium allows them to obtain greater opportunities for satisfaction (Dimmick et al., 2000), is conceptually similar to outcome expectations. Gratification opportunities, but not more conventionally phrased gratification measures, predicted the majority of cases in which e-mail displaced the telephone as a mode of interpersonal communication, confirming this analysis. However, the predictions were weak,[1] perhaps because the gratification opportunities were operationally defined in terms of general beliefs about the opportunities afforded by a medium (e.g., "It fits peoples' work schedules") rather than personal beliefs about expected outcomes (e.g., "If fits my work schedule") and so may have lacked the necessary precision.

Even when outcome expectancies are properly assessed, uses and gratifications studies may fail to fully account for Internet usage because they neglect variables that are important in social-cognitive theory, notably self-efficacy. Self-efficacy is a determinant of behavior and a mediator between outcome expectations and performance, as we have seen, so this is a noteworthy oversight. Ferguson and Perse's (2000) finding that expertise (a measure of prior experience with the computers and the Web) predicted Web usage suggests a role for self-efficacy in that prior experience is an important determinant of beliefs about one's capability to perform a behavior (Bandura, 1997). Eastin and LaRose (2000) found that Internet self-efficacy was a powerful predictor of Internet usage.

Self-regulation also has been neglected in uses and gratifications research, although the process of assessing and modifying gratification and/or outcome expectations in light of experience (as described in Palmgreen et al., 1985) implies certain aspects of self-

monitoring and judgmental processes. Habit is an element of uses and gratifications theory (Palmgreen et al., 1985) that interacts with beliefs and expectations about media alternatives to help formulate gratifications sought. In practice, however, Internet researchers treated habit as a type of gratification, for example, "Because it is part of my usual routine" (Korgaonkar & Wolin, 1999); "Because it is a habit, just something I do" (Ferguson & Perse, 2000; Kaye, 1998); and "Always do something because it is a 'habit'" (Perse & Greenberg-Dunn, 1998). The single-item measures of habit used by Kaye (1998) and Charney and Greenberg (2001) were dropped from those studies due to insufficient loadings on gratification factors. A parallel item in Korgaonkar and Wolin (1999) was only a minor component of a seemingly unrelated socialization factor and a minor component of Ferguson and Perse's (2000) entertainment factor. Perse and Greenberg-Dunn (1998) kept their single-item "habit gratification" item as a distinct element in their analysis, and for them, it was a significant predictor of computer use. So whereas habit is a predictor of behavior, it does not appear to be a gratification. Rather, we interpret habit as an indicator of deficient self-regulation within social-cognitive theory (Bandura, 1991b) and propose (after LaRose, 2001) that the symptoms of so-called Internet addictions are really indicators of habitual use stemming from ineffective self-regulation.

HYPOTHESES

Thus, we expect outcome expectations to predict Internet usage, with independent contributions from distinct types of incentives. This is the basic premise of the uses and gratifications paradigm as seen through the lens of social-cognitive theory. By assessing outcome expectations in terms of likely behavioral consequences instead of typical uses, we expect to replicate prior successes (i.e., Charney & Greenberg, 2001; Compeau et al., 1999; Lin, 1999) while achieving moderate multiple correlations with usage.[2]

Hypothesis 1: Expected activity, pleasing sensory, novel sensory, and social outcome expectations will be positively related to Internet usage.

Expected negative outcomes of Internet usage should also affect behavior according to social-cognitive theory. Levy (1977) and Becker (1979) raised this possibility with conventional mass media behavior, but negative outcomes have not been conceptualized in prior Internet gratifications studies.

Hypothesis 2: Expectations of negative Internet outcomes will be negatively related to Internet usage.

Internet self-efficacy should explain additional variation in Internet usage. This proposition has been established in a related behavioral domain (Compeau et al., 1999) and in prior Internet research (Eastin & LaRose, 2000) but is not found in uses and gratifications research. Internet self-efficacy should be related to the expected outcomes of Internet use. The relationship is reciprocal: Belief in one's ability to use the medium to attain important goals should precede the achievement of desired outcomes, and successful attainment of desired outcomes should also strengthen beliefs in one's ability. However, within social-cognitive theory, self-efficacy should make a unique contribution to behavior over and above outcome expectations (Bandura, 1997).

Hypothesis 3: Internet self-efficacy will be positively related to Internet usage, independent of the effects of outcome expectations.

Finally, self-regulatory mechanisms should also impact usage, another set of variables not found in previous research on Internet uses and gratifications. Self-disparagement and self- slighting with respect to Internet-related abilities should decrease usage. Self-perception of an Internet addiction should be positively associated with Internet use because it indicates deficient self-regulation.

Hypothesis 4: Self-disparagement will be negatively related to Internet usage.
Hypothesis 5: Self-slighting will be negatively related to Internet usage.
Hypothesis 6: Self-perceptions of Internet addiction will be positively related to usage.

METHOD

Participants

The participants were 171 undergraduate students from an introductory communication class at a large Mid-Western university. Of the respondents, 35% were freshman, 22% were sophomores, 18% were juniors, and 25% were seniors. Furthermore, 60% were male, 40% were female, and the mean age was 21 years old.

Procedure

Questionnaires were administered over a 2-week period. Respondents picked up the questionnaire on the first day of class each week and returned it the second day of class that same week. In the interim, respondents kept a diary of their total Internet use (i.e., amount and type of use). Respondents were offered extra credit for participating in the study; an alternate extra credit assignment was provided for those who chose not to participate.

Operational Measures

Respondents indicated the likelihood of each Internet outcome using Likert-type scale items ranging from *very likely* (scored 7) to *very unlikely* (scored 1). The four-item Activity Outcomes ($\alpha = .87$) Scale measured the likelihood of finding enjoyable activities on the Internet (e.g., "feel entertained"). The six-item Novel Sensory Outcomes ($\alpha = .83$) Scale assessed the likelihood of finding information on the Internet (e.g., "get immediate knowledge of big news events"). Items used to measure these outcome expectancies were obtained from among top loading items on gratification factors identified by Charney and Greenberg (2001). These were supplemented with items corresponding to incentive categories established within social-cognitive theory.

However, the prior research did not include a satisfactory measure of pleasing sensory outcomes in that all of three of the items in its sights and sounds factor were double-barreled questions (e.g., "to look at graphics or animation"). Five new items were constructed for the Pleasing Sensory Outcomes ($\alpha = .80$) Scale indicating the likelihood of encountering aesthetically pleasing visuals online (e.g., "see Web pages with bright colors"). The candidate items were developed from focus group interviews conducted with participants who attended the same class as the current respondents in the term prior to this research. Charney and Greenberg (2001) also did not include a social outcome factor. Items for the Social Outcomes ($\alpha = .86$) Scale were assembled from several different factors to compose a five-item scale assessing the likelihood of developing relationships over the Internet (e.g., "find com-

panionship"). The four-item Negative Outcomes ($\alpha = .61$) Scale assessed the likelihood of encountering negative outcomes associated with Internet use (e.g., "have trouble finding what I am looking for"). It drew on two items from Charney and Greenberg's Net Frustrations Scale and two suggested by a well-known survey of Web users (Graphic, Visualization and Usability Center, 2000).

Internet self-efficacy was assessed with eight Likert-type items ranging from *strongly agree* (scored 7) to *strongly disagree* (scored 1). Respondents rated their confidence that they could use the Internet in each of the ways specified, for example, to troubleshoot Internet problems ($\alpha = .93$; see Eastin & LaRose, 2000). Three self-regulatory constructs were also measured. Self-disparagement ($\alpha = .71$; e.g., "I feel my computer skills are inadequate") and self-slighting ($\alpha = .77$; e.g., "I feel helpless when I can't find what I am looking for on the Internet") consisted of three items, whereas perceived Internet addiction ($\alpha = .77$) had four items (e.g., "I use the Internet so much it interferes with other activities"). All three of these constructs were measured with Likert-type items ranging from *strongly agree* (scored 7) to *strongly disagree* (scored 1). The items for the latter measure were derived from the definition of the Internet addiction disorder (Internet Addiction Support Group, 2000).

Internet usage was an additive index of four self-reported items ($\alpha = .82$). Participants were asked on a typical weekend day and on a typical weekday about how much time they spent on the Internet (both items coded 1 if none, 2 if less than an hour, 3 if 1 to 2 hours, 4 for more than 2 and up to 5 hours, and 5 if more than 5 hours), about how many days in a typical week they went on the Internet (responses ranged from 0 to 7), and how much time they spent surfing the Web each week (Coded 1 for none, 2 for less than an hour, 3 for 2 to 4 hours, 4 for 5 to 7, 5 for 7 to 9, 6 for 10 to 20, and 7 for more than 20 hours).

Analyses

All analyses were performed using the Statistical Package for the Social Sciences, Version 7.5 (SPSS, 1997). Pearson product-moment correlations (with pairwise exclusion of missing cases) were used to test all hypotheses. To assess the relative predictive utility of the various independent variables, they were entered into a multiple regression analysis, with Internet usage as the dependent variable.

When proposing new variables for an existing model, it is customary to perform stepwise regression in which previous variables (outcome expectations, in this case) are introduced first. However, the reciprocal causation mechanism posited by social-cognitive theory made this inappropriate, so all independent variables were entered in a single step.

RESULTS

A correlation matrix showing the relationships between variables is presented in Table 2. As stated in Hypothesis 1, activity outcomes ($r = .48$, $p < .001$), pleasing sensory outcomes ($r = .37$, $p < .001$), novel sensory outcomes ($r = .32$, $p < .001$), and social outcomes ($r = .37$, $p < .001$) were all positively related to Internet usage. Negative Internet outcomes were inversely related to Internet usage ($r = –.16$, $p < .05$), thus supporting Hypothesis 2.

Hypotheses 3, 4, 5, and 6 were also confirmed. Self-efficacy ($r = .65$, $p < .001$), self-disparagement ($r = –.48$, $p – .001$), self-slighting ($r = –.46$, $p < .001$), and self-perceptions of Internet addiction ($r = .65$, $p < .001$) were all found to be significantly related to Internet usage in the directions hypothesized.

TABLE 2
Pearson Product Moment Correlation Coefficients

	Variable										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	M	SD
1. Internet use										14.21	4.35
2. Activity outcomes	.48**									20.31	5.43
3. Pleasing sensory outcomes	.37**	.53**								27.74	5.45
4. Novel sensory outcomes	.32**	.46**	.68**							16.00	4.21
5. Social outcomes	.39**	.23**	.03	.10						10.92	6.42
6. Negative outcomes	–.16*	–.08	.00	–.11	–.09					19.20	5.48
7. Internet self-efficacy	.65**	.37**	.39**	.40**	.37**	–.24**				54.50	16.80
8. Self-disparagement	–.48**	–.31**	–.32**	–.30**	–.06	.31**	–.61**			9.29	4.24
9. Self-slighting	–.46**	–.26**	–.19*	–.22**	–.07	.45**	–.58**	.62**		10.12	4.80
10. Perceived addiction	.65**	.37**	.20**	.27**	.51**	–.11	.57**	–.29**	–.30**	15.00	5.80

*$p < .05$. **$p < .001$.

To gain an overall understanding of how each of these variables predicted Internet usage while controlling for the other variables, a multiple regression was conducted. The combination of these variables significantly predicted Internet usage ($F_{9,168} = 26.512$, $R^2 = .60$, $p < .001$; see Table 3).

Internet Self-Efficacy ($b = .652$), perceived addiction ($b = .411$), activity outcomes ($b = .208$), and self-disparagement ($b = -.144$) each uniquely predicted Internet use at an alpha level of $p < .05$. In combination, self-slighting, pleasing sensory outcomes, social outcomes, novel sensory outcomes, and negative outcomes only explained an additional .8% of the variance in Internet usage.

DISCUSSION

The conceptualization of Internet usage as a social-cognitive process received considerable confirming evidence. Positive outcome expectations, Internet self-efficacy, and perceived Internet addiction were directly related to Internet usage as expected. Negative outcome expectations, self-disparagement, and self-slighting were negatively related to usage. As was the case for Charney and Greenberg (2001) and Lin (1999), defining gratifications as outcome expectations once again explained greater variation in Internet usage than in uses and gratifications studies employing conventional operational definitions stressing frequent uses. Internet self-efficacy and perceived addiction explained considerable additional variance in usage beyond outcome expectations, indicating the value of adding these concepts to models of Internet usage.

Whereas negative outcome expectations were negatively related to usage, the relationship was a weak one that disappeared in multiple regression analysis. The measure of negative outcomes had marginal reliability, which could account for the disappointing result. However, scatterplot analysis indicated the possibility of a curvilinear relationship between negative consequences and usage (although modeling this element as a nonlinear component did not affect the results). That is, high expectations of negative consequences, such as having one's computer freeze up or having trouble finding information on the Internet, were associated with both very high levels of usage and very low levels of usage. It may be that encountering frustrations such as these while using the Internet may cause some users to give up, whereas others persist. Persistent users may spend more time on the Internet as they work their way back to the location where the freeze-up occurred or apply new search strategies, actions that would increase the amount of time spent on the Internet. This type of persistence is an inherent characteristic of individuals with high self-efficacy. There was a moderate correlation between negative outcome expectations and self-slighting and also a significant correlation with self-disparagement. Both variables had moderate negative correlations with the dependent variable, suggesting that these self-regulatory processes mediated the relationship between negative outcomes and usage.

Self-disparagement and self-slighting had strong negative correlations with Internet usage in themselves, although the latter relationship was attenuated in multiple regression analysis (possibly as a result of multicollinearity between the two). Both had moderately high negative correlations with Internet self-efficacy, suggesting it is an intervening variable between self-regulation and usage. Sociocognitive processes are dynamic and iterative ones in which conceptions of one's ability to perform a behavior are continually re-evaluated through experience. In the process of mastering the Internet, users may proceed from a state of helplessness and despair to one of confidence and mastery. Deficient self-regulatory beliefs may be critical in the early stages of Internet experience but lose their power after a sense of self-efficacy is attained.

TABLE 3
Stepwise Regression of Sociocognitive Variables on Internet Usage

Variable	*Beta*	R^2 *Change*	F *Change*	*Significant* F *Change*
Internet self-efficacy	.652	.425	123.504	.000
Perceived addiction	.411	.114	41.055	.000
Activity outcomes	.208	.035	13.733	.000
Self-disparagement	–.144	.013	5.048	.026
Self-slighting	–.080	.004	1.401	.238
Pleasing sensory outcomes	.056	.002	.846	.359
Social outcomes	.038	.001	.398	.529
Novel sensory outcomes	–.031	.001	.282	.596
Negative outcomes	.005	.000	.010	.922

NOTE: F (9, 168) = 26.512. $p < .001$. $R^2 = .600$.

Although outcome expectations were significant predictors of Internet usage, as they were in Charney and Greenberg (2001), the relative importance of the categories differed. In the previous research, an informational dimension (termed *novel sensory expectations* here) was the most powerful predictor of Internet usage, whereas here, the activity dimension (identified as a diversion-entertainment factor in the prior research) dominated. One possible explanation is that in the 3 years between the two studies, the character of the Internet changed so that it became increasingly regarded as a source of enjoyable activities rather than a repository of information. In that time, a variety of new enjoyable activities became widely accessible over the Internet, at least among the college populations surveyed in these two studies. These include playing multi-user games and downloading music from the Internet, both time-consuming, enjoyable activities that could greatly increase total Internet usage.

Pleasing sensory outcome expectations were distinguished from novel sensory ones to preserve the identity of informational gratifications that are a common element in uses and gratifications research. However, the two were highly correlated ($r = .68$, $p < .001$), suggesting that they may reflect the same underlying construct. In social-cognitive theory, both novel and pleasing stimuli fall under the category of sensory incentives, with novelty viewed as a means of enhancing the impact of pleasing stimuli (Bandura, 1986).

Perceived addiction was a powerful predictor of usage. There was a moderate correlation between perceived addiction and social outcome expectations ($r = .51$, $p < .001$), suggesting that excessive use of e-mail, newsgroups, or chatrooms may be implicated. *Addiction* may be a misnomer in that high scores on this measure were often associated with moderate levels of Internet use, whereas the term is usually reserved for deviant levels of use. This variable was conceptualized in terms of an acknowledged failure to regulate one's own Internet behavior, and that mechanism could operate across a wide range of usage as opposed to something that occurs only after a very high threshold is crossed. Thus, a more appropriate term might be *deficient self-regulation*, as proposed in LaRose (2001).

Limitations

This research used retrospective self-reports of media behavior, an approach that is frequently criticized as being inherently unreliable. However, a comparison of diary data and retrospective self reports of Internet usage made in this study revealed a .65 correlation between the two measures. This finding is consistent with studies comparing self-reports of

computer use with electronic log data (Deane, Podd, & Henderson, 1998) and with comparisons of self-reported and objectively measured computer game activity (Zielke, Schildmann, & Wirausky, 1995). It may be that the interactive nature of Internet consumption makes it more salient than passive consumption of conventional mass media, so that retrospective reports of behavior are more accurate. Within social-cognitive theory, retrospective self reports take on a special significance and reflect a self-monitoring process. "If [people] want to exert influence over their own actions, they have to know what they are doing" (Bandura, 1986, p. 336).

This study relied on a convenience sample of college students. College students are a logical choice for exploratory research on Internet usage because they enjoy ready access to Internet resources. However, these results should be replicated with more diverse populations. Other populations may be differentially affected by the various types of incentive dimensions that frame outcome expectations. Self-efficacy might play less of a role in populations with smaller proportions of relatively new Internet users.

A cross-sectional survey method was used to assess the explanatory power of new predictors of Internet usage behavior. Social-cognitive theory stresses dynamic relationships between these variables and reciprocal causation (e.g., among self-efficacy, outcome expectations, and the performance of behavior), mechanisms that are better understood through structural modeling or time-series approaches.

Theoretical Implications

Can uses and gratifications and social-cognitive theory coexist? If the uses and gratifications mechanism is identical with the enactive learning mechanism under social-cognitive theory, it is possible to conclude that the latter subsumes the former or makes it redundant. However, uses and gratifications has a 30-year tradition that offers many insights into communication behavior. As Blumler (1979) pointed out, the definite articles *the* or *a* are not appropriate prefixes to uses and gratifications theory: There are many theories of the phenomenon. So perhaps there is room to modify and expand uses and gratifications to incorporate sociocognitive constructs instead of pitting one theory against the other.

The incentive categories (i.e., social, status, activity, sensory, monetary, and self-reactive) found in social-cognitive theory could provide a consistent theoretical framework in which to explore the outcome expectations associated with the Internet. The Internet studies reviewed here drew on the same sources for their initial gratification items: studies of television use completed some two decades ago (Greenberg, 1974; Rubin, 1981, 1983). These were supplemented with items suggested by the authors' own theoretical analyses of the Internet and/or from suggestions elicited from focus groups. Following the accepted practice of uses and gratifications research, the items were subjected to exploratory factor analysis, yielding widely varying factor structures with as few as three (in the case of Lin, 1999) or as many as eight (Charney & Greenberg, 2001) separate dimensions. Our earlier analysis of the incentive dimensions found in Internet gratification factors (presented in Table 1) revealed that most gratification factors reflected multiple incentives, with some drawing from as many as five different categories. Future uses and gratifications research might build on the incentive categories from social-cognitive theory, probing for both positive and negative instances in exploratory qualitative research and then applying confirmatory rather than exploratory factor analysis (Hunter & Gerbing, 1982).

The reliance on past television research as the foundation for Internet usage studies may obscure variables that are salient in the online environment but not present in the television-viewing environment of a generation ago. Status cues have been overlooked, perhaps because there is little enhancement of status associated with television viewing (the typical

focus of prior research), at least not in the span of years (since the early 1970s) that uses and gratifications has been with us. Status cues may be highly salient for Internet use, however. The widely publicized Digital Divide effect heightens status cues by stressing the association between social status and Internet use. Monetary incentives are another category that are not very salient when talking about "free" (i.e., advertising-supported) media but are important to the many Internet users who still pay by the minute (e.g., when dialing long distance to reach an Internet service provider) and when examining pay sites and electronic commerce (as in Korgaonkar & Wolin, 1999). Finally, self-efficacy is unlikely to be an important determinant of television exposure, a behavior that does not involve complex skills to perform, but is an important variable when analyzing a complex and continually changing interactive medium such as the Internet.

The self-regulatory mechanisms described by social-cognitive theory may clarify the role of habitual media behavior. As we have seen, uses and gratifications researchers conceptualized habit as a distinctive element in the model (e.g., Palmgreen et al., 1985) but in Internet research have operationalized the construct as a gratification, with single items that are obscured by overall factor structures. However, habit or perceived addiction, as it was called in this study, explains unique variance in usage distinct from outcome expectations. Thus, including habit in factor analyses of gratifications and/or outcome expectations obscures the role of the variable.

Further conceptual elaboration of self-regulatory mechanisms in media behavior should explore media addictions in terms of faulty self-monitoring, failure to apply standards to media behavior, and inability to generate self-reactive incentives. Conventional mass media usage, particularly television viewing, also has a habitual quality that some have termed *addictive* (e.g., Finn, 1992; McIlwraith, 1998; Winn, 1985). Finn found evidence that models of addiction predicated on self-control and personal responsibility provided the most plausible explanations of television addiction. McIlwraith found correlations between television addiction and both poor attention control and daydreaming about guilt and failure, both possible indications of faulty self-regulation.

Research Implications

Thus, reformulating gratifications as outcome expectations (rather than as gratifications sought or gratifications obtained) in operational terms and organizing them into incentive categories may improve the predictive validity of uses and gratifications research in general. This would bring greater consistency to the measurement of gratifications (or incentive categories, in social-cognitive terms) as well. Status and monetary incentives should receive greater emphasis in future Internet-related studies.

Moreover, the identity and relative importance of the various categories of incentive motivators should be periodically reassessed when examining the Internet, perhaps beginning with a fresh approach to item elicitation and scale construction that does not rely on 20-year-old television research. There were indications of a shift in the primacy of incentive categories over a relatively short period time when comparing this study to Charney and Greenberg's (2001), completed only 3 years earlier.

A new approach to assessing negative outcomes may be in order. Each incentive dimension presents the possibility of both positive and negative outcomes. For example, the receipt of unwanted e-mail would presumably be a negative instance of a social incentive that mitigates against Internet use when weighed against the positive social outcomes, such as finding companionship. Inclusion of both negative and positive instances might further improve the variance explained.

However, negative instances do not relate consistently to positive outcomes, so creating a single additive scale for each incentive would reduce reliability and violate accepted procedures of scale construction. Another possibility is to weight outcome expectations by the corresponding evaluation of each outcome and then add the products across all outcomes, a procedure recommended by the theory of planned behavior (Ajzen, 1985) and emulated in certain uses and gratifications studies (e.g., Babrow & Swanson, 1988; Rayburn & Palmgreen, 1984). Outcome evaluations were obtained in this study but failed to produce any overall improvement in either the reliability of the outcome scales or the variance explained in Internet usage. Future efforts might develop measures of negative outcomes for each dimension of expected outcomes, beginning with qualitative elicitations of negative instances along each incentive dimension.

Future research should include self-efficacy and self-regulatory measures as part of a comprehensive model of media attendance. Social-cognitive theory recommends two further mechanisms that are known to communication researchers (Bandura, 1991a) but have not been conceptualized in uses and gratifications models of media exposure: vicarious reinforcement and social diffusion of innovations. In the context of Internet use, these mechanisms call attention to such variables as the competence levels, social characteristics, and role relationships of the peers, mentors, and experts on whom individuals rely for formal and informal Internet training.

Social Implications

The Digital Divide is described as a problem of differential access to Internet resources separating rich from poor and minorities from whites (Hoffman & Novak, 1998; National Telecommunications and Information Administration, 1999). However, uncertainty about how to get started on the Internet and how to deal with the complexities of computers may be nearly as important deterrents as economics and lack of access (Katz & Aspden, 1996). Efforts to bolster Internet self-efficacy using techniques proven effective in other behavioral domains may be needed to completely close the Digital Divide. These include enactive mastery, vicarious experience, verbal persuasion, and control of disruptive physiological responses that frustrations with the Internet may provoke. Moreover, the divide will not necessarily close even if Internet access is assured if minorities do not persist in their usage. The inability to find anything the user wants rivals cost as a factor preventing Internet usage in minority homes once they have computers (National Telecommunications and Information Administration, 1999). In this context, this is a problem of low outcome expectations. Users do not have sufficiently strong beliefs about positive outcomes—or perhaps believe too strongly in negative outcomes—to motivate continuing use (Eastin & LaRose, 2000).

The current policy of subsidizing the cost of access (e.g., through the e-rate program) reduces a monetary disincentive to Internet use, but there are other economic disincentives that are not being addressed, including the cost of support and the cost of access to information. Other monetary incentives take the form of e-commerce "bargains," comparison shopping for big-ticket items and profits from stock transactions (cf. Korgaonkar & Wolin, 1999); these are not available to those who lack consumer credit resources.

Providing content that is specifically tailored to low-income and minority communities addresses sensory and activity incentives, an approach exploited by Web sites that cater to specific minority groups (e.g., `http://www.bet.com` and `http://www.blackvoices.com`). Some stress interactive communication, providing a potential source of social incentives (`http://www.blackplanet.com`). The participation of prominent community leaders in Web-related projects addresses the status dimension. Thus,

many of the components for a social-cognitive "solution" to the Digital Divide are becoming available, and perhaps promoting greater awareness of these options is all that is required.

However, efforts to publicize the Digital Divide as a social problem may have an unfortunate side effect by reinforcing self-disparagement and self-slighting among minority Internet users. The Digital Divide could become a dysfunctional social standard against which minorities compare their own behavior, concluding (erroneously) that performance failures are to be expected relative to their own community standards. To counteract this, interventions could provide realistic but encouraging feedback about the attainments of minority users as they explore the Internet and provide relevant points of social comparison from within minority communities (see Eastin & LaRose, 2000, for a further discussion of intervention strategies).

Social-cognitive theorists propose self-regulation as the key to understanding physiological addictions (Bandura, 1999), and behavioral addictions might be conceptualized with similar mechanisms, absent the physiological craving. Internet addictions represent a suspension of normal self-regulatory processes. The self-described "addicts" are aware that their usage is excessive but fail to apply standards that could supply the self-incentive to modify the behavior. Self-regulation might be restored by encouraging problem users to track their own behavior, setting realistic near-term goals for behavior change, developing multiple coping strategies, and identifying incentives that motivate abstinence (Perri, 1985). Social support is also important, including developing social support for abstinence and severing ties with other addicts (Bandura, 1999).

Coping self-efficacy, or one's belief that she or he can successfully engage in actions that will overcome an addiction, affects the success of these self-regulatory efforts. As outlined above, coping self-efficacy may be bolstered through enactive mastery, verbal persuasion, vicarious experience, and control of physiological states. Thus, social-cognitive theory emphasizes self-help coupled with appropriate social support. The treatment paradigm for online addictions might stress the development of self-awareness of excessive usage and online communities that support moderation.

NOTES

1. However, the prediction may not have been very robust. The gratification opportunities analysis predicted 58.4% of the cases in which the respondent (self reported) used e-mail more than the telephone. However, with only two possible categories (used the phone the same or less; only a handful reported using it more), the chance level was 50%. No conventional estimate of the variance explained was provided.

2. This research builds on the gratification factors of Charney and Greenberg (2001) because that study produced the most successful predictions of actual Internet usage to date. However, their factors did not include any with major components for monetary, status, or self-reactive incentives, so these were not reflected in this study.

REFERENCES

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds), *Action-control: From cognition to behavior* (pp. 11-39). New York: Springer.

Babrow, A. S., & Swanson, D. L. (1988). Disentangling antecedents of audience exposure levels: Extending expectancy-value analyses of gratifications sought from television news. *Communication Monographs*, *55*, 1-21.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*, 1175-1184.

Bandura, A. (1991a). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 61-90). Hillsdale, NJ: Lawrence Erlbaum.

Bandura, A. (1991b). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*, 248-287.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (1999). A sociocognitive analysis of substance abuse: An agentic perspective. *Psychological Science*, *10*, 214-217.

Becker, L. B. (1979). Measurement of gratifications. *Communication Research*, *6*, 54-73.

Blumler, J. G. (1979). The role of theory in uses and gratifications studies. *Communication Research*, *6*, 9-36.

Charney, T., & Greenberg, B. (2001). Uses and gratifications of the Internet. In C. Lin & D. Atkin (Eds.), *Communication, technology and society: New media adoption and uses* (pp. 383-406). Cresskill, NJ: Hampton.

Chou, C., & Hsiao, M. C. (2000). Internet addiction, usage, gratification, and pleasure experience: The Taiwan college students' case. *Computers and Education*, *35*, 65-80.

Compeau, D., & Higgins, C. (1995). Computer self-efficacy: Development of a measure and initial test. *Management Information Systems Quarterly*, *19*, 189-211.

Compeau, D., Higgins, C., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *Management Information Sciences Quarterly*, *23*, 145-158.

Deane, F. P., Podd, J., & Henderson, R. D. (1998). Relationship between self-report and log data estimates of information system usage. *Computers in Human Behavior*, *14*, 621-636.

Dimmick, J., Kline, S., & Stafford, L. (2000). The gratification niches of personal e-mail and the telephone. *Communication Research*, *27*, 227-248.

Eastin, M. A., & LaRose, R. L. (2000). Internet self-efficacy and the psychology of the digital divide. *Journal of Computer Mediated Communication* [Online serial], *6*(1). Retrieved November 29, 2000, from the World Wide Web: `http://www.ascusc.org/jcmc/vol6/issue1/eastin.html`

Eighmey, J., & McCord, L. (1998). Adding value in the information age: Uses and Gratifications of sites on the World Wide Web. *Journal of Business Research*, *41*, 187-194.

Ferguson, D. A., & Perse, E. M. (2000). The World Wide Web as a functional alternative to television. *Journal of Broadcasting and Electronic Media*, *44*, 155-174.

Finn, S. (1992). Television "addiction?" An evaluation of four competing media-use models. *Journalism Quarterly*, *69*, 422-535.

Flanagin, A. J., & Metzger, M. J. (2001). Internet use in the contemporary media environment. *Human Communication Research*, *27*, 153-181.

Graphic, Visualization and Usability Center. (2000). *GVU's tenth annual WWW user's survey.* Atlanta: Georgia Institute of Technology. Retrieved July 20, 2000, from the World Wide Web: `http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/graphs/use/`

Greenberg, B. S. (1974). Gratifications of television viewing and their correlates for British children. In J. Blumler & E. Katz (Eds.), *The uses of mass communication: Current perspectives on gratifications research* (pp. 71-92). Beverly Hills, CA: Sage.

Hindman, D. B. (2000). The rural-urban digital divide. *Journalism and Mass Communication Quarterly*, *77*, 549-560.

Hoffman, D. L., & Novak, T. P. (1998, April 17). Bridging the racial divide on the Internet. *Science*, pp. 390-391.

Hunter, J., & Gerbing, D. (1982). Unidimensional measurement, second order factor analysis, the causal model. *Research in Organizational Behavior*, *4*, 267-320.

Internet Addiction Support Group. (2000). *Are you suffering from Internet addiction disorder?* Retrieved July 20, 2000, from the World Wide Web: `http://www.iucf.indiana.edu/~brown/hyplan/addict.html`

Katz, E., Blumler, J., & Gurevitch, M. (1974). Utilization of mass communication by the individual. In J. Blumler & E. Katz (Eds.), *The uses of mass communications: Current perspectives on gratifications research* (pp. 19-32). Beverly Hills, CA: Sage.

Katz, J. E., & Aspden, P. (1996, October 6). *Motivations for and barriers to Internet usage: Results of a national public opinion survey.* Paper presented to the 24th annual Telecommunications Policy Research Conference, Solomons, Maryland.

Kaye, B. (1998). Uses and gratifications of the World Wide Web: From couch potato to Web potato. *New Jersey Journal of Communication*, *6*, 21-40.

Korgaonkar, P., & Wolin, L. (1999). A multivariate analysis of Web usage. *Journal of Advertising Research*, *39*, 53-68.

LaRose, R. (2001). On the negative effects of e-commerce: A sociocognitive exploration of unregulated on-line buying. *Journal of Computer Mediated Communication* [Online serial], *6*(3). Retrieved May 12, 2001, from the World Wide Web: `http://www.ascusc.org/jcmc/vol6/issue3/larose.html`

Levy, M. R. (1977). The uses and gratifications of television news. *Journal of Communication*, *27*, 112-117.

Lin, C. (1999). Online-service adoption likelihood. *Journal of Advertising Research*, *39*, 79-89.
McIlwraith, R. D. (1998). "I'm addicted to television": Personality, imagination, and TV watching patterns of self-identified TV addicts. *Journal of Broadcasting and Electronic Media*, *42*, 371-386.
Mitchell, P. (2000). Internet addiction: Genuine diagnosis or not? *The Lancet*, *355*, 632.
Morris, M., & Ogan, C. (1996). The Internet as a mass medium. *Journal of Communication*, *46*, 39-50.
National Telecommunications and Information Administration. (1999). *Falling through the net: Defining the Digital Divide*. Retrieved May 12, 2001, from the World Wide Web: `http://www.ntia.doc.gov/ntiahome/fttn99/contents.html`
National Telecommunications and Information Administration. (2000). *Falling through the net: Toward digital inclusion*. Retrieved May 12, 2001, from the World Wide Web: `http://www.ntia.doc.gov/ntiahome/fttn00/contents00.html`
Palmgreen, P., Wenner, L. A., & Rayburn, J. D. (1981). Gratification discrepancies and news program choice. *Communication Research*, *8*, 451-478.
Palmgreen, P., Wenner, L., & Rosengren, K. (1985). Uses and gratifications research: The past ten years. In K. Rosengren, L. Wenner, & P. Palmgreen (Eds.), *Media gratifications research* (pp. 11-37). Beverly Hills, CA: Sage.
Papacharissi, Z., & Rubin, A. M. (2000). Predictors of Internet usage. *Journal of Broadcasting and Electronic Media*, *44*, 175-196.
Parker, B. J., & Plank, R. E. (2000). A uses and gratifications perspective on the Internet as a new information source. *American Business Review*, *18*, 43-49.
Perri, M. G. (1985). Self-change strategies for the control of smoking, obesity and problem drinking. In T. A. Wills & S. Shiffman (Eds.), *Coping and substance use* (pp. 295-317). San Diego, CA: Academic Press.
Perse, E., & Greenberg-Dunn, D. (1998). The utility of home computers and media use: Implications of multimedia and connectivity. *Journal of Broadcasting and Electronic Media*, *42*, 435-456.
Rayburn , J. II, & Palmgreen, P. (1984). Merging uses and gratifications and expectancy-value theory. *Communication Research*, *11*, 537-562.
Rubin, A. M. (1981). An examination of television viewing motives. *Communication Research*, *8*, 37-51.
Rubin, A. M. (1983). Television uses and gratifications: The interactions of viewing patterns and motivations. *Journal of Broadcasting*, *27*, 37-51.
Rubin, R. A., Perse, E., & Barbato, C. (1988). Conceptualization and measurement of interpersonal communication motives. *Human Communication Research*, *14*, 602-628.
Shaffer, H. J., Hall, M. N., & Vander Bilt, J. (2000). "Computer addiction": A critical consideration. *American Journal of Orthopsychiatry*, *70*, 162-168.
SPSS, Inc. (1997). *Statistical package for the social sciences* (Version 7.5 for Windows). Chicago: Author.
Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*. New York: Simon & Schuster.
Walther, J. B., & Reid, L. D. (2000, February 4). Understanding the allure of the Internet. *Chronicle of Higher Education*, pp. B4-B5.
Winn, M. (1985). *The plug-in drug: Television, children and the family*. New York: Viking.
Young, K. S. (1999). Evaluation and treatment of Internet addiction. In L. VandeCreek & T. Jackson (Eds.), *Innovations in clinical practice: A source book* (Vol. 17, pp. 19-31). Sarasota, FL: Professional Resource.
Zielke, A., Schildmann, S., & Wirausky, H. (1995). *Spiel-und Sozialverhalten im Morgengrauen* [Play and social behavior in the MUD Morgengrauen] [Online]. Retrieved April 5, 2000, from the World Wide Web: `http://www.mud.de/Forschung/verhalten.html`

Robert LaRose is a professor in the telecommunication department at Michigan State University. He has a Ph.D. in communication theory and research from the Annenberg School at the University of Southern California.

Dana Mastro is an assistant professor in the Department of Communication at Boston College. She holds a Ph.D. in communication from Michigan State University.

Matthew S. Eastin is an assistant professor in the School of Journalism and Communication at Ohio State University. He has a Ph.D. in mass media from Michigan State University.

Affective Dimensions of Internet Culture

ADAM B. KING
Indiana University

This study examines online culture by constructing an affective portrait of Internet users. Respondents were recruited through a highly visible advertisement on the Yahoo! search engine, and their sentiments were collected with a Web-based survey instrument using Osgood's semantic-differential technique. The study had three particularly engaging findings: First, Internet users hold extremely intense and nuanced affective sentiments toward the components of their online world. Second, these sentiments vary with the amount of time Internet users spend online as well as their cumulative years of Internet experience. Third, men and women have slightly different affective responses toward many components of online culture, suggesting that despite the supposedly gender-free nature of the Internet, men and women experience somewhat different online social worlds.

Keywords: Internet, online culture, affect, cultural sentiments, cybersociety, social

Early reports of online communication generally claimed that the expressive limitations of the medium prevented users from sharing emotional displays and affective sentiments. The Internet—a text-only medium at that time—was viewed as a classic "mass society," populated by a swarm of individual users who were effectively insulated from one another's personalities. Any emotionality that may have occurred in online interaction was seen as idiosyncratic, dependent on transient moods of individual users rather than on the influence of shared situations or of a shared affective culture.

Recent examinations of online social life offer a different picture. An "online culture" has developed as online social relationships have deepened and matured (Jones, 1995; Walther, 1992). This culture and the structure of shared affective sentiments that it implies provide several new points of departure for research regarding online affect.

First, research can examine the online affective culture and thereby focus on an aspect of the online social world that is more immediate than the long-term affective consequences of online activity and more stable than the momentary emotions accompanying online interactions. Social interaction is motivated, mediated, and successfully accomplished through the influence of culturally grounded sentiments held toward the components of a shared social world (MacKinnon, 1994). Online interaction is no different from offline interaction in this regard, so examining Internet cultural sentiments provides insight into the affective bases of online social interaction.

Second, sentiments held by Internet users can be compared with the sentiments held by members of other cultures. This comparison is possible because the affective meanings of concepts can be measured rigorously with graphic rating scales assessing three connotative

AUTHOR'S NOTE: Adam King would like to thank David R. Heise for his indispensable advice and support during this project. He also thanks the members of the Indiana University Sociology Department Social Psychology Seminar for their constructive criticism and helpful suggestions.

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 414-430

dimensions: evaluation, potency, and activity (EPA) that are cross-cultural universals (Osgood, May, & Miron, 1975). Research directed at the EPA dimensions during the past 30 years (MacKinnon, 1994; Osgood & Tzeng 1990) has shown that when aggregated into group measures, individuals' EPA ratings define cultural sentiments.

The study reported here used an online survey instrument to measure cultural sentiments held by Internet users. In particular, graphic rating scales, presented to users via the World Wide Web (Bertot & McClure 1997), gathered data on the EPA profiles of typical online identities, settings, and behaviors. These data were then aggregated to reveal the affective sentiments that Internet users hold toward components of the online social environment. The measurements are examined in this article to determine if online culture is reflected coherently in users' sentiments, to assess ways in which users' sentiments are the same as or different than sentiments in the general culture, and to explore how users' sentiments are related to personal characteristics such as sex and amount of experience on the Internet.

SELECTION OF STIMULI

Internet concepts denoting identities, behaviors, and settings were selected from a number of offline and online Internet-related dictionaries (Barry, 1991; Crumlish, 1995; Fahey, 1994; Raymond, 1993). Two criteria were used in selecting stimuli.

First, the word or phrase had to be familiar enough to elicit meaningful responses from a cross section of Internet users. This criterion was articulated primarily through the frequency with which the term appeared in the various online and offline dictionaries consulted. Second, the term ideally referred exclusively to online activities rather than offline activities. Exceptions had to be allowed (e.g., chat, ignore, post) where it was impossible to eliminate a term with an offline referent without removing a major online phenomenon from consideration.

After the initial set of stimuli words was assembled, a dozen sociology graduate students (a group with average Internet savvy) were surveyed to identify the most familiar online terms, and items were ranked according to the number of students indicating familiarity with each term. With this information, the length of the list of Internet-related stimuli was reduced from more than 60 to a more manageable 30, distributed relatively evenly between online identities, behaviors, and settings. The 30 stimuli are listed and defined in Tables 1, 2, and 3.

RECRUITMENT

Respondents were recruited on the World Wide Web by an advertising banner placed on the Yahoo! search engine. The animated banner contained drawings of computer monitors with cartoon faces. As the animation progressed, the facial expressions of the computers changed from happy to sad and back again. Between the monitors was a text message: "Make your feelings known," interchanging with "Click here to participate in the 1998 Internet User Survey."

The researcher had originally intended to purchase advertising space on Yahoo!, but such advertising was prohibitively expensive. Fortunately, Yahoo! Inc. devotes a portion of its advertising space to nonprofit and public service advertisements. After a period of review and negotiation, Yahoo! Inc. agreed to give 6 months' worth of pro-bono advertising to the survey project. (See Snyder & Rosenbaum, 1997, for a discussion of the issues surrounding the use of Internet advertising by not-for-profit organizations.)

Yahoo! was the most visited Internet site in the world in the months preceding the survey, with hundreds of millions of "visits" a year (RelevantKnowledge, 1998). The broad interna-

TABLE 1
Glossary of Online Behaviors

Online Behavior	*Definition*
Chat	To communicate with someone online in a real-time, typically text-based fashion
Crosspost	To post a single message simultaneously to multiple newsgroups or discussion lists
Cybersex	A two-way text-based interaction involving vivid descriptions of sexual activities, the Internet equivalent of phone sex
E-mail	To communicate with someone via electronic mail
Flame	To use e-mail or public postings to criticize someone personally or to attack their online utterances in an overly harsh, disproportionate fashion
Flamebait	To make outrageous statements over e-mail or in public postings in hope of getting someone to flame you and bring about a general flame war (a lengthy, overheated argument)
Ignore	To ignore someone's online communication or behavior
Mailbomb	To send gigantic e-mails or multitudes of e-mails to someone in an attempt to crash their e-mail program and make it impossible for them to read their regular e-mail
Post	To display a written set of observations, comments, or other information in a newsgroup or discussion list (much like posting information on a real-world bulletin board)
Shout	To type in all capital letters, typically in a posting or an e-mail
Spam	To e-mail or post unwanted or inappropriate information, most often refers to unsolicited advertisements
Surf the Web	To peruse resources on the World Wide Web, especially in an unplanned, nonfocused way

TABLE 2
Glossary of Online Identities

Identity	*Definition*
AOL'er	A member of the major Internet service provider America Online™
'Bot	A program most often found on Internet relay chat and multiple-user domains and written to perform useful online functions; Bots are sometimes designed to emulate a real person in social interactions
Cracker	Someone who uses technical prowess to illegally access and/or modify computer resources (often contrasted with the similarly capable although more benevolent hacker)
Cyberpunk	A kind of online desperado; the term comes from the cyberpunk science fiction genre of the late 1980s and early 1990s
Hacker	A technical virtuoso motivated by curiosity about computers and free access to information
Lurker	Someone who spends time in chatrooms, Internet relay chat channels, multiple-user domains, or newgroups but just "watches" rather than contributing anything to the discourse; some estimates have lurkers outnumbering actual participants 10 to 1, although measuring the unseen is always difficult
Netizen	A citizen of the online community (the net)
Newbie	A newcomer to the online environment
Webmaster	An individual who administers Web sites, maintains access, organizes resources, and generally helps to keep the online environment up and running

TABLE 3
Glossary of Online Settings

Online Setting	*Definition*
Chatroom	A typically Web-based environment in which individuals can adopt pseudonymous "handles" and engage in text-based, real-time interaction
Cyberspace	A blanket term used to refer to the entirety of the online environment, commonly attributed to William Gibson's (1984) novel *Neuromancer*
Information Superhighway	A term loosely describing the entirety of the online environment, often attributed to Al Gore's early 1990s speeches supporting the National Information Infrastructure
Internet	The vast telecommunications network that links together computer networks from around the world
Internet relay chat (IRC)	A worldwide system of real-time text-based discussions organized into arenas called channels and mediated by a large number of IRC servers at various locations; unlike a chatroom, IRC usually refers to a Telnet-based rather than Web-based discussion process
Multiple-user domain (MUD)	Originally a multiple-user dungeon, text-based online settings ranging from medieval fantasies to science fiction locales in which users create personae and interact with one another; MUDs differ from IRC and chatrooms in their emphasis on fictional personae and pseudo-physical settings
Newsgroup	A discussion forum to which people may post messages and comments; like a bulletin-board, the interaction is not real time; the term usually refers to usenet, a system of thousands of hierarchically arranged discussion groups on a multitude of topics
Real world	The offline (non-Internet) world
World Wide Web	The huge system of interlinked Internet content pages written in HTML format and displayed by Web browser software

tional exposure of Yahoo! helped the survey advertisement reach a large and diverse audience. In addition, due to the Web-based nature of Yahoo! and its popularity among Internet users, it seemed to be a location that would attract respondents with Internet savvy.

The advertising banner was placed into general circulation by Yahoo! Inc., meaning that it came up randomly as people used the Yahoo! site. Upon seeing the survey advertisement, a Yahoo! user had the option of clicking on the advertisement or ignoring it. If the Yahoo! user clicked on the advertisement, he or she was taken to the survey consent form (a standard human participants informed-consent form), which concluded with a hyperlink taking the respondent to the survey form itself.

The Yahoo! advertisement ran from May 1, 1998, to October 15, 1998. During that period, the banner advertisement was displayed 6,230,660 times, a mean of 37, 087 times per day. (This number was much lower on weekends due to the lower traffic on Yahoo! during weekends.) During the advertisement period, the banner was clicked 28,388 times. In other words, someone went to the survey site itself from Yahoo! about 170 times per day or about once out of every 200 times that the advertisement appeared (automated personal communication, Yahoo! Inc., October 31, 1998.).

Of the approximately 4,500 people per month who went from Yahoo! to the consent form at the survey site, approximately 40% (about 1,750 per month) viewed the survey form after reading the consent form (automated personal communication, Indiana University Information Technology Services, October 31, 1998.) About a quarter of the people who looked at the survey page completed the survey (about 442 per month, or about 15 per day). The total sample size from the survey during the 6-month run was 2,431 cases.

INSTRUMENTATION

Figure 1 is the example item given to respondents at the beginning of the survey. (The complete instrument is available at `http://php.indiana.edu/~adking/survey_form.htm` or upon request.) Below each stimulus were rating scales for evaluation (good or bad), potency (powerful or weak), and activity (active or inactive). The endpoints were anchored by words that in normal English tend to indicate the dimension that is being considered. For example, the first scale, the Evaluation dimension, is anchored by *bad, awful* on the negative end and *good, nice* on the positive end. The respondents were asked to click a position on each of the three scales below each item to indicate their ratings.

Concept ratings were coded on a 9-point scale, with *infinitely bad* (in the case of the Evaluation dimension) being coded as –4, *infinitely good* being coded as 4, and the *neutral* midpoint coded as 0. The intermediate values were marked off in whole numbers in each direction from the midpoint; that is, *slightly good* was coded as 1, *slightly bad* was coded as –1, *quite good* was coded as 2, *quite bad* was coded as –2, and so forth. The Potency and Activity dimensions were coded the same way.

To the left of each stimulus was a button marked with a question mark, to be used when the respondent did not know or otherwise could or would not respond to a stimulus item. The respondents were told in the survey instructions that they were to assign either ratings or a question mark (not applicable) response for each item. If respondents did not respond to every stimulus, the data-recording computer program would notify them and refuse to accept the data until all items were assigned either ratings or question mark responses.

In addition to concept ratings, the survey also included a set of demographic questions. One demographic variable (gender) preceded the main stimulus-rating portion of the instrument, whereas the rest of the demographic variables appeared at the end of the survey. The demographic questions were as follows:

What is your gender? Responses were coded 0 for male and 1 for female.

How many years have you spent online? Respondents could select one of several response categories, for example, "less than 6 months," "6 months to a year," "1 to 3 years," and so forth. Responses were converted into a continuous-type variable for the purposes of analysis by assigning a midpoint value of the category selected. It was then log transformed in subsequent analyses to correct for sizable positive skew in the data.

How much time do you spend online per week? (categorical variable; "less than 1 hour," "1 to 3 hours," and so forth). Responses were converted to a continuous-type variable for analysis by assigning a midpoint value of the category selected.

How old are you? (categorical variable; "less than 10 years old," "10 to 15 years old," and so forth). Responses were converted to a continuous-type variable for analysis by assigning a midpoint value of the category selected.

How do you access the Internet? Respondents could select one of several forms of access: local Internet access provider, major online service (e.g., AOL, AT&T, MSN), educational provider, and so forth.

In addition to questionnaire responses, some extra information was recorded with each survey response by the computer program. This information included the exact time and date of the survey response, the type of Web browser being used, and the Internet domain where the survey response originated. Neither the survey instrument itself nor the data-recording program gathered any individual identification information (e-mail addresses, names, and so forth) from respondents.

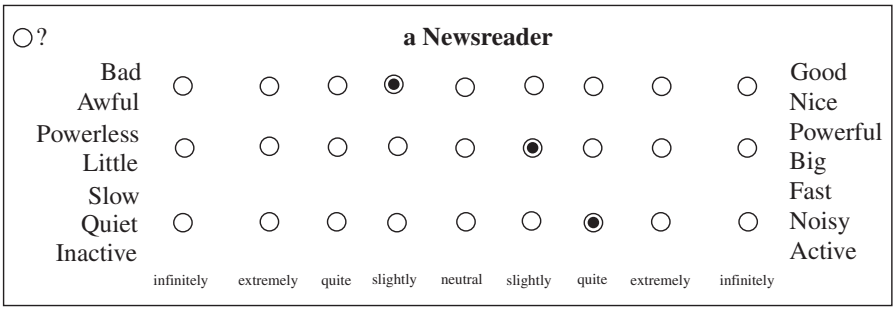

Figure 1: Survey Item Example

SAMPLE DEMOGRAPHICS

At the time of the survey, the Internet was still viewed as a "men-only camp," with men far outnumbering women. From that perspective, the gender distribution of the sample—56% male, 44% female— would seem rather surprising. However, in recent years, the gender ratio of the Internet has quickly leveled off (Pew Research Center, 1999).

Other characteristics of the sample also paralleled known Internet user characteristics. For example, 73% of Internet users in the sample were between the ages of 18 and 44 years, peaking between 25 and 35 years. This distribution corresponds well with the age distributions discovered in studies of Internet demographics (CyberAtlas, 1999). Of the respondents, 90% reported being online for less than 7 years, which also is typical of Internet users. Although most were relative newcomers, only 2% of the respondents reported spending less than 1 hour online per week, and most claimed between 10 and 20 hours online per week. Hence, the respondents were almost uniformly people who are invested and interested in the online environment.

The sample followed a global distribution typical for Internet users, with the United States of America accounting for 82% of the responses and Canada, Europe, and Asia making up much of the rest. That the survey was only available in English may explain why respondents from many areas of the world were even more rare than would be expected from their limited Internet access. An Antarctica case was recorded; however, it was likely a joke (the Internet domain reported for that case does not match the domains of the Antarctic research stations).

Three quarters of the respondents reported accessing the Internet through a local Internet access provider (50%) or a major online service (25%). These proportions have been changing during the past several years, as new Internet users are more likely to be users of a major online service than are Internet veterans (CyberAtlas, 1999).

In summary, the "typical" respondent to the survey was an American between the ages of 25 and 35 who spends between 10 and 20 hours online per week, accesses the Internet through a local Internet service provider, has 1 to 3 years of online experience, and is marginally more likely to be male than female.

ONLINE CULTURAL SENTIMENTS

Tables 4 through 6 report mean EPA ratings of all stimuli for both men and women. Because respondents could skip stimuli instead of rating them, the sample size for each stim-

(text continues on p. 423)

TABLE 4
EPA Ratings of Online Behaviors

Concept	N	*Evaluation*	*Potency*	*Activity*	*Words With Similar Profiles*
Ignore					
Male	1,295	–0.47	0.47	–0.34	Shush, reproach, patronize, admonish, hush
Female	1,002	–0.63*	0.73***	–0.41	Silence, reproach, subdue, shush, admonish
Surf the Web					
Male	1,351	2.58	2.17	1.49	Save, make love to, cheer, rescue, delight
Female	1,048	2.82***	2.36***	1.68**	Delight, rescue, love, amuse, enjoy
Spam					
Male	1,234	–3.03	–0.41	1.08	Steal from, poison, disrespect, double-cross
Female	809	–3.20**	0.14***	1.39***	Slay, steal from, hate, torment, murder
Cybersex					
Male	1,299	–0.12	–0.11	0.37	Butter up, eye, beseech, implore, coddle
Female	993	–1.10***	–0.45***	0.32	Deprecate, berate, misjudge, displease, deride
Chat					
Male	1,329	2.03	1.18	1.17	Laugh with, desire sexually, amuse, court
Female	1,051	2.45***	1.53***	1.50***	Delight, rescue, amuse, dance with, enjoy
Crosspost					
Male	913	–0.58	0.11	0.65	Cajole, butter up, bluff, exhibit, prod
Female	559	–0.01***	0.49***	0.65	Kid, waken, halt, question, entreat
Flame					
Male	1,059	–1.57	–0.08	1.17	Quarrel with, cuss, heckle, pester, mock
Female	606	–2.19***	0.01	1.26	Harass, curse, humiliate, disrespect, insult
Flamebait					
Male	778	–1.63	–0.63	0.46	Nag, mislead, distress, hoodwink, browbeat
Female	395	–2.19***	–0.40*	0.61	Bribe, shame, ridicule, degrade, nag
E-mail					
Male	1,346	3.14	2.62	1.80	Rescue, make love to, love, cheer, delight
Female	1,050	3.41***	2.79***	2.05***	Delight, love, rescue, dance with, amuse
Shout					
Male	1,303	–1.66	–0.59	0.96	Nag, heckle, distress, annoy, disgrace
Female	1,001	–1.45***	0.01***	1.30***	Pester, taunt, bother, mace, aggravate

TABLE 4 Continued

Concept	N	*Evaluation*	*Potency*	*Activity*	*Words With Similar Profiles*
Post					
Male	1,307	1.99	1.50	1.12	Delight, laugh with, amuse, speak to
Female	992	2.15**	1.67**	1.18	Delight, amuse, dance with, interest, entertain
Mailbomb					
Male	1,118	–2.74	0.27	0.86	Hate, humiliate, persecute, harm, disable
Female	707	–3.19***	0.24	1.06*	Steal from, slay, hate, murder, disable

NOTE: EPA = evaluation, potency, and activity.
*$p < .05$. **$p < .01$. ***$p < .001$ (t test of sex differences).

TABLE 5
EPA Ratings of Online Identities

Concept	N	*Evaluation*	*Potency*	*Activity*	*Words With Similar Profiles*
Newbie					
Male	1,171	–0.15	–1.84	–0.76	Waif, orphan, beginner, simpleton, klutz
Female	770	0.32***	–1.50***	–0.65	Victim, shrimp, nobody, dummy, easy mark
Hacker					
Male	1,333	–0.93	2.00	1.42	Bouncer, tough, rival, racketeer, he-man
Female	1,017	–1.91***	2.04	1.48	Gangster, mobster, desperado, assailant
AOL'er					
Male	1,259	–1.28	–1.48	–0.47	Blockhead, bum, vagrant, bonehead
Female	915	–0.55***	–0.83***	–0.15***	Ninny, hanger-on, dimwit, scapegoat
Cracker					
Male	831	–0.80	1.12	0.81	Highwayman, big-shot, safecracker, spy
Female	334	–0.94	0.82**	0.66	Nark, ghoul, bookie, firebug, safecracker
Netizen					
Male	1,040	1.41	0.94	0.94	Prodigy, chum, magician, wit, man
Female	579	1.54*	1.20***	1.12**	Prodigy, guy, spouse, teammate, man
Bot					
Male	838	0.41	0.83	0.54	Lookout, saloon keeper, graduate student
Female	386	0.33	0.94	0.58	Lookout, insider, referee, barkeeper

(continued)

TABLE 5 Continued

Concept	N	*Evaluation*	*Potency*	*Activity*	*Words With Similar Profiles*
Cyberpunk					
Male	1,060	–0.92	–0.35	0.82	Gigolo, firebug, cynic, crackpot, weirdo
Female	703	–1.42***	–0.34	1.07***	Cad, fugitive, crackpot, maniac, scamp
Lurker					
Male	933	–0.39	–0.49	–1.10	Square, loner, bookworm, homebody
Female	642	–0.77***	–0.80***	–1.31**	Drudge, zombie, introvert, wallflower
Webmaster					
Male	1,262	1.72	1.89	0.96	Brain, expert, lover, hero, airline pilot
Female	865	1.85*	2.06**	1.27***	Fireman, heroine, hero, coach, brain

NOTE: EPA = evaluation, potency, and activity.
*$p < .05$. **$p < .01$. ***$p < .001$ (t test of sex differences).

TABLE 6
EPA Ratings of Online Settings

Concept	N	*Evaluation*	*Potency*	*Activity*	*Words with Similar Profiles*
Chatroom					
Male	1,329	0.48	0.23	1.40	Locker room, hangout, bull session, saloon, bar
Female	1,028	0.60*	0.40**	1.49	Bull session, nightclub, cocktail lounge, hangout, bar
Newsgroup					
Male	1,205	1.15	0.99	1.00	Debate, stage, classroom, seminar, banquet
Female	799	1.07	0.96	1.03	News stand, stage, cocktail lounge, computer center
Cyberspace					
Male	1,324	2.24	2.48	2.14	Celebration, fire station, ball game, college, Christmas
Female	1,005	2.34	2.74***	2.38***	College, airport, football stadium, car, pep rally
Real world					
Male	1,312	1.59	2.05	2.10	Football stadium, airport, car, festival, ball game
Female	999	1.69	2.11	2.24*	Football stadium, airport, pep rally, festival, university
Internet					
Male	1,351	2.92	3.10	2.58	Celebration, Christmas, fire station, college, ball game
Female	1,056	2.96	3.30***	2.82***	College, Christmas, beach, vacation, celebration

TABLE 6 Continued

Concept	N	*Evaluation*	*Potency*	*Activity*	*Words with Similar Profiles*
Information Superhighway					
Male	1,333	2.23	2.42	2.10	Celebration, fire station, ball game, college, Christmas
Female	1,024	2.82***	3.03***	2.59***	College, Christmas, beach, vacation, celebration
Multiple-user domain					
Male	938	1.08	0.80	1.01	Debate, journey, waterbed, movie, boardwalk
Female	546	1.28**	1.25***	1.34***	Television studio, mall, clubhouse, penthouse, stage
World Wide Web					
Male	1,343	2.87	2.94	2.32	Christmas, celebration, fire station, college, ball game
Female	1,046	3.03**	3.23***	2.67***	Christmas, college, beach, vacation, celebration
Internet relay chat					
Male	1,069	1.27	1.13	1.54	Athletic club, festival, debate, nightclub, bar
Female	685	1.52***	1.48***	1.75**	Festival, ball game, parade, New Years' Eve

NOTE: EPA = evaluation, potency, and activity.
*$p < .05$. **$p < .01$. ***$p < .001$ (t test of sex differences).

ulus is also given. Statistically significant differences between mean ratings given by men and women on each stimulus are indicated next to the female ratings. The significance results are based on two-tail t tests.

Because the EPA mean ratings are difficult to interpret as mere numbers, words with similar EPA profiles are given for each stimulus. Semantic equivalents were derived by matching online survey EPA ratings to common English words with comparable ratings found in a database of EPA profiles (Heise, 1998). For example, the semantic equivalents of *newbie* (which refers to someone who is new to the online environment) include the word *beginner*, and the semantic equivalents of *cracker* (which refers to people who use their technical prowess to gain illegal access to computing resources) intriguingly include the word's predecessor, *safecracker*.

The results in Table 4 indicate that e-mail, post, surf the Web, and chat were consistently given positive ratings on all three EPA dimensions. Negative online behaviors generated more subtly differentiated mean ratings than positive behaviors, and an apparent rank order developed: Most mildly, to "ignore" someone online seems a somewhat negative and powerful behavior. More negative than online ignoring is engaging in cybersex. Crossposting, flamebaiting, and flaming, staple behaviors of usenet newsgroups, are the next-worse group of online behaviors. SHOUTing, mailbombing, and spamming, in descending order of "goodness," round out the bottom of the scale. The most extreme sentiments, both positive and negative, relate to e-mail, suggesting that the most common application of the Internet also is the one with the most cathexis.

Table 5 shows sentiments associated with Internet identities. At the top of the good-bad (evaluation) hierarchy are webmaster, which seems to be the grand magician of the Internet, and netizen, which refers to fellow Internet users and perhaps the respondents themselves. In

the middle of the range, near the neutral point of the scales, is bot, which refers to the relatively innocuous yet helpful programs that are used for a variety of purposes on the Internet. Lower down on the Evaluation dimension come identities such as newbie, lurker, cyberpunk, AOL'er, hacker, and cracker. Most of these identities are described in dictionaries of online terminology as negative in character, and that judgment is confirmed by these data. In general, a kind of xenophobic rejection seems to produce one class of deviants (newbie, lurker, AOL'er), and a concern over secretive and perhaps damaging online activities produces a more extreme class of deviants (hacker, cracker).

Table 6 suggests that Internet users consider the online environment to be a marvelous place that is overwhelmingly good, potent, and dynamic. To draw on some of the semantic equivalents, the Internet is a "celebration," as magical as "Christmas," as enjoyable as a "vacation" at the "beach." Although still high on all three EPA dimensions, "the Real World" was rated lower on all dimensions than the Internet, the Information Superhighway, the World Wide Web, or Cyberspace.

An interesting gap appears between general terms for the online world (Internet, Information Superhighway, World Wide Web, Cyberspace) and terms referring to specific subareas of the Internet (Internet relay chat, chatrooms, multiple-user domains, newsgroups): Without exception, general terms for the online world receive more intense ratings on all dimensions than do the specific online location terms. This gap could indicate that respondents' sentiments toward the online world as a whole are not simply the average of their sentiments toward the parts.

GENERAL PATTERNS

A striking pattern evident in the survey results is the extremity of the ratings assigned to most of the stimuli. EPA results are usually dominated by mean ratings in the range of –1.5 to 1.5, with ratings higher than 2.5 or lower than –2.5 tending to be reserved for archetypical examples of the dimension. Identities normally receiving such extreme ratings include God and Devil, settings receiving such extreme ratings include Christmas and Skid Row, and behaviors receiving such extreme ratings include rescue and annihilate. Many of the mean ratings for online identities, behaviors, and settings end up in ranges that would be considered unusually extreme in other data sets.

The mean ratings gathered by this survey were often so extreme that finding semantic equivalents with matching EPA profiles was difficult. For instance, many online settings have EPA profiles that only match the extremes of celebration, Christmas, college, fire station, and ball game. Similarly, EPA profiles for several online social behaviors (e-mail and surf the Web) curiously match those for sexualized semantic equivalents such as love and make love to. The online ratings are so intense that finding the closest semantic equivalent was tantamount to searching for common English words with the most intense affective associations.

GENDER DIFFERENCES

The relative intensity of the online EPA ratings varied somewhat across genders. Of the 90 mean ratings gathered by the survey, 65 (72%) had differences across genders that were statistically significant at the .05 level or better. Of these differences, 24 were on the Evaluation dimension, 23 were on the Potency dimension, and 18 were on the Activity dimension.

The vast majority of the 65 gender-related concept variations were differences of intensity. Although agreeing on the general direction of the ratings, women rated most items

TABLE 7
Gender Correlations With Concept Ratings (1 = female)

	Evaluation	*Potency*	*Activity*
Cybersex	–0.19**		
Crosspost	0.13**		
E-mail	0.12**		
Flame	–0.14**		
Flamebait	–0.13		
Mailbomb	–0.11**		
Cyberpunk	–0.11**		
Hacker	–0.20**		
Spam		0.10	
Shout		0.13**	
Internet		0.08	
Multiple-user domain		0.11	
Webmaster			0.08
Newbie	0.12**	0.09	
AOL'er	0.16**	0.15**	
World Wide Web		0.11**	0.10**
Chat	0.13**	0.10**	0.10**
Information superhighway	0.15**	0.17**	0.12**

NOTE: Due to the large number of nonsignificant correlations with concept ratings, this table only reports correlations where $p < .05$ or better. The following concepts had no significant correlations ($p < .05$) with gender: ignore, surf the Web, post, cracker, netizen, bot, lurker, chatroom, newsgroup, cyberspace, real world, Internet relay chat.
**$p < .01$ (Bonferoni-adjusted pairwise correlations).

slightly more intensely than did men. Male-female differences were more complicated in a few cases. For example, men feel that cybersex is a harmless, flirtatious behavior, whereas women feel it is an aggressive, threatening behavior. Gender differences in online interaction generally reflect rather subtle differences in online cultural sentiments, but important variations do exist in male and female sentiments toward online concepts. Exactly why men and women feel differently toward certain online cultural objects that are seemingly "gender-neutral" is a matter requiring further study. Structural factors, linguistic influences, and differences in prior experiences (Kiesler & Sproull, 1985) may play a factor in these sentiment variations as they do for all cultural sentiments. In any case, although men and women share a great deal of common online culture, gender differences in sentiments are quite visible, and these differences presumably are relevant to gender differences in social interaction in cyberspace (Herring, 2000; Kendall, 1998).

CONCEPT DEMOGRAPHICS

Tables 8, 9, and 10 show correlations of behavior, identity, and concept ratings with measures of respondent age, time spent online per week, and total years spent online. The age of Internet users significantly correlated with a number of Internet behavior ratings (see Table 8). In particular, younger respondents were more positive about chat, ignore, and mailbomb. Table 9 shows that younger Internet users also more positively rated three Internet identities (hacker, cracker, and cyberpunk). On the other hand, older users tend to rate newcomer identities such as newbie and AOL'er more positively than do younger users. When it comes to Internet settings (see Table 10), older users tend to denigrate every aspect

TABLE 8
Correlations: Behaviors and Demographic Items

Dimension		*Age*	*Hours Online per Week*	*Years Online (log)*
Ignore	E	–0.11**	0.16**	0.08**
Chat	E	–0.09		–0.14**
	P			–0.12**
	A	–0.10**		–0.07**
Crosspost	E		–0.16**	–0.19**
	P		–0.16**	–0.12**
Shout	E		–0.12**	–0.22**
	P		–0.12**	–0.18**
Mailbomb	E	–0.10	0.10	
	P	–0.11**		
	A			0.08**
Spam	E			–0.09**
	A			0.09**
Post	E			–0.06**
	P			–0.04

NOTE: E = evaluation; P = potency; A = activity. Due to the large number of nonsignificant correlations with concept ratings, this table only reports correlations where $p < .05$ or better. Ratings for surf the Web, cybersex, flame, flamebait, and e-mail had one or fewer correlations with the demographic variables significant at the .05 level.
**$p < .01$ (Bonferoni-adjusted pairwise correlations).

TABLE 9
Correlations: Identities and Demographic Items

Dimension		*Age*	*Hours Online per Week*	*Years Online (log)*
Newbie	E	0.16**	–0.11**	–0.13**
	P	0.16**	–0.12**	–0.16**
Hacker	E	–0.18**	0.18**	0.19**
	P	–0.14**		0.08
AOL'er	E	0.16**	–0.20**	–0.24**
	P	0.15**	–0.19**	–0.26**
	A			–0.04
Cracker	E	–0.22**		
	P	–0.14**		0.13
	A			0.07
Cyberpunk	E	–0.16**	0.12**	0.16**
Lurker	E		0.12**	0.19**
	A			–0.09

NOTE: E = evaluation; P = potency; A = activity. Due to the large number of nonsignificant correlations with concept ratings, this table only reports correlations where $p < .05$ or better. Ratings for bot, Webmaster, and netizen had one or fewer correlations with the demographic variables significant at the .05 level.
**$p < .01$ (Bonferoni-adjusted pairwise correlations).

(every EPA dimension) of chat rooms but otherwise tend to rate most online settings the same or slightly more positively than do younger users.

Immersion in the Internet environment (as indicated by time spent online every week) seems to affect Internet behavior sentiments, but only in regard to generally negative behav-

TABLE 10
Correlations: Identities and Demographic Items

Dimension		*Age*	*Hours Online per Week*	*Years Online (log)*
Chatroom	E	–0.19**	0.12**	–0.08**
	P	–0.13**	0.10**	
	A	–0.12**		0.09**
Newsgroup	E	0.09		0.05
	P			0.06**
	A			0.11**
Real world	E		–0.09**	–0.04
	P		–0.11**	–0.05
	A		–0.13**	–0.05
Information Superhighway	E	0.10**		–0.15**
	P			–0.13**
	A			–0.10**
Multiple-user domain	E			–0.07**
	P			–0.13**
	A			–0.06
World Wide Web	E	0.09		–0.05
	P			–0.05**
Internet relay chat	E			–0.08**
	P			–0.06**

NOTE: E = evaluation; P = potency; A = activity. Due to the large number of nonsignificant correlations with concept ratings, this table only reports correlations where $p < .05$ or better. Cyberspace, Internet, multiple-user domain, and Internet relay chat had one or fewer correlations with the demographic variables significant at the .05 level.
**$p < .01$ (Bonferoni-adjusted pairwise correlations).

iors: Users who spend more time online are more positively disposed toward ignoring and mailbombing but more negatively disposed toward crossposting and shouting behaviors. Time spent online also seems to make users evaluate generally negative identities such as hacker, cyberpunk, and lurker more positively but rate newcomer identities such as newbie and AOL'er more negatively. Users spending more time on the Internet every week seem to be more positive than average about chat rooms but more negative about every aspect (EPA dimension) of the real world.

Cumulative experience on the Internet, as measured by the natural log of the years spent online, is negatively correlated with age ($r = –.07$, $p < .01$). The years spent online has an impact on concept ratings that occasionally parallels the effects of age. Like age, years spent online has a negative correlation with the behaviors of chatting and crossposting. Years spent online also resembles age in its negative correlation with ratings for chat room and its positive correlation with ratings for newsgroup. However, unlike age, cumulative years spent online is correlated negatively with ratings for SHOUT and World Wide Web.

These inconsistencies and the negative correlation between age and years spent online are puzzling. One would think that age and cumulative experience at an activity would be strongly positively correlated. A possible answer lies in the unique makeup of the online population: The youngest respondents often had a great deal of accumulated experience for their age, although their accumulation of experience is ultimately limited by their youth. On

TABLE 11
Correlations: Demographic Measures

	Gender (female = 1)	*Age*	*Hours Online per Week*
Age	–0.02		
Hours online per week	–0.12***	–0.09***	
Years online (log)	–0.23***	–0.07**	0.23***

*$p < .05$. **$p < .01$. ***$p < .001$.

the other hand, older respondents tended to adopt Internet technology more slowly than younger users and often had little accumulated experience for their age. In between these age groups is the large set of respondents between the ages of 25 and 35 who possess both age and accumulated Internet experience. Due to these cohort-related differences, when discussing age and Internet experience patterns, focusing on the different age and experience cohorts would give a more accurate picture than assuming a simple linear relationship between these two factors.

The effects of years spent online on concept ratings have a close symmetry to the effects of immersion: Like years spent online, time spent online is negatively correlated with newbie and AOL'er but positively correlated with hacker, cyberpunk, and lurker. This symmetry of effects reflects the significant positive correlation ($r = .23$, $p < .001$) between the years spent online and the hours spent online per week. The correlation between years spent online and hours spent online per week can be interpreted in two different ways: (a) People who have spent many years on the Internet are likely to become more immersed in the Internet culture on a daily basis, and (b) people who are more immersed in the Internet culture on a daily basis are more likely to continue to use it for a long period of time. Some combination of these two tendencies is the most likely explanation.

Worth noting are the correlations between gender and the demographic variables (see Table 11): Time spent online per week and years spent online are negatively correlated with gender (female = 1), suggesting that female users are less immersed in and experienced with the Internet environment. These gender-related differences may be the legacy of the early male-dominated days of the Internet, and as such, they may decline or even disappear as the Internet achieves gender parity.

DISCUSSION

Internet researchers have long argued that a unique culture underlies online social interaction, a culture similar to but separate from the everyday culture of the offline world. If such a culture existed, it would be revealed by a set of shared affective sentiments held by Internet users toward online concepts. This study measured the affective sentiments of Internet users toward behaviors, settings, and identities common to the online world using a Web-based survey instrument. These data, when aggregated across a sample of Internet users, could be used to empirically evaluate the claim of a unified online culture.

The findings of this study indicate that Internet users share a large set of common affective sentiments toward Internet-related concepts. These shared sentiments can be thought of as "cultural sentiments," which codify and make meaningful the online social interactions of Internet users. Although there are slight variations in these cultural sentiments related to gender, age, immersion, and cumulative Internet experience, the overall picture is one of cultural

consistency and stability. Despite its transitory appearance, the Internet social world does have a memory, one that keeps and is kept by the shared affective sentiments comprising online culture.

The study also found that the online world is a potent cultural environment subjectively built on fundamental sentiments much more extreme than those typically invoked in other environments. One might expect that as these intense settings, identities, and behaviors are manipulated during online interactions, outrageous emotional displays and profound mood changes would be commonplace. Indeed, casual observation and popular culture suggest that the online social environment fits this expectation, with extreme emotional displays and disinhibited behavior being the norm rather than the exception. However, these accounts tend to exaggerate the disinhibited elements of online culture (Lea, O'Shea, Fung, & Spears, 1992) and overlook the cultural stability of the online environment. Although the online social environment is powerfully charged with affect, that affective charge appears to be invoked very judiciously in online interaction.

Fundamental cultural sentiments, such as those found in this study, do not mandate extraordinary social responses toward cultural objects. Instead, extreme sentiments held toward the actors, behaviors, and settings of social interactions yield familiar behaviors and emotions that are very distinctive for the situation (Heise, 1977; Smith-Lovin & Heise, 1988). Thus, the extreme sentiments of online culture work no differently than less extreme sentiments do in other cultures. On the other hand, in the face of an extremely destabilizing influence or event, the powerful sentiments held by Internet users may accentuate socioemotional expressions and behaviors to levels much more extreme than those seen in the offline world. "Flame wars" are real, although they are less common than many accounts suggest. Still, this potential for extreme responses does not totally undermine cultural stability. Indeed, shared cultural sentiments, even if they are extreme, encourage cultural stability because they make visible the behavioral expectations embedded in that culture.

This study is somewhat limited by its dependence on Yahoo! for the recruitment of the sample. It is possible that some types of Internet users may be more inclined than others to use Yahoo!, may be more inclined to notice or respond to advertisements, or may be more likely than other users to respond to online surveys. These biases are problematic but also might weight the sample with Internet-savvy Web users who are integrated into online experiences and are inclined to discuss and participate in online activities, shifting the sample in a desirable direction for the purposes of this study. This issue could best be addressed by replications of this research using other sampling methods.

By investigating the shared affective sentiments underlying online culture, this study provides a unique perspective on social interaction on the Internet. However, many questions remain: The survey found that gender differences in online cultural sentiments are slight but nonetheless meaningful. Further inquiries into the relationship between gender and online culture are called for. The findings of this study also suggest that online cultural sentiments are typically much more extreme than offline cultural sentiments. Examining the situational and long-term consequences of this affective extremity would be a valuable next step. Finally, this study found suggestive relationships between various demographic characteristics and online cultural sentiments, relationships that should be investigated in more detail. Although the overarching online culture is unified by a set of shared cultural sentiments, online subcultures and special interest groups may exhibit unique patterns of sentiments toward online identities, behaviors, and settings. Further research into the cultural and affective roots of online social interaction is strongly encouraged.

REFERENCES

Barry, J. (1991). *Technobabble*. Cambridge, MA: MIT Press.

Bertot, J. C., & McClure, C. R. (1997). Electronic surveys: Methodological implications for using the World Wide Web to collect survey data. In S. Hardin (Ed.), *Proceedings of the 59th annual meeting of the American Society for Information Science* (pp. 173-185). Medford, NJ: Information Today, Inc.

Crumlish, C. (1995). *The Internet dictionary.* San Francisco: Sybex.

CyberAtlas. (1999). *AOL surpasses local ISPs* [Online]. Retrieved March 25, 1999, from the World Wide Web: `http://www.cyberatlas.com/big_picture/traffic_patterns/aol.html`

Fahey, T. (1994). *Net.Speak Internet dictionary.* Indianapolis, IN: Hayden.

Gibson, W. (1984). *Neuromancer.* New York: Ace Books.

Heise, D. R. (1977). Social action as the control of affect. *Behavioral Science*, *22*, 163-177.

Heise, D. R. (1998). *An online dictionary of cultural sentiments* [Online]. Retrieved January 4, 1999, from the World Wide Web: `http://www.indiana.edu/~socpsy/ACT`

Herring, S. (2000). Gender differences in CMC: Findings and implications. *Computer Professionals for Social Responsibility Newsletter* [Online serial], *8*(17). Retrieved August 16, 2001 from the World Wide Web: `http://www.Cpsr.org/publications/newsletters/issues/2000/Winter2000/herring.html`

Jones, S. G. (Ed.). (1995). *CyberSociety: Computer-mediated communication and community.* Thousand Oaks, CA: Sage.

Kendall, L. (1998). Meaning and identity in "cyberspace": The performance of gender, class, and race online. *Symbolic Interaction*, *21*, 129-153.

Kiesler, S. B., & Sproull, L. S. (1985). Pool halls, chips, and war games: Women in the culture of computing. *Psychology of Women Quarterly, 9*, 451-462.

Lea, M., O'Shea, T., Fung, P., & Spears, R. (1992). "Flaming" in computer-mediated communication: Observations, explanations, implication. In M. Lea (Ed.), *Contexts of computer-mediated communication* (pp. 89-112). London: Harvester-Weatsheaf.

MacKinnon, N. (1994). *Symbolic interactionism as affect control*. Albany: State University of New York Press.

Osgood, C. E., May, W., & Miron, M. (1975). *Cross-cultural universals of affective meaning*. Urbana: University of Illinois Press.

Osgood, C. E., & Tzeng, O.C.S. (1990). *Language, meaning, and culture: The selected papers of C. E. Osgood*. New York: Praeger.

Pew Research Center. (1999). *Who's online in America* [Online]. Retrieved March 25, 1999, from the World Wide Web: `http://www.cyberatlas.com/big_picture/demographics/who.html`

Raymond, E. (1993). *The new hacker's dictionary.* Cambridge, MA: MIT Press.

RelevantKnowledge. (1998). *Top dot.com domains, April 1998* [Online]. Retrieved May 11, 1998, from the World Wide Web: `http://www.iw.com/daily/stats/1998/05/1101-rk.html`

Smith-Lovin, L., & Heise, D. (1988). *Analyzing social interaction: Advances in affect control theory.* New York: Gordon & Breach Science.

Snyder, H., & Rosenbaum, H. (1997). Advertising on the World Wide Web: Issues and policies for not-for-profit organizations. In S. Hardin (Ed.), *Proceedings of the 59th annual meeting of the American Society for Information Science* (pp. 186-191). Medford, NJ: Information Today, Inc.

Walther, J. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, *19*, 52-90.

Adam B. King is a doctoral candidate in sociology at Indiana University. His major interests include social psychology, culture, science, and technology. His dissertation, from which this article is adapted, examines variations in the affective cultural sentiments of Internet users.

Cyberslacking and the Procrastination Superhighway:

A Web-Based Survey of Online Procrastination, Attitudes, and Emotion

JENNIFER A. A. LAVOIE
Simon Fraser University

TIMOTHY A. PYCHYL
Carleton University

This study was designed to explore the extent to which time spent online was related to self reports of procrastination. A sample of 308 participants (Mean age = 29.4 years, *SD* = 12.0, 198 females) from various regions of North America completed a survey posted to the World Wide Web. Data collected included demographic information, attitudes toward the Internet, amount of time spent online (at home, work, and school), trait procrastination, and measures of positive and negative emotion. Results demonstrated that 50.7% of the respondents reported frequent Internet procrastination, and respondents spent 47% of online time procrastinating. Internet procrastination was positively correlated with perceiving the Internet as entertaining, a relief from stress, and paradoxically, as an important tool. Internet procrastination was also positively correlated with trait procrastination and negative emotions. Implications regarding Internet procrastination are discussed in relation to procrastination theory and research as well as Neil Postman's critique of technology.

Keywords: Internet, procrastination, technology, volitional action, stress, coping

Internet use has evolved to become an essential method of communication, permeating academic, vocational, and domestic domains. Global Internet use is pervasive and is currently estimated at 300 million subscribers (Braziller, 2000), of which the United States and Canada have the highest Internet usage (59% and 56% of the national populations, respectively). In that the Internet provides a medium of electronic services to which the majority of North Americans have access and devote a considerable amount of time, the effect of Internet technology is certainly profound. It appears that Internet subscribers have embraced the Internet in daily living to "communicate, transact, entertain, educate and improve their connectivity and productivity" (Scruby, 1999, p. 2). Yet, with all of the acclaimed advantages for enhanced information management and the exponential adoption rate, does this new Internet technology fulfill the promises of technological efficiencies and progress, or is it subverted by some of our more mundane motivational or volitional problems such as procrastination?

AUTHORS' NOTE: The authors thank Lawrence Wong for his preparation of the online questionnaire, programming assistance, and pilot study.

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 431-444

In Postman's (1992) book *Technopoly: The Surrender of Culture to Technology*, he suggests that a mistake occurs in presuming that a technological innovation generates only a one-sided effect on society. He asserted that the notion that "progress" is a natural and inevitable advancing of humankind is an example of technological prejudice. Postman argued that the benefits of progress are believed to be attained through scientific and technological innovation and, more specifically, through the generation of information. Due to the associated link between progress and the speedy processing of data, a perception results that convenient production and distribution of information through technology intrinsically improves industriousness. In fact, Postman suggested that North American culture has developed a belief system that computer technology manifests a superior form of wisdom permitting increased efficiency for the user. Postman, however, questioned whether the generation of information through technology contributes to the solution of problems that progress is presumed to overcome.

A potential ramification of this false image of productivity promoted by computer use is the establishment of an environment conducive to wasting time. Recently, a lack of productivity in jobs in which Internet technology is involved has been acknowledged in popular literature. For example, a poll taken by a career Web site, `http://www.vault.com`, found that greater than 90% of the 1,244 employees sampled surfed the Internet for information irrelevant to their jobs while at work (Marron, 2000). Furthermore, the same poll found up to 25% of surveyed respondents indicated that while online, precautionary measures were taken to avoid being caught by supervisors. This growing trend known as "cyberslacking," the wasting of time and company resources by entertaining oneself on the Internet when one should be working (Marron, 2000), appears to be a modern, technologically mediated manifestation of procrastination. Further evidence of misappropriation of time spent using the Internet includes the finding by university administrators in New York that 43% of freshmen attrition was related to greater Internet access and inability to curtail personal computer over use (Wallace, 1999).

One theory addressing this notion of Internet procrastination has been proposed by Silver and Sabini (1981), who suggested that a particular style of procrastination involves the irrational fragmentation of time into short intervals to delay working toward a task. In this situation, the decision to work is not withdrawn. However, in a choice of an activity for the next immediate time period (e.g., the next 5 minutes), the individual justifies engaging in some minor pleasure instead of committing to the intended task. This decision is based on a rationalization that the amount of work that can be completed in the few minutes required to engage in a more pleasurable activity is minimal and justifies the procrastination. This cycle of "rational" task postponement can continue until the individual perceives that a 5-minute interval will be costly to the completion of the intended task. Silver and Sabini warn of the particular attractiveness of pleasurable activities that are brief and can be easily ceased as potential candidates for these short-term tasks choices. Notably, Internet activities such as managing e-mail or surfing the Web may be particularly alluring because these activities are quick and immediately rewarding and can be discontinued at will. The rationalization that checking one's email will only take a few minutes may be a popular form of procrastination.

Another aspect of procrastination relevant to cyberprocrastination suggested by Silver and Sabini (1981) is the notion of dramatizing commitment to a task. In the case in which an individual has not decided to "break" from working on a task, procrastination occurs through searching for off-task distractions while maintaining an appearance of accomplishing the intended task. According to Silver and Sabini, a "procrastination field" results through the attempt to display and maintain oneself in a position of readiness to work (to continue convincing oneself that one is working) and to seek diversions because the task is not actually

being accomplished. They suggested that an individual maintaining a procrastination field is especially vulnerable to any activity that does not require leaving the present location, involves minimum commitment, and lacks immediate averseness. Certainly, the Internet fits these criteria. While attempting to accomplish a task using the Internet, distractions in the form of flashing hyperlinks provide tempting distractions to an individual who is less focused on completing the task at hand. Minimum commitment is experienced due to the fact that these distractions require only a click of a mouse button to resume the original task. Finally, it is easy for the user to remain convinced that the task is still being accomplished because the user never left the computer desk to procrastinate.

Given the potential of the Internet as a tool for procrastination, the purpose of this research was to explore the nature of procrastination on the Internet. Information concerning feelings and attitudes toward the Internet, Internet time use, measures of general procrastination, and measures of composite affect (i.e., emotion) were collected via the World Wide Web. This research was guided by four major hypotheses relating procrastination to the Internet.

The first hypothesis for our study was that Internet-based procrastination would be related to a perception that the Internet is an important tool despite self-acknowledged procrastination through Internet use. This hypothesis was based on Postman's (1992) notion of the technological bias in which properties of immediate advantageousness and productivity were attributed to technology. It was expected that the misconception that Internet use automatically enhances efficiency would contribute to an environment in which procrastinatory activities were easier to justify and prolong.

The second hypothesis was that Internet-based procrastination would be related to a perception of the Internet as entertainment. This hypothesis was based on a concept Postman (1985) revealed in an earlier work, *Amusing Ourselves to Death*. Here, Postman argued that North American television facilitates the presentation of discourse as entertainment. Information becomes trivialized as it is articulated as short and continuously dynamic "snippits." According to Postman, televised information is presented in a fragmented format in which decontextualized facts are delivered at speeds permitting only a moment's thought. Postman suggested that data delivered in this irrelevant manner transforms knowledge into entertainment.

We also reasoned that Internet procrastination would be related to perceiving the Internet as entertaining in that research has demonstrated that procrastinators will engage in activities they find pleasant while actively postponing the completion of a task (Ellis & Knaus, 1977; Knaus, 1973, 1998, 2000; McCown & Johnson, 1991, as cited in Ferrari, Johnson, & McCown, 1995). Research by McCown (1986, as cited in Haycock, McCarthy, & Skay, 1998) suggests that "procrastination is a learned habit developing from a human preference for pleasurable activities and short-term rewards" (p. 31). More recent research by Pychyl, Lee, Thibodeau, and Blunt (2000) underscores the role of pleasurable, short-term, specious rewards as an important aspect of procrastination. Thus, the Internet may be particularly conducive to procrastination due to the provision of entertainment and pleasure through distraction in online activities.

The third hypothesis was that Internet procrastination would be associated with stress relief. Research by Ferrari et al. (1995) indicates that procrastination in task avoidance serves to relieve experiences of anxiety. Anxiety may be a general response to the task at hand due to perceived task averseness or fear of failure. McCown and Johnson (1991, as cited in Ferrari et al., 1995) advocated that to the extent that this anxiety is adverse, procrastinators will engage in some other less stressful task.

The tendency to seek distraction when experiencing anxiety associated with an aversive task is supported by clinical work (e.g., Knaus, 1973, 1998, 2000), which indicates that peo-

ple often avoid aversive tasks by resorting to "escapist" activities such as watching television or sleeping to obtain temporary relief from anxiety. In the same manner, procrastination through Internet use was expected to be related to temporarily diminishing stress through entertaining distractions.

The final hypothesis for the study was that procrastination through Internet use would be related to feelings of negative affect. This hypothesis was based on a study by Kraut et al. (1998), who found that Internet use was associated with negative subjective well-being. Kraut et al. demonstrated that the greater the time spent online, the greater are the feelings of depression and loneliness in first time Internet users. Negative well-being was shown to have increased with the introduction of Internet access to the participants.

Moreover, general procrastination has been linked to negative affect (Burka & Yuen, 1990; Ellis & Knaus, 1977; Ferrari et al., 1995; Pychyl et al., 2000). Burka and Yuen (1990) stated that those who needlessly delay tasks suffer negative internal consequences, feel miserable, and are frustrated. They suggested that the cyclical nature of procrastination builds pressure and causes enormous strain and self-doubt within the individual as the task is continuously postponed. While those engaging in procrastination attempt to distract themselves with pleasurable activities, any enjoyment ultimately subsides and is replaced by regret, apprehension, and guilt (Burka & Yuen, 1990).

METHOD

Participants

Respondents were 314 individuals who completed a survey on the World Wide Web and submitted the questionnaire results via e-mail. All questionnaires with more than three questions left incomplete were not used in the survey. In total, 308 surveys were included in this study. There were 198 females (mean age = 29.7, *SD* = 12.0) and 104 males (mean age = 28.5, *SD* = 11.7) in the sample (6 respondents failed to indicate gender). All participants resided in North America, where 56.7% lived in the United States and 43.3% resided in Canada. In reporting occupations, 51.9% of the sample indicated they were students (male $n = 55$, mean age = 20.02, *SD* = 2.72; female $n = 97$, mean age = 20.90, *SD* = 5.01), 42.5% were employed (male $n = 42$, mean age = 39.10, *SD* = 9.82; female $n = 89$, mean age = 38.87, *SD* = 10.27), and 3.6% were unemployed.

Design and Procedure

Participation in this study was voluntary and occurred through completion of the survey accessed through the Procrastination Research Group Web site, posted at `http://www.carleton.ca/~tpychyl`. Participants visited the Web site for various reasons, such as general curiosity, self-help, and research purposes. Several participants indicated that participation in the questionnaire reflected a fulfillment of a university course requirement or resulted from encouragement from university instructors at the University of Alberta. One method of accessing the Web site was through entering the word *procrastination* into a search engine and subsequently selecting the Procrastination Research Group Web site link among the available search hits. Search engines such as Altavista, Lycos, and Excite listed the Procrastination Research Group site first when the word *procrastination* was entered into the search field.

To complete the questionnaire, participants accessed the Procrastination Research Group Web site and selected the participate in research option link from the home page. Immediately following this selection, a Web page displaying an informed consent form appeared. The informed consent form described the voluntary method of participation, assured strict confidentiality, and offered the freedom to withdraw from participation at any time. Participation requirements and the purpose of the research were also displayed on the informed consent page. In addition, the informed consent form offered guidance in obtaining additional information concerning this research.

Participants who accepted the terms of the informed consent form then selected the questionnaire link presented at the bottom of the consent form page. Upon completion of the survey, participants were provided with the option to send the questionnaire, which resulted in the submission of the questionnaire via e-mail. Debriefing information outlining the purpose of the research and the availability of results was accessed through selection of the hypertext debriefing information link displayed at the end of the final section of the questionnaire. The debriefing page also included a list of references for further reading on procrastination.

It is important to note that participants were not actively recruited and were limited to those individuals with access to the Internet. As this was a nonrandom sample, this limited the interpretation of any statistical tests performed on the data, given that tests of significance cannot be interpreted in conventional terms of probabilities or replication of strength of relationship under resampling. We have conducted and presented statistical tests using a standard criterion for statistical significance (i.e., $\propto < .05$) for their value as a useful, if arbitrary, cut-off criterion in this exploratory research.

Measures

Participants completed a self-report research questionnaire composed of a total of 48 questions divided into five sections: demographic information, personal feelings and attitudes toward the Internet, Internet time-use evaluation, Lay's (1986) General Procrastination (GP) Scale, and Diener and Emmons (1985) Composite Affect Scale.

DEMOGRAPHIC INFORMATION

This section consisted of six questions including gender, age, occupation, citizenship, country of residence, and state or province if residing in the United States or Canada, respectively.

PERSONAL FEELINGS AND ATTITUDES TOWARD THE INTERNET

This segment of the survey contained six questions measuring overall feelings and attitudes concerning the Internet. Items were scored using a 5-point Likert-type scale (1 = *false of me*; 5 = *true of me*). Examples included "I feel the Internet makes me more productive" and "I often use the Internet to relieve stress."

INTERNET TIME-USE EVALUATION

This section consisted of five open-ended questions requiring the participants to provide an estimate of the number of daily hours spent using the Internet in three different locations (i.e., home, work, school). The number of hours spent using the Internet at home, work, and school was summed to obtain an estimate of total time spent using and procrastinating on the Internet. Estimates regarding the amount of time spent procrastinating on the Internet per

day in these same locations were also requested. Participants indicated in their own words the reasons underlying their personal Internet use. Participants provided information regarding how they think about time spent on the Internet. Finally, respondents indicated whether they were procrastinating at the moment in a "yes" or "no" format.

GP SCALE

The GP Scale (Lay, 1986) is a self-report measure of trait procrastination. This unidimensional inventory consists of 20 items describing general daily tasks, such as "I often find myself performing tasks that I had intended to do days before." Each item is scored on a 5-point Likert-type scale (1 = *false of me*; 5 = *true of me*). Ten items are reversed scored and responses are summed to obtain a single composite score, with high scores reflecting procrasinatory behavior. The GP Scale has been demonstrated to have a test-retest reliability of 0.80 (Ferrari, 1989) and a Cronbach's alpha coefficient of 0.82. (Lay, 1986). According to Ferrari (1992), the GP Scale is effective in measuring habitual task delay across several situations. The scale has been positively correlated with low self-esteem, defensive avoidance, disorganization, self handicapping, and individual need for achievement (Ferrari, 1992; Ferrari et al., 1995; Lay, 1986). This scale was included in the study to determine an association between Internet procrastination and general procrastination.

DIENER AND EMMONS' COMPOSITE AFFECT SCALE

The Diener and Emmons' (1985) Composite Affect Scale consists of nine monopolar adjectives describing mood. Five adjectives represent negative affect (i.e., depressed, worried and/or anxious, angry and/or hostile, unhappy, frustrated) and four adjectives denote positive affect (i.e., joyful, happy, pleased, enjoyment and/or fun). Participants rated the degree to which each adjective described their feelings toward using the Internet on a 7-point Likert-type scale (0 = *not at all*; 7 = *extremely much*). Scores were obtained by summing positive-affect ratings and negative-affect ratings to attain a composite score of positive and negative affect for each participant.

Two additional items were added to supplement the scale: motivated and guilty. Motivation, considered a positive adjective of affect, was included in the scale to reflect the negative relation between motivation and procrastination (Pychyl et al, 2000). Similarly, a demonstrated positive relation between procrastination and guilt led to the addition of the guilty adjective to the negative affect scale (Pychyl et al., 2000; Pychyl & Little, 1998). Both the temporal stability and internal consistency coefficients approached 0.90 in previous research (Emmons & Diener, 1985, 1986), and the scale has a demonstrated reliability across several situations (Omodei & Wearing, 1990; Pychyl & Little, 1998).

RESULTS

A series of Pearson's product-moment correlation coefficients, paired samples *t* tests, and analyses of variance (ANOVAs) were conducted to explore the hypothesized relations among the variables and the differences between groups. All statistical tests conducted were two-tailed. Due to the multiple *t* tests conducted in this study, an alpha level of .001 was set for all *t* tests to compensate for the increased risk of a Type I error (Evans, 1998).

ANOVA's and *t* tests revealed no significant differences in gender in any of the variables. Likewise, there were few significant differences between students and employed respondents (differences noted are discussed later in this section). Given the similarity of the groups, the sample was pooled and treated as a single group for the remainder of the analyses.

Internet Time-Use Evaluation

The mean time per day spent using the Internet and the average amount of time spent per day procrastinating through Internet use was calculated by adding the respondent's estimates of time spent using the Internet and amount of time spent procrastinating online in each of the three domains (home, work, school) and dividing by the total number of participants. The total mean time spent using the Internet in all domains was 3.36 hours per day (*SD* = 3.4). The daily modal time spent using the Internet was 1.5 hours. The mean time spent procrastinating while using the Internet was 1.59 hours per day (*SD* = 2.38). As shown in Table 1, respondents spent just less than half (47%) of the online time procrastinating. It is interesting to note the wide variability around these means, indicating that many participants spent considerably more (and less) time online than the reported average.

GP Scale

To determine the nature of Internet procrastination, GP Scale scores were calculated and correlated with reports of online procrastination. Participants were found to procrastinate above the midpoint of the scale. The mean GP Scale total score was 64.6 (*SD* = 14.6), ranging from a minimum of 25.0 to a maximum of 94.0. In comparison with previous research, it was found that the GP Scale mean score in this study was slightly higher than some of the past research (e.g., *M* = 58.7, *SD* = 11.9 for males; *M* = 53.9, *SD* = 12.9 for females; Blunt & Pychyl, 1998) and relatively consistent with other research (e.g., *M* = 64.81, *SD* = 14.91; Zamanpour & Pychyl, 2000), indicating that this sample was generally normal. Excessive trait procrastination was not evidenced in this sample despite the voluntary nature of participation through the Procrastination Research Group Web site.

An ANOVA revealed that a significant difference in GP scores existed between respondents with different occupations, $F(2, 296) = 5.871$, $MSE = 1181.82$, $p = .003$. Participants who were employed ($M = 66.89$, $SD = 14.63$) demonstrated a significantly higher GP Scale score than did participants who were students (M = 61.88, *SD* 14.14).

Furthermore, a series of Pearson's product-moment correlation coefficients was calculated to examine the relation between procrastination on the Internet and general procrastination as measured by the GP Scale. Internet procrastination and GP Scale scores were moderately correlated, $r = .39$, $p < .0001$. GP Scale scores were also significantly positively correlated with mean negative affect, $r = .38$, $p < .0001$; spending more time using the Internet than expected, $r = .32$, $p < .0001$; and stress relief, $r = .28$, $p < .0001$.

Feelings and Attitudes Toward the Internet

An analysis of the frequency and percentage of responses of the total sample regarding personal feelings and attitudes toward the Internet indicates that the Internet is a potent environment conducive to time wasting, as half of the sample (50.7%) reported engaging in frequent procrastination while using the Internet. Furthermore, 83.5% of the respondents spent more time than anticipated when online. As expected, most of the sample (69.5%) found the Internet to be very entertaining. In addition, in accordance with Postman's (1992) argument relative to the technological bias, the majority of participants (85.6%) perceived the Internet as an important tool.

A series of Pearson's product-moment correlation coefficients was calculated to explore the interrelations among the six Internet feelings and attitudes variables, with specific emphasis on Internet procrastination. As predicted, Internet procrastination and believing

TABLE 1
Mean Hours of Daily Time Spent Using the Internet and Procrastinating Online

	Mean Hours (N = 305)	SD	*Time Spent Pocrastinating (%)*
Home			
Hours spent using the Internet	1.96	2.15	57
Hours spent procrastinating	1.12	1.89	
Work			
Hours spent using the Internet	0.75	1.91	32
Hours spent procrastinating	0.24	0.66	
School			
Hours spent using the Internet	0.65	2.06	40
Hours spent procrastinating	0.26	0.95	

the Internet was an important tool were positively correlated, $r = .24$, $p < .0001$. This outcome lends support to Postman's (1992) argument regarding technological bias: Despite the participant's recognition of procrastination via the Internet, an associated attribution of importance toward the Internet remains.

Similarly, Postman's argument for wasting time in technological entertainment was supported by the finding that Internet procrastination was significantly related to a perception of the Internet as entertaining, $r = .35$, $p < .0001$. Like television, Internet procrastination is clearly related to the entertainment quality of the medium.

Internet procrastination and perceived online stress relief were examined to determine whether procrastination through computer technology was similar to procrastination related to the reduction of anxiety. As expected, Internet procrastination and stress relief were significantly correlated, $r = .57$, $p < .0001$. This outcome is consistent with previous research regarding the association of procrastination and stress relief (e.g., Ferrari et al., 1995). Furthermore, using the Internet to relieve stress is also positively and independently associated with online entertainment ($r = .46$). Notably, online stress relief differed significantly between student and employee respondents, $F(5, 281) = 3.311$, $MSE = .790$, $p < .005$. Participants who were employed ($M = 3.14$, $SD = 1.39$) reported a significantly higher amount of online stress relief than did students ($M = 2.61$, $SD = 1.30$).

Internet Procrastination and Negative Affect

Finally, negative affect toward the Internet was examined to determine whether online task postponement was related to feelings of guilt, unhappiness, anxiety, anger, depression, and frustration. The results indicated that positive affect adjectives (mean = 4.43, $SD = 1.14$) received significantly higher ratings than negative affect adjectives (mean = 2.77, $SD = 1.28$; $t = 14.43$, $p < .0001$), suggesting that most respondents find spending time on the Internet to be a more positive than negative experience.

An ANOVA revealed that a significant difference existed between respondents of differing occupations in describing feelings about Internet use, $F(2, 296) = 3.355$, $MSE = .5.363$, $p < .05$. Employed respondents ($M = 2.80$, $SD = 1.26$) reported a greater mean negative affect than students ($M = 2.67$, $SD = 1.25$). Furthermore, as predicted, mean negative affect and

Internet procrastination were positively correlated ($r = .27$, $p < .0001$). Internet procrastination was not significantly related to positive affect ($r = –.05$, $p = .36$). Mean positive affect was significantly related to perceiving the Internet as an important tool ($r = .37$, $p < .001$), finding the Internet entertaining ($r = .38$, $p < .001$), Internet productivity ($r = .38$, $p < .001$), and stress relief ($r = .19$, $p < .001$). Furthermore, mean negative affect was positively correlated with spending more time than anticipated ($r = .16$, $p < .001$), Internet procrastination ($r = .27$, $p < .001$), and the GP Scale score ($r = .38$, $p < .001$). More specifically, higher levels of procrastination were related to the feeling of guilt. The guilty adjective was significantly correlated with Internet procrastination ($r = .50$, $p < .0001$). This finding suggests that although participants frequently engaged in procrastination on the Internet, they were feeling guilty about doing it.

DISCUSSION

The purpose of this study was to explore the extent to which time spent online was related to self reports of procrastination and attitudes held regarding the Internet. Results verified that cyberslacking is prevalent among Internet users in that 50.7% of the sampled participants procrastinated through Internet use on a frequent basis and that 47% of the time spent online for all participants involved self reports of procrastination. Although a simple generalization of these results is problematic, as we discuss below, the trends are disturbing. Based on the recent estimate of 108 million Internet users in the United States (Braziller, 2000), as many as 54 million subscribers may use the Internet frequently for procrastination, and nearly half of the time that anyone spends online may involve procrastination in some form.

Reports of excessive and serious time wasting through Internet activities are not uncommon. Brenner (1998, as cited in Wallace, 1999) reported that of 185 respondents of an online survey, 30% reported an inability to reduce Internet use time and more than half of those surveyed had been told by others that they were spending a disproportionate amount of time online. Behavior of an addictive and compulsive nature resulting in intense attachment to Internet computer use has also been documented (Griffiths, 1998, 1999; Huang & Alessi, 1997; Stein, 1997; Swadley, 1995; Wallace, 1999). For example, Swadley described the "addiction" to multiple-user domains as a general phenomenon, citing a daily commitment of 16 hours as typical for some Internet users. This illustrates the attraction of interactive technology as well as the negative effects of online use. There remains little doubt that the Internet has become a tool for procrastination in that a new generation of mouse potatoes (Scruby, 1999) or cyberslackers (Marron, 2000) is emerging.

The potent seduction of the Internet as a medium for cyberslacking or procrastination can be explained through Postman's (1992) notion of the technological bias, the misconception that the use of technology intrinsically allows for greater human efficiency. The results of our study confirmed the hypothesized positive relation between Internet-based procrastination and the perception of the Internet as an important tool ($r = .24$). This paradoxical finding supports Postman's assertion regarding the negative effects of technological bias in which information technology represents an ultimate form of wisdom. Technological prejudice is evidenced through the participants' general perception that the Internet is an important tool (85.6%), despite the fact that those in the sample spent almost half of the time online procrastinating. In fact, only 24% of sampled Internet users felt Web use made them unproductive. Thus, technological bias was illustrated through the respondents' continued association of Internet use with industriousness, irrespective of the personal acknowledgement of frequent procrastination online.

Perceptions of the Internet as an effective instrument despite self-recognized procrastination may be easily justifiable to the user because the procrastination activity involves exposure to information. Believing that access to an enormous amount of information via the Internet intrinsically facilitates industriousness may hinder self-regulation against procrastination. It is possible that the respondents felt they were still extracting value from their online activities even while procrastinating, thereby facilitating the deceptive perception that work was being accomplished.

The establishment of the "mouse potato" generation can be further explained through the application of Silver and Sabini's (1981) theory of procrastination. It will be recalled that Silver and Sabini suggested that procrastination involves searching for off-task distractions while dramatizing an appearance of readiness to work. This creates a situation or "field" in which procrastination is facilitated. According to Silver and Sabini, individuals attempting to convince themselves that work is being continued are particularly susceptible to any activity characterized by convenience, minimal obligation, and lack of immediate unpleasantness. The Internet is particularly seductive in that it encompasses all of these properties. Thus, to the extent that the Internet is a virtual procrastination field, the online environment is the most modern medium available for task postponement.

Virtual Entertainment

Internet subscribers also appear to be amusing themselves into procrastination through the availability of virtual entertainment. Statistical analysis demonstrated that Internet-based procrastination was related to a perception of the Internet being very entertaining ($r = .36$). The majority of the sample (70%) reported their online experience as amusing. This finding is not unexpected given that the Internet has much to offer in the way of instant entertainment, including sites devoted exclusively to entertainment and time wasting. Many sites serve as control "portals," providing a list of links to the most interesting online locations including shopping, music, gaming, and jokes. A particularly salient example of this is the "I-should-be-working" Web site found at `http://www.ishouldbeworking.com`. The social and interactive aspects of chat rooms and multiple user domains also make the Internet an exciting instrument for users.

In accordance with Postman's (1985) prediction that television facilitates the presentation of information as entertainment, the form and content of information available through Internet technology correspondingly represents entertainment in that data are presented in decontextualized "snippits." Evidence supporting the idea that the Internet is evolving into the newest medium for procrastination is found in a PricewaterhouseCoopers poll in which 38% of Canadians surveyed indicated they would be watching television if the Internet did not exist (Lyman, 2000). This finding suggests that the Internet may be replacing the television as a form of entertainment and possibly as a form of procrastination.

Again, this association between Internet procrastination and entertainment may best be understood through Silver and Sabini's (1981) theory of procrastination, specifically their discussion of time fragmentation. Silver and Sabini postulated that one method to delay working toward task completion is to perceive time as segments of short intervals. Procrastination occurs when the individual justifies engaging in a minor amusement instead of committing to the task, even when the decision to work is not withdrawn. This justification is founded on the rational belief that the task can wait the few minutes while one engages in a short-term pleasure, during which time little long-term cost is experienced. This form of procrastination is easily applied to Internet use due to the fact that online activities typically do

not require more than a few minutes to complete (e.g., viewing e-mail, searching the Internet for the day's weather).

The Internet is an especially risky instrument for idleness given this type of time fragmentation because this technology encompasses the properties that place the user at risk for procrastination: speed, accessibility, and "tip-of-your-fingers" convenience. There exists a particular attractiveness inherent to digital indulgences given that the pleasurable distraction is brief and can be ceased, at least ostensibly, by an act of one's volition or will. Rationalizations such as "Just one more game" or "I'm just checking my e-mail" are easily justified when perceiving time in short intervals because engaging in quick, minor distractions does not seem harmful relative to completing work.

In sum, Internet users appear to be entertaining themselves into task postponement. Internet activities are particularly seductive for users given this type of time fragmentation, coupled with the availability of characteristics conducive to procrastination.

Surfing and Stress Relief

It appears that there is a positive side to online procrastination. As hypothesized, procrastination via Internet use is related to perceived feelings of stress relief ($r = .57$). This correlation supports the previous research demonstrating that stress relief is a prominent motivator for task avoidance. For example, Ferrari et al. (1995) indicated that procrastination served to relieve feelings of anxiety related to both task avoidance associated with completing boring or difficult tasks and provided temporary protection from confirmation of feared lack of abilities.

Due to the relation between stress relief and online procrastination, access to the Web places the anxious Internet subscriber into a tempting predicament. This is especially true for employee Internet users who, as demonstrated in this study, reported greater stress relief in online distraction than did students. A click of a mouse button provides ready task avoidance for Internet users who find a task aversive or frustrating. Initial attempts to complete assigned work on the computer (e.g., writing) can be thwarted by engaging in Internet distractions (e.g., Web surfing). This convenient remedy for anxiety through online procrastination is maladaptive in that the relief from stress is only temporary and may produce feelings of guilt or negative affect more generally.

Negative Affect and Cyberguilt

Generally, subjective well-being is at risk if the Internet is used as a method to escape tasks that require completion. A positive correlation of mean negative affect associated with Internet procrastination ($r = .27$) suggests that the time spent using the Internet to procrastinate can be characterized by frustration, guilt, depression, and unhappiness. Given that 47% of the time devoted to Internet use is spent procrastinating, it can therefore be estimated that almost half of the time spent online is accompanied by the experience of negative emotions. This finding supports Kraut et al.'s (1998) conclusions that Internet use is associated with negative subjective well-being as well as with feelings of depression. Insofar that procrastination has been empirically related to anxiety, depression, and low self-esteem (Ferrari et al., 1995), it is not surprising to find that Internet procrastination is characterized by the same negative emotional properties.

Among the negative emotions found to be related to procrastination, guilt accounted for 25% of the variance in Internet procrastination ($r = .50$). Based on the relation between

online procrastination and feelings of guilt, it is apparent that the procrastinator is not successful in deriving continuous stress relief through task postponement. It appears that employed Internet procrastinators may be at a greater risk for suffering negative affect from Internet use given that employed participants reported higher mean negative affect ratings than did students.

Limitations of the Study and Future Research

The use of the Internet for research is still in its infancy, and although it provides interesting opportunities, it presents methodological problems as well. The major issue regarding Internet research is sampling bias (Coomber, 1997). Studies conducted via the Internet are selective and can only reach people who have access to the Web and who use it. Another sampling difficulty associated with this Web-based study was the lack of active recruitment of participants. Participants in the study were required to discover the Web site independently and to make the choice to participate based on personal interest alone. It is possible that most people visiting the Procrastination Research Group Web page were procrastinating; thus, the sample may be less representative of Internet users. However, the average score on trait procrastination, as measured by the GP Scale (Lay, 1986), indicated that the participants in the study did not represent an abnormally high procrastination group. In fact, the average procrastination score closely resembled other research samples (e.g., Zamanpour & Pychyl, 2000).

Despite these limitations, Internet surveys have much potential. The generalizability of the findings for this study were not problematic because the population of interest was Internet users. Moreover, because many of the participants in the study were students who were asked to complete the survey by instructors at the University of Alberta, the frequency of procrastinators independently accessing the Web site was reduced. Furthermore, in terms of the self-report element of the survey, the anonymity associated with responding to questionnaires via the Internet may have facilitated honest answers and less pressure to respond in a socially favorable manner.

Given the nature of these limitations, future research would benefit from a more active approach to collecting data to reduce sampling bias. Conducting an observational study in which the Web sites of the participants are monitored electronically would remove the limitation of self-report data. If this experimental design could be employed through random sampling, the susceptibility of targeting procrastinators would be reduced relative to having participants discover a questionnaire on a Web page.

Another method may involve visiting newsgroups of the specific populations of interest to canvas for research participation (Coomber, 1997). Targeting the work environment is a viable research endeavor given the percentage of those sampled who engage in online procrastination and the serious consequences associated with idleness on the job.

Summary and Conclusions

The purpose of this study was to explore the extent to which time spent online was related to self reports of procrastination. Results verified that cyberslacking is pervasive among sampled participants. The attractiveness of the Internet as an environment conducive to task postponement was explained through Postman's (1992) notion of technological bias and Silver and Sabini's (1981) theories of procrastination. It is apparent that efforts to increase productivity through generating and distributing more information at faster rates through technology have not been entirely successful. The convenience, speed, and accessibility of the

Internet and accompanied bias of technological productivity have served to create a tool for procrastination. The implications of time wasted through Internet procrastination are significant for individuals and businesses alike, and the impact of online procrastination may only increase as more of our activities move into this new computer-mediated environment. Given the preliminary nature of this study, further research in the area of Internet procrastination is needed to provide greater insight into the phenomenon of cyberslacking.

REFERENCES

Blunt, A., & Pychyl, T. A. (1998). Volitional action and inaction in the lives of undergraduate students: State orientation, procrastination and proneness to boredom. *Personality and Individual Differences*, *24*, 837-846.

Braziller, C. (March 22, 2000). Face of the Web study pegs global Internet population at more than 300 million [Online]. Retrieved March 22, 2000, from the World Wide Web: `http://www.angusreid.com/media/content/displaypr.cfm?id_to_view=1001`

Burka, J. B., & Yuen, L. M. (1990, May/June). Procrastination: Causes and consequences. *New Realities*, 42-45.

Coomber, R. (1997). Using the Internet for survey research. *Sociological Research Online* [Online serial], *2*(2). Retrieved March 19, 2000, from the World Wide Web: `http://www.socresonline.org.uk/2/2/2.html`

Diener, E. & Emmons, R. A. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47*, 1105-1117.

Ellis, A., & Knaus, W. J. (1977). *Overcoming procrastination*. New York: Penguin.

Emmons, R. A., & Diener, E. (1985). Personality correlates of subjective well-being. *Personality and Social Psychology Bulletin*, *11*, 89-97.

Emmons, R. A., & Diener, E. (1986). A goal-affect analysis of everyday situational choices. *Journal of Research Personality*, *20*, 309-326.

Evans, A. N. (1998). *Using basic statistics in the social sciences* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Ferrari, J. R. (1989). Reliability of academic and dispositional measures of procrastination. *Psychology Reports*, *64*, 1057-1058.

Ferrari, J. R. (1992). Psychometric validation of two procrastination inventories for adults: Arousal and avoidance measures. *Journal of Psychopathy and Behavioral Assessment*, *14*, 97-100.

Ferrari, J. R., Johnson, J. L., & McCown, W. G. (1995). *Procrastination and task avoidance: Theory, research and treatment*. New York: Plenum.

Griffiths, M. (1998). Internet addiction: Does it really exist? In J. Gackenbach (Ed.), *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 61-75). San Diego, CA: Academic Press.

Griffiths, M. (1999). Internet addiction. *Psychologist*, *12*, 246-250.

Haycock, L. A., McCarthy, P., & Skay, C. L. (1998). Procrastination in college students: The role of self-efficacy and anxiety. *Journal of Counseling and Development*, *76*, 317-324.

Huang, M. P., & Alessi, N. E. (1997). Drs. Huang and Alessi reply. *American Journal of Psychiatry*, *153*, 890.

Knaus, W. J. (1973). Overcoming procrastination. *Rational Living*, *8*, 2-7.

Knaus, W. J. (1998). *Do it now: How to stop procrastinating* (Rev. ed.). New York: John Wiley.

Knaus, W. J. (2000). *Procrastination, blame and change* [Special issue]. *Journal of Social Behavior and Personality*, *15*, 153-166.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being. *American Psychologist*, *53*, 1017-1031.

Lay, C. H. (1986). At last, my research article on procrastination. *Journal of Research in Personality*, *20*, 474-495.

Lyman, P. (2000, January 10). *PricewaterhouseCoopers study shows Canadians lead Europeans in Internet usage* [Online]. Retrieved February 21, 2000, from the World Wide Web: `http://www.pwcglobal.com/extweb/ncpressrelease.nsf/DocID/521DCDD66F5F0EFB852568600070234E`

Marron, K. (2000, January 20). Attack of the cyberslackers. *The Globe and Mail*, p. T5.

Omedei, M. M., & Wearing, A. J. (1990). Need satisfaction and involvement in personal projects: Toward an integrative model of subjective well being. *Journal of Personality and Social Psychology*, *59*, 762-769.

Postman, N. (1985). *Amusing ourselves to death: Public discourse in the age of show business*. New York: Penguin.

Postman, N. (1992). *Technopoly: The surrender of culture to technology*. New York: Vintage.

Pychyl, T. A., Lee, J. M., Thibodeau, R., & Blunt, A. (2000). Five days of emotion: An experience sampling study of undergraduate student procrastination [Special issue]. *Journal of Social Behavior and Personality*, *15*, 239-254.

Pychyl, T. A., & Little, B. R. (1998). Dimensional specificity in the prediction of subjective well-being: Personal projects in pursuit of the Ph.D. *Social Indicators Research*, *45*, 423-473.

Scruby, D. (1999, September 30). *1999 PricewaterhouseCoopers four country consumer technology survey* [Online]. Available: `http://www.pwcglobal.com/extweb/ncpressrelease.nsf/DocID/extweb/ncpressrelease.nsf/DocID/E08266B81BE91F9A852568010063B693?OpenDocument`

Silver, M., & Sabini, J. (1981). Procrastinating. *Journal of Theory of Social Behavior*, *11*, 201-221.

Stein, D. J. (1997). Internet addiction, Internet psychotherapy. *American Journal of Psychiatry*, *153*, 890.

Swadley, R. K. (Ed.). (1995). *The Internet unleashed* (2nd ed.). Indianapolis, IN: Sams-net.

Wallace, P. M. (1999). *The psychology of the Internet*. Cambridge, UK: Cambridge University Press.

Zamanpour, H., & Pychyl, T. A. (2000, June 30). *The cognitive-affective antecedents of avoidance achievement motivation and academic procrastination*. Poster presented at the annual convention of the Canadian Psychological Association, Ottawa, Ontario, Canada.

Jennifer A. A. Lavoie is a graduate student in the Department of Psychology at Simon Fraser University, British Columbia. She completed this research as a thesis project at Carleton University, Ottawa.

Timothy A. Pychyl is an associate professor of psychology at Carleton University, with a research focus on personal well-being and the breakdown in action known as procrastination.

Going Beyond the Code

The Production of Hypermedia Ethnography

BRUCE MASON
Memorial University of Newfoundland

BELLA DICKS
Cardiff University

New developments in multimedia and hypermedia technology have raised important questions for the conduct and dissemination of sociological research. Attention has traditionally focused on uses of new media for teaching, but their application to research is now also becoming clearer. There has been particular interest—although little practical work so far—on how ethnographers might use hypermedia in planning, design, analysis, and presentation of ethnographic work. This article offers findings from the authors' current ESRC project, which examines the appropriateness of hypermedia for the production of ethnography. The theoretical and methodological implications for sociological knowledge of this new technology are discussed. The authors also address arguments about the potential for transforming the production of sociological knowledge through granting equal weight to the audio-visual plane of meaning as that given to the verbal and situate their discussion within debates over new forms of reading and authoring offered to sociology through electronic media.

Keywords: hypermedia, hypertext, ethnography, representation, coding

This article derives from a simple belief: There must be something more interesting for social scientists to do with their computers than coding data (Coffey, Holbrook, & Atkinson, 1996; Dicks & Mason 1998; Weaver & Atkinson 1994). This proposition has lead the authors to engage in a 2-year project, funded by the Economic and Social Science Research Council, in which we explored the uses of new media in all phases of ethnography, from fieldwork to (re)presentation. The research was intended to exploit new media technologies so as to allow a new integration of visual and "textual" ethnography that bridges the current boundaries between the two. Consequently, our interest is not limited to ethnography as a product (whether textual, visual, or other) but as a holistic process. Our intent here is not to provide a how-to guide or a theoretical analysis but to comment on the various issues and challenges arising from our experience of authoring an ethnography of a local industrial heritage museum in hypermedia form on CD-ROM.

The aim of our work is to use technology in bridging the gap between "experimental ethnography" (Richardson, 1994; Van Maanen, 1995) and the seemingly more prosaic need to handle and analyze fieldwork data. We believe that there is a convergence of theory and new computer-mediated communication technology (cf. Landow 1997) that offers ways of

AUTHORS' NOTE: A version of this article was presented at the British Sociological Association, April 1999. We wish to thank the ESRC and Cardiff School of Social Sciences for financial support. We would also like to acknowledge the input of Paul Atkinson and Amada Coffey. The flaws, however, remain our own.

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 445-457

addressing the oft-mentioned crisis of representation in ethnography (Atkinson & Coffey, 1995; Clifford, 1988; Marcus & Fischer, 1986). New computer-mediated communications technology makes it easier to represent the messy complexities of the postmodern era. In a sense, the technology we explore serves a prosthetic (cybernetic) function, enabling the researcher to capture and preserve the interaction with the "field" more adequately than currently possible and to aid with the analysis and representation of the material. However, such technologies should not simply be naturalized. A camcorder is not just a tool for enabling us to remember better; a computer is not just a labor-saving device for the analysis of data. As researchers, we interact differently with the field depending on the tools we employ, and the field reacts differently to us. Walter J. Ong (1992) claimed that "writing is a technology that structures thought," and digital tools, when used in ethnography, are as meaningful and as powerful as the acts of writing, participating, and observing. While avoiding naïve forms of technological determinism, we will make various observations about the use of digital technology in all phases of the ethnographic process: from inception to representation.

USING HYPERMEDIA IN THE FIELD

A hypermedia ethnography requires reconceptualizing the nature of social research from its outset. Our research featured, among other more traditional interviewing and participant-observation techniques, "fly-on-the-wall" style filming of the guides at a local heritage park as they conducted their visitor tours. Simultaneously, we handed visitor groups a digital camera so that they could record their own images and provide us with an alternative perspective. Many studies on the use of technology in the field focus on the practicalities of employing the equipment (e.g., Agar, 1980; Hammersley & Atkinson, 1983; Howard, 1988; Jackson, 1987; Fischer, 1994); it is our contention that the employment of new technologies informs every aspect of the project, opening up new possibilities and modes of interaction and potentially closing others.

Hypermedia technology offers researchers the potential to exploit a range of media in ethnographic projects and to appreciate how each medium—whether printed word, sound, video, still image, or audio—brings its own characteristic meaning-making capabilities (cf. McLuhan, 1964). Rather than seeing the visual and aural dimensions as merely illustrations or add-on extras that complement the printed word, hypermedia allows these media to act as equal (although not equivalent) sites for the production of meaning (Kress, 1998; Marcus, 1994). However, there are certain issues that arise in the ethnographer's encounter with video in a hypermedia environment. Some of these are familiar to visual ethnographers, who have been working with film and video for decades. Others are posed more particularly by the hypermedia challenge itself and arise through the particular demands of combining different media into one environment.

The first dilemma arises as soon as filming begins. Whereas the traditional fieldworker will take copious field notes, using the full range of human senses and supplementing this with technological back up where needed (e.g., in the form of tape recorders), video ethnographers will spend considerable periods merely filming. Where they are the sole fieldworkers, the demands of operating the camera can all too easily obscure the social activity happening all around and narrow the field of vision to merely the action that is "filmable" (Barbash & Taylor, 1997). The video ethnographer is always confronted by a tension between the demands of visual recording (for visual impact and visual significance as well as more mundane technical requirements such as good lighting, the positioning of subjects, adequate sound, and so forth) and the challenges posed by the need for ethnographic rigor (Biella, 1993a, 1993b; Weinberger, 1994). This tension between the aesthetic and the record-

ing functions of film and video technology is a key issue for visual research. The camera is never a mere extension of the ethnographer's eye, as the visual plane of meaning works through sense-making conventions that are bound up with both a cultural aesthetics and a politics of looking (Hastrup, 1992; Morphy & Banks, 1997).

There is another issue to touch on briefly: the nature of team work in hypermedia-based social research. The traditional model of research features a lone researcher who makes a sortie into the field to collect data, brings back field notes to analyze, and then writes a monograph. A hypermedia-based project is probably too complex for one person: It requires the ability to use a camera, proficiency with a wide range of audio-visual equipment, and facility with a wide range of software. In many ways, these are the problems that visual ethnographers have faced for many years (Asch, 1988; Chiozzi, 1989). We are working on a team-of-generalists model in which the authors have overlapping skills as both ethnographers and cultural analysts and different areas of expertise: one in hypertext theory and one in visual ethnography. A team-of-generalists model allows us to try to avoid setting up the boundaries that may result from the ethnographer–camera operator split that often characterizes visual ethnography (Barbash & Taylor, 1997) while simultaneously recognizing that each researcher needs to bring something unique to the project.

USING HYPERMEDIA IN THE OFFICE

Following Weaver and Atkinson (1994), we propose that hypermedia-authoring programs have the potential to enable a richer and more creative interaction with data than the familiar code-and-retrieve paradigm seen in standard Computer Assisted Qualitative Data Analysis Software (CAQDAS) programs. There has been much debate about the implications of CAQDAS, its failings, and its abilities (Coffey et al., 1996; Dicks & Mason, 1998; Lee & Fielding, 1996; Weaver & Atkinson, 1994). Rather than repeating such debates here, we will instead describe some of the methods we have used in our project.

Hypertext and Analysis

The intent of the hypertext strategy is to take advantage of the associative nature of hypertext linking. As originally envisioned by Vannevar Bush (1945), hypertext was a form of informational retrieval. Bush was concerned that the amount of data being generated was outstripping researchers' accessing abilities. He reasoned that as human cognition tended to work in an associative manner, then a method of retrieval and indexing that featured associative links would be easier to use. In social research, hypertext enables the in situ linking of data segment to data segment as well as to interpretative texts authored by the researcher, whereas the traditional code-and-retrieve paradigm links data to hierarchical codes that run the risk of decontextualization from the discourse as a whole (Weaver & Atkinson, 1994). Data are no longer considered to be a series of quotable excerpts but are coherent networks of associative links. Hypertext thus potentially enables a more holistic interaction with the data.

We used two related strategies in pursuing hypertextual data analysis: reference nodes (for paradigmatic categorization) and path creation (for syntagmatic linking of data elements and analytic texts into interpretative pathways). The data were imported into *StorySpace*, a hypertext authoring program originally developed for the construction of hypertext fiction. The program allows for rapid import of plain text files and has fairly sophisticated linking abilities. Each data element, whether textual report, still photograph, or transcribed interview, was then attached to an associated reference node and links were created to the data. Each link was given a name based on what it referenced. For example, whenever the subject

of memorialism is mentioned in the data, a link named *memorialism* is created from the relevant part of the data to the reference node for that piece of the data. In a manner not dissimilar to coding, the researcher can progressively reread and/or revisit the data and refine the various links. *StorySpace* can then depict these links and the data's relation to the reference nodes with various graphical views. Figure 1 below shows a screen capture of the reference node and the various data links for an interview with a participant named Ivor.

The lower window is a part of the interview showing the end point of a link. The left-hand side of each window shows which nodes belong on the path named **memorialism*. A similar process can be used with still images, with portions (or all) of the image being capable of being linked to other segments of data or reference nodes. The version of *StorySpace* we used dealt only with text and still images[1], and we had to incorporate our video data at a later stage using the hypermedia-authoring program *Authorware* (from Macromedia). This software allows video and sound data to be synchronized on screen to develop visually arresting presentations while still allowing multidirectional hypertext navigation. However, the incorporation of video data confronts the researcher with a new set of problems, as we shall now see.

Video and Analysis

The question of how to analyze video data to construct scholarly argumentation is a particularly difficult and underdeveloped area (Biella, 1993a, 1993b). Visual ethnography does not traditionally include a period of analysis that precedes the postproduction or editing stage (Barbash & Taylor, 1997). As with field notes, however, the ethnographer's videotapes will need to be sifted through in a process of selection (by digital video capture on computer), which identifies particular sequences and shots that represent interpretative categories. This is a process that represents the video equivalent of immersing oneself in the field notes and categorizing the data. With video film, the researcher has a record of many different aspects of the field setting. The video camera records an enormous amount of data, which cannot be matched by even the most rigorous field notes. Unlike interview transcripts, these will include nonverbal communication such as facial expression and gesture as well as spatial features such as topography, movement, and so forth. The selecting out of film data for capture has to balance three aspects: the desire to exploit the full range of video data available, the equal need to ensure "viewability" and aesthetic and/or scholarly sense making, and the humdrum fact that computer storage capacity for video is still extremely limited.

In theory, decisions about which sections of video data to capture could correspond to the decisions made when coding an interview transcript. However, the point of coding a printed transcript is to categorize the data into related chunks that can then be analyzed together (through decontextualization and recontextualization or cut and paste). This is the process with which CAQDAS programs assist so effectively. In visual ethnography, however, it is simply not feasible to manipulate digital video data in this way, at least given current limitations on computer memory. Because video data have to be written to computer disc before they can be cut, pasted, or otherwise manipulated, researchers cannot afford the space to capture a fraction of the data that their coding frame would require. Instead, analysis of the video data will need to be done before any of the data are captured to computer disk. As software development for electronically tagging and labeling video data on computer screen is still in its infancy, researchers will probably have to rely on traditional techniques of video logging, with a pen and pencil at hand.

Furthermore, the question of whether traditional coding is at all appropriate for video analysis will first need to be confronted. In our own project, we had video data of two kinds: video of sit-down interviews and video of participants engaged in natural action. Although

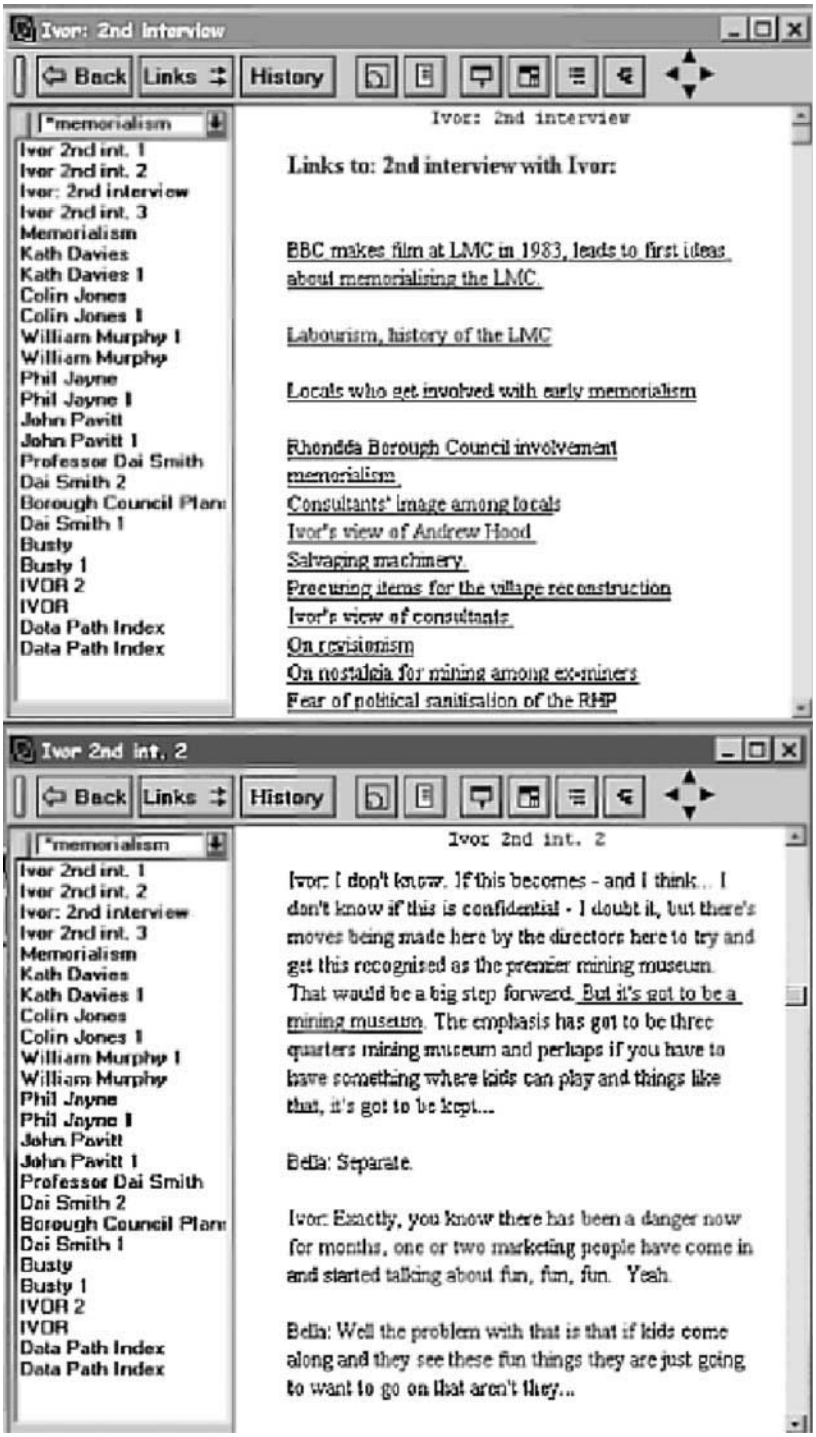

Figure 1: *StorySpace* Screen View Showing Data Links Created for a Sample Interview Transcript

the former can be treated in a similar way to traditional interview transcripts, with the added level of meaning that nonverbal communication conveys, natural action cannot easily be reduced to a finite number of categories. Instead, researchers will need to adopt a more flexible and interpretative approach to the data, which does not try to break it down into categories based on the same procedures as used for textual analysis.

USING HYPERMEDIA FOR REPRESENTATION

The intent behind our project is to explore the creative as well as analytic potential of digital technology. Experimental ethnography has attempted to broaden the representational repertoire of ethnographies, and hypermedia expands those possibilities. Similarly, recent ethnographies have tried to represent the complexities of the social world through a more fragmentary and reflexive literary style (Borchard, 1998; Dorst, 1989; Marcus, 1994b). Thus, there are already moves that challenge "traditional" linear print-based monographs. Using hypermedia, we have tried to balance the innovative nature of the medium with the requirements of rigorous academic writing. In ethnographic writing, we suggest, these will inevitably include the need for linearity and sequentiality, at least in places, to furnish the credibility of academic argumentation (Dicks & Mason, 1999; see also Biella, 1993a, 1993b). In what follows, we explore different means of organizing narratives and sequences in ethnographic hypermedia, first, by considering its written elements (hypertext) and second, by exploring what happens when we add in video and sound (hypermedia).

Hypertext

At its simplest, a hypertext consists of two or more textual entities (nodes), with some method that allows the reader to progress from entity to entity (links). Links between nodes may be basic in that they merely close one node and open another at its beginning, or they may be anchored in that either the source or the target (or both) of the link may be a subpart of the node (often known as *text links*). For example, selecting a link that is a quote from an informant might take the reader directly to the section of the interview in which the quote is uttered. Consequently, any node may have multiple links to and from other nodes. It is this interaction between multiple linking and nodes that creates the greatest potential, and greatest challenge, for hypertext authors.

Hypertext writing opens up many varied potentials for the representation of scholastic argumentation. As with any medium of expression, hypertext is amenable to specific rhetorical strategies and literary devices. To date, the major investigations into experimental hypertext writing have been carried out in the field of fiction, most notably, the authors Stuart Moulthrop (1991) and Michael Joyce (1996), although explorations of the use of hypertext for academic writing are beginning to appear, for example, David Kolb's (1994) philosophical hypertext *Socrates in the Labyrinth*, Diane Greco's (1995) *Cyborg,* and Wendy Morgan's (1999) postfeminist hypertextual reworking of *Monstrous Angels*. We will briefly explore some of the issues addressed by these authors and the strategies they have employed.

The first dilemma in hypertext authoring tends to be the tension between freedom and control. Hypertext potentially opens up the text through multiple linking, allowing readers the opportunity to generate unpredictable reading paths. Given this, how do authors, especially those dealing with scholastic argumentation, simultaneously orient readers toward intended readings while allowing readers to discover their own pathways through the hypertext?

Second, how do authors ensure that the multiple choices are not so bewildering that readers find the hypertext impenetrable? For fiction writers, this tension causes other difficulties. For example, consider a character, Lucretia, who has just died (i.e., the reader has just read the node containing Lucretia's death). How can the hypertext be constructed so that, having encountered her death, the reader does not then stumble on a node in which she is still alive? In a complex hypertext (e.g., Joyce's [1996] *Afternoon: A Story* contains 539 nodes and 951 links, Deena Larsen's [1996] *Samplers* contains 238 nodes and 2,038 links), ensuring that Lucretia stays dead is a major task. Whether we write as ethnographers with a scholarly interpretation to present or as fiction authors with a story to reveal, we are faced with the same issue: How do we create narrative progression in hypertexts? Creating navigational strategies has been the major concern of hypertext authors, and we can see emerging two major foci of attention: linking strategies and hyperstructures.

LINK TYPES

As noted above, links can be subdivided into two types: basic and text links. Any adequate hypertext-authoring system is capable of making such a differentiation. Furthermore, links can be named, and consistent, principled use of link names can help construct a semantic map of the hypertext. Finally, links can be conditional or dynamic rather than static. A conditional link is one that has different targets depending on whether certain conditions have been met; for example, some links could have different targets depending on whether the node containing Lucretia's death has been read. Dynamic links allow the use of various other kinds of limitations for restricting navigational directions.

Basic links can be used to create a default option at each node. For example, Joyce's (1996) *Afternoon* starts with a node that muses about poetry and recollection to an unnamed listener. Subsequently, 19 "hidden" links lead to various parts of the hypertext, and following any of them can disorient an incautious reader. The 20th link, however, is a basic link triggered by taking the default action of pressing the enter key; if the reader takes this option, then he or she may progress through 32 nodes taking the default option each time. By so doing, the reader gains a base-level insight into the characters and plot of the novel. Once the end of this default path has been reached, the reader may then back track and investigate various text links to experience the story in more depth.

A more complex use of links comes from using link naming to create paths. A path is a collection of nodes that are referenced by links of a certain name. Each node is then conceptually part of a theme or subject, and the reader may choose to follow this path in whole or part to explore its subject. For example, one of the paths in our hypertext, which we call the *identity* path, concerns the transformation of an old colliery into a living heritage museum: Every node that is connected to another node by a link named *identity* is part of this path. Several different paths exist in our work, which we call *tours*, and the reader's first interaction with the hypertext is in an introductory node called the *Tour Hall* in which the reader can choose which tour to explore. Paths thus provide local structure within a hypertext and are used extensively by Kolb (1994) in his philosophical hypertext to structure the argumentation.

Links can also serve a rhetorical function. For example, it is possible to create links that support, refute, exemplify, or otherwise comment on propositions contained in nodes. By using the appropriate link name, for example, *supports*, the reader can be made aware of what relationship the node at the other end of the link bears to the current node (see Barbules, 1998). There are, of course, many different possible link functions. Links can be used disjunctively, for instance, taking readers to other nodes without obvious relationships as a challenge to narrative closure. Links can also function lyrically, aesthetically, and narratively.

It should be noted, however, that link technology is still in its infancy. The most commonly used form of hypertext, HTML on the World Wide Web, can handle basic and text links but has no simple way of managing conditional or dynamic linking. Similarly, HTML is currently incapable of creating names for links. The most widely used authoring package for hypertext, *StorySpace*, features most of the features discussed above, but hypertext authors are always looking for new forms of link capabilities, such as the ability to preview the destination of a link before selecting it. Also, as will be discussed later, hypertext linking is somewhat challenged when authors attempt to include nontextual media in the piece.

HYPERTEXT STRUCTURES

Hypertext authors are beginning to explore the different possibilities of what Ted Nelson (1987)—a hypermedia pioneer—called *nonsequential* writing, but which we suggest thinking of as multisequential writing. In particular, complex linking strategies enable new narrative structures. Kolb's (1994) *Socrates in the Labyrinth* details various structures and experiments with them in four subsidiary hypertexts. Perhaps the most convincing structure he uses, and the one that has become most widely adopted, is the cycle. Here, after following a path, the reader ends up back at the beginning—but when the cycle is begun again, the path mutates: By inserting conditions into base links, a different set of nodes is encountered each time around. Kolb expanded this in "Earth Orbit," which has cycles of argumentation that spawn other cycles, forming what appears to be a series of orbits around a central theme. Figure 2 shows how Kolb demonstrated this.

Kolb (1994) also explored the possibility of a pyramidal structure of hypertext in his hypertext "The Habermas Pyramid." The top level consists of a series of postulates, each one of which is a link. Selecting any of the links takes the reader down a step of the pyramid into a more detailed set of postulates, each of which is also clickable, allowing the user to move down a further level until the bottom of the pyramid is revealed. For example, Figure 3 shows the top level of the pyramid (Part 1) and the window (Outline 1) that opens up when the first postulate "For Habermas, modernity brings differentiation and self-consciousness" is clicked.

An alternative approach is taken by Moulthrop (1991) in his hyperfiction text *Victory Garden*. Written in response to Joyce's (1987) heavily controlled *Afternoon*, Moulthrop's hypertext has a map as its main point of departure. By selecting various parts of the map, readers can explore the garden, meeting a variety of characters and encountering a series of events as they gradually makes sense of the narrative.

Each of these strategies has been described to give an idea of some of the possibilities open to authors in hypertext; it is by no means an exhaustive survey. For our work, we combined several of these strategies. Our overarching metaphor is that of ethnographer as guide, treating the reader as a visitor to the hypertext. The heart of the hypertext consists of several tours focused on specific issues. Each tour features a basic link from node to node so that a reader can follow the tour in a simple way. Text nodes can take the reader into less-structured excursions on optional parts of the tour. Through the consistent use of a hypertext structure, enabling readers to understand where they are in relation to the whole, as well as consistent navigation strategies to aid in understanding where their choices will lead them, we hope to minimize reader disorientation.

The use of such strategies allows authors to attempt to more fully portray the rich complexity of social phenomena than is possible in a printed linear text. Atkinson (1990) has noted the poetics of ethnographic writing and its role in conveying narrative and authenticating statements; hypertext offers new poetics for the ethnographer to explore (Dicks &

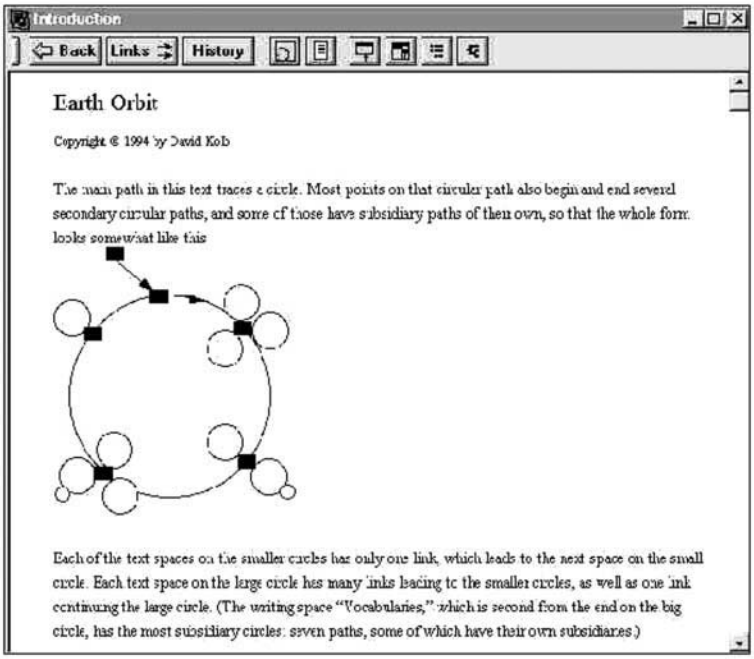

Figure 2: Introductory Node From Kolb's (1994) Hypertext, *Socrates in the Labyrinth*

Mason, 1999). The promise of hypertext also lies in its ability to open up scholarly texts. In social research, this can be realized by making much, if not all, of the data accessible to readers. When using the hypertext computer environment to help analyze the data, developers create various links and structures that may not appear in the final, authored hypertext. It is a relatively simple task to make these available to readers so that they can trace the creation of ideas and interpretations developed by ethnographers. In addition, readers can, in most hypertext reading programs, create their own links and add in their own notes and comments. Thus, readers can explore the narratives constructed by the authors, browse through the hypertext guided by their own interests, or even move into the "backstage" area of the data and create their own interpretations. Consequently, we can say that hypertext not only aids researchers' interaction with the data but also provides for a richer interaction between readers, researchers, and data.

Hypermedia

Hypermedia integrates multimedia components such as video and sound into hypertextual structures. In terms of presentation, hypermedia allows a certain amount of cross-media linking (e.g., from printed word to video clip or still image; see Figure 4). In practice, however, printed-medium material can be far more easily hypertextualized than video data—the technology is simply more advanced. Consequently, we run the risk of letting the video dimension slip back into its conventional, merely illustrative role. This is a problem that may well be solved by new technological developments on the horizon—new versions of various hypertextual data analysis and authoring programs claim to be able to handle all kinds of media.

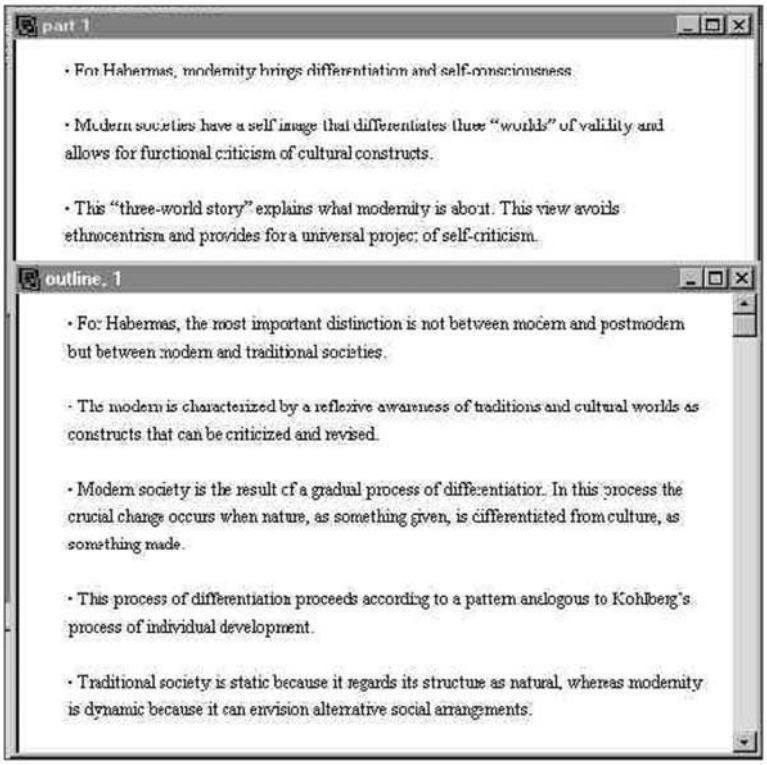

Figure 3: A Top-Level and Second-Level Node From the Habermas Pyramid, Part of Kolb's (1994) Hypertext *Socrates in the Labyrinth*.

The problem of linking video to text and vice versa can best be understood through references to Lynda Hardman's (1997) conceptualization of space and time in hypermedia. An image on a computer screen exists in a spatial dimension but has no temporal dimension. On the other hand, a sound clip played by a computer has no spatial presence but exists purely along a temporal axis. A video clip exists along both dimensions. To create a link from or to an image requires the ability to select a spatial area; to create a link to or from a sound clip requires setting up a temporal link. Integrating spatial, temporal, and spatiotemporal links has proved technologically challenging. Furthermore, text links function on neither of these dimensions. To integrate multimedia into hypertext requires every multimedia object to be specified in terms of relations of space, time, and textuality. At present, there is no software that manages this, although the proposed convergence between synchronized multimedia integration language for the Web (SMIL) with the Xlink working group drawing up protocol to extend the range and complexity of linking on the Web shows promise.

Finally, the presentation stage of research throws up a further set of decisions to be made for the incorporation of video material into the hypermedia environment. Hypermedia tends to encourage a fragmented approach to presentation, with writing broken down into small screens of information and argumentation that are easy on the eye, as in Figure 4. Clearly, video clips too could be kept brief, even to the level of single-shot sequences. However, the meaning of the moving image is produced through the syntagmatic combination of shots via montage as well as the paradigmatic plane of mise-en-scĕne (Marcus 1994a). As words gain meaning through their combination into creative sentences and narratives with their own rhe-

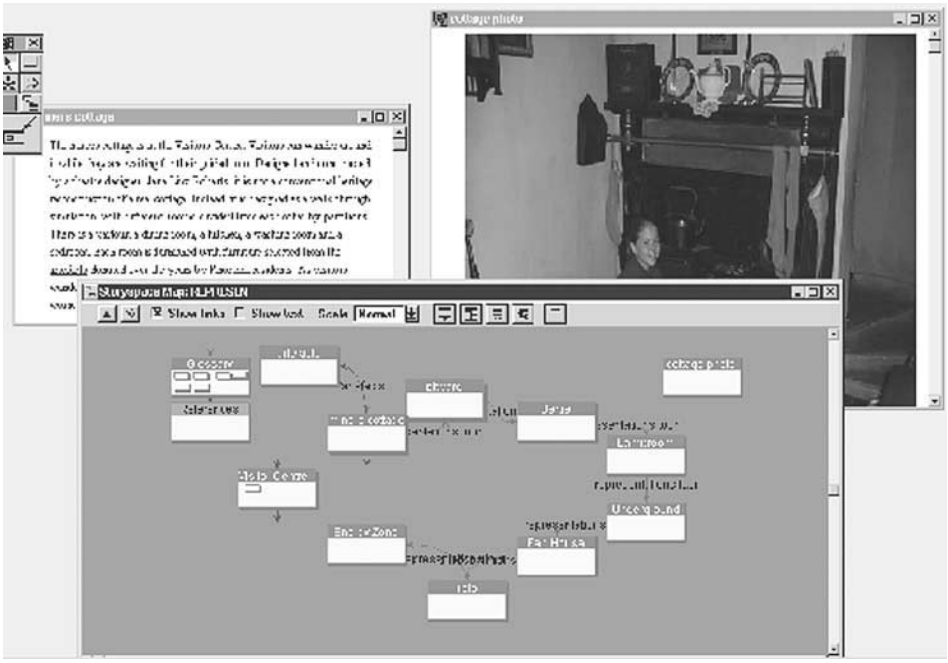

Figure 4: A Screen View From Our Own Ethnographic Hypermedia Environment in *StorySpace* Showing the Integration of Images and Text

torical devices, so does audio-visual material produce meaning through sequences of clips edited together. One option is to take advantage of this by using a video-editing program (such as Adobe Premiere) to create mini films to convey meaning. Again, this decision has to be balanced against the problems of computer storage requirements and the rudimentary state of current linking technology.

CONCLUSION

At this time, it is still not possible to store, analyze, interpret, and author hypermedia data in one medium. Storage problems make the analysis of video problematic while there is still no program that can handle complex hypertextual linking and synchronize multimedia elements. In our project, we are using *StorySpace* to analyze data and develop the hypertext structure and then *Authorware* to handle complex synchronization for authoring. However, new technology is rapidly appearing.

- Digital Versatile Disc (DVD) can store approximately 10 times the amount of information than a compact disc can, making it possible to store and organize larger amounts of video data.
- Hypertext-authoring packages such as *StorySpace* are gradually incorporating more and more multimedia elements.
- Web-based delivery of multimedia is also improving. In particular, the emergence of SMIL, allied to new innovations in link technology for the Web, means that Web-based delivery of hypermedia environments is beginning to be feasible.
- It is also likely that collaborative hypertextual systems, which are still somewhat experimental, will become more robust, allowing for the possibility of an online ethnography that could be added to by other researchers.

Hypermedia ethnography poses challenges and offers opportunities for the employment of digital technology. It seems likely that a lone ethnographer cannot produce hypermedia ethnography, and the collaborative possibilities offered with online hypermedia presentations encourage us to think of new forms and styles of research. The technology offers new media to explore, but its greatest effects may ultimately be on the culture of academic research, its methodology and forms of assessment, and the poetics of ethnographic writing.

NOTE

1. StorySpace 2, the latest version, supports images, Quicktime video, and sound. It also translates Storyspace documents to Web-readable HTML files, and Web addresses and links can be incorporated into the hypertext itself.

REFERENCES

Agar, M. (1980). *The professional stranger: An informal introduction to ethnography.* San Diego, CA: Academic Press.

Asch, T. (1988). Collaboration in ethnographic filmmaking: A personal view. In J. R. Rollwagen (Ed.), *Anthropological filmmaking: Anthropological perspectives on the production of film and video for general public audiences*. Chur, Switzerland: Harwood Academic.

Atkinson, P. (1990). *The ethnographic imagination*, p. 453. London: Routledge.

Atkinson, P., & Coffey, A. (1995). Realism and its discontents: On the crisis of cultural representation in ethnographic texts. In B. Adam & S. Allan (Eds.), *Theorising culture*. Los Angeles: UCLA Press.

Authorware 4.0 [Computer software]. (1997). San Francisco: Macromedia.

Barbash, I., & Taylor, L. (1997). *Cross cultural filmmaking*. Los Angeles: University of California Press.

Barbules, N. C. (1998). Rhetorics of the Web: Hyperreading and critical literacy. In I. Snyder (Ed.), *Page to screen: Taking literacy into the electronic era*. London: Routledge.

Biella, P. (1993a). Beyond ethnographic film: Hypermedia and scholarship. In J. R. Rollwagen (Ed.), *Anthropological film and video in the 1990s* (pp. 132-176). Brockport, NY: The Institute.

Biella, P. (1993b). The design of ethnographic hypermedia. In J. R. Rollwagen (Ed.), *Anthropological film and video in the 1990s* (pp. 293-341). Brockport, NY: The Institute.

Borchard, K. (1998). Between a hard rock and postmodernism: Opening the hard rock hotel and casino. *Journal of Contemporary Ethnography*, *27*, 242-269.

Bush, V. (1945). As we may think. *Atlantic Monthly*, *176*, pp. 101-108.

Chiozzi, P. (1989, April). Reflections on ethnographic film with a general bibliography (D. Dresner, Trans.). *Visual Anthropology*, 1-82.

Clifford, J. (1988). *The predicament of culture: 20th century ethnography, art and literature*. Cambridge MA: Harvard University Press.

Coffey, A., Holbrook, B., & Atkinson, P. (1996). Qualitative data analysis: Technologies and representations. *Sociological Research Online* [Online serial], *1*(1). Retrieved from the World Wide Web: `http://www.socresonline.org.uk/socresonline/1/1/4.html`

Dicks, B., & Mason, B. (1998). Hypermedia and ethnography: Reflections on the construction of a research approach. *Sociological Research Online* [Online serial], *3*(3). Available: `http://www.socresonline.org.uk/socresonline/3/3/3.html`

Dicks, B. & Mason, B. (1999). Cyber ethnography and the digital researcher. In J. Armitage & J. Roberts (Eds.), *Exploring cybersociety: Social, political, economic and cultural issues. Proceedings of the Exploring Cybersociety Conference* (Vol. 1). Newcastle, UK: University of Northumbria Press.

Dorst, J. D. (1989). *The written suburb: An American site, an ethnographic dilemma*. Philadelphia: University of Pennsylvania Press.

Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles in practice*. London: Routledge.

Hardman, L. (15.05.01). *Authoring hypermedia* [Online]. Unpublished doctoral dissertation, National Research Institute for Mathematics and Computer Science, the Netherlands. Retrieved from the World Wide Web: `http://www.cwi.nl/~lynda/thesis`

Hastrup, K. (1992). Anthropological visions: Some notes on visual and textual authority. In P. I. Crawford & D. Turton (Eds.), *Film as ethnography.* Manchester, UK: Manchester University Press.

Joyce, M. (1987) 1st edition. *Afternoon: A story* [Hypertext 3.5 inch disc]. Watertown, MA: Eastgate Systems Inc.

Kolb, D. (1994). *Socrates in the labyrinth: Hypertext, argument, philosophy* [Hypertext CD-ROM]. Watertown, MA: Eastgate Systems Inc.

Kress, G. (1998). Visual and verbal modes of representation in electronically mediated communication: The potentials of new forms of text. In I. Snyder (Ed.), *Page to screen: Taking literacy into the electronic era (*pp. 53-79). London: Routledge.

Landow, G. P. (1997). *Hypertext 2.0: The convergence of contemporary critical theory and technology* (Rev. ed.). Baltimore, MD: Johns Hopkins University Press.

Larsen, D. (1996). *Samplers: Nine vicious little hypertexts* [Hypertext CD-ROM]. Watertown, MA: Eastgate Systems Inc.

Lee, R. M., & Fielding, N. (1996). Qualitative data analysis: Representations of a technology: A comment on Coffey, Holbrook and Atkinson. *Sociological Research Online* [Online serial], *1*(4). Retrieved from the World Wide Web: `http://www.socresonline.org.uk/socresonline/1/4/lf.html`

Marcus, G. E. (1994a). The modernist sensibility in recent ethnographic writing and the cinematic metaphor of montage. In L. Taylor (Ed.), *Visualising theory.* London: Routledge.

Marcus, G. E. (1994b). What comes (just) after "post"? The case of ethnography. In N. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research*. Thousand Oaks, CA: Sage.

Marcus, G. E., & Fischer, M. (1986). *Anthropology as cultural critique: An experimental moment in the human sciences*. Chicago: University of Chicago Press.

McLuhan, M. (1964). *Understanding media: The extensions of man*. Boston: Routledge Kegan Paul.

Morgan, W. (1999). Tools for dismantling the master's house: Postructuralist feminist research and hypertext poetics. In K. Tochterman, J. Westbomke, U. K. Will, & J. Leggett (Eds.), *Hypertext 99: Returning to our diverse roots, Proceedings of the 10th ACM Conference on Hypertext and Hypermedia*. New York: Association of Computing Machinery.

Morphy, H., & Banks, M. (1997). Introduction: Rethinking visual anthropology. In M. Banks & H. Morphy (Eds.), *Rethinking visual anthropology* (pp. 1-35). New Haven, CT: Yale University Press.

Moulthrop, S. (1991). *Victory garden* 1st edition. [Hypertext 3.5 inch disc]. Watertown, MA: Eastgate Systems Inc.

Nelson, T. (1987). *Computer Lib.: Dream machines*. Seattle, WA: Microsoft Press.

Ong, W. J. (1992). Writing is a technology that restructures thought. In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 293-319). Amsterdam: John Benjamins.

Richardson, L. (1994). Writing: A method of inquiry. In N. K.Denzin & Y. S. Lincoln (Eds.), *A handbook of qualitative research* (pp. 923-948). Thousand Oaks, CA: Sage.

StorySpace 1.7.5 for Windows [Computer software]. (1999). Watertown, MA: Eastgate.

Van Maanen, J. (1995). An end to innocence: The ethnography of ethnography. In J. Van Maanen (Ed.), *Representation in ethnography.* Thousand Oaks, CA: Sage.

Weaver, A., & Atkinson, P. (1994). *Microcomputing and qualitative data analysis*. Aldershot, UK: Avebury.

Weaver, A., & Atkinson, P. (1995). From coding to hypertext: Strategies for microcomputing and qualitative data analysis: In R. C. Burgess (Ed.), *Studies in qualitative methodology.* Greenwich, CT: JAI.

Weinberger, E. (1994). The camera people. In L. Taylor (Ed.), *Visualising theory: Selected essays from V.A.R. 1990-1994* (pp. 3-26). London: Routledge.

Bruce Mason is teaching and completing his Ph.D. in the folklore department at Memorial University of Newfoundland, Canada.

Bella Dicks is a lecturer in sociology at Cardiff School of the Social Sciences, Cardiff University, United Kingdom.

REPORTS AND COMMUNICATIONS

A Digital Library for the Dissemination and Replication of Quantitative Social Science Research

The Virtual Data Center

MICAH ALTMAN
LEONID ANDREEV
MARK DIGGORY
GARY KING
AKIO SONE
SIDNEY VERBA
Harvard University

DANIEL L. KISKIS
MICHAEL KROT
University of Michigan

The Virtual Data Center software is an open-source, digital library system for quantitative data. The authors discuss what the software does, how it provides an infrastructure for the management and dissemination of distributed collections of quantitative data, and the replication of results derived from these data.

Keywords: open-source, digital library, quantitative data, replication

Researchers in social sciences, and in academia in general, increasingly rely on large quantities of numeric data. The analysis of such data appears in professional journals, scholarly books, and more and more often, popular media. For the scholar, the connection between research articles and data is natural. We analyze data and publish results. We read the results of other analyses, learn from them, and move forward with our own research.

But these connections are sometimes difficult to make. Data supporting an article are often difficult to find and even more difficult to analyze. Archiving, disseminating, and sharing data is crucial to research but is often costly and difficult (Sieber, 1991). Consequently, our ability to replicate the work of others and to build on it is diminished. Researchers, university data centers, and students all face challenges when trying to find and use quantitative research data.

AUTHORS' NOTE: Supported by National Science Foundation award No. IIS-9874747.

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 458-470

The Virtual Data Center (VDC) software is a comprehensive, open-source, digital library system designed to help curators and researchers face the challenges of sharing and disseminating research data in an increasingly distributed world (Altman et al., 2001). The VDC is also a first step toward better citation of data. Current citations of data are typically ad hoc, fragile, and shallow. Ultimately, digital libraries such as the VDC will serve to make citations more robust and research more replicable.

VDC FEATURES

The system provides five areas of functionality:

1. *Study preparation:* unique naming and conversion tools for multiple data and documentation formats and tools for preparing catalog records for datasets;
2. *Study management:* file-system independent data set and documentation storage, archival formatting, cataloging;
3. *Interoperability:* Dublin Core, MARC, and DDI (Data Documentation Initiative) metadata import and export and OpenArchives and Z39.50 query protocol support;
4. *Dissemination:* extract generation, format conversion, and exploratory data analysis;
5. *Distributed and federated operation:* location-independent name resolution, distributed virtual collections, federated metadata harvesting, repository exchange and caching, and federated authentication and authorization.

The VDC provides functionality for users, curators, and producers of data. For users, it enables online search, data conversion, and exploratory data analysis facilities. For curators, it provides facilities to create virtual collections of data that bring together and organize data sets from multiple producers. For producers, it offers naming, cataloging, storage, and dissemination of data.

Consider the following use cases: First, an undergraduate is writing a term paper on the 1996 U.S. presidential election; next, a graduate student in the School of Public Health is researching the epidemiology of heart disease in France; finally, a senior professor in the economics department is, for the first time, testing new models of the political factors affecting economic growth. At first look, all three users appear to have very different research needs. The student needs to find a single number—the percentage of women in the Northeast who voted for Clinton. The graduate student is attempting to extract a large subset of data from a larger study along with an accompanying geographic map of the data; the senior professor, meanwhile, needs to develop an extensive set of data comprising interrelated variables from dozens of data sets. Although the magnitudes of these research tasks are different, each researcher faces the same set of core tasks. These tasks all involve searching for a relevant data set, extracting an appropriate subset of the data, and constructing a summary of the data. (These tasks are illustrated in Figure 1.)

For the curator of the university data center and/or library, the VDC provides efficient and flexible dissemination of the collection. The curator can use the VDC system to make all data sets available online through a consistent set of user interfaces. The VDC also assists with preservation of the collection by converting datasets into XML (Extensible Markup Language), preservation-friendly formats and by separating the methods used to access data sets from the storage technology.

For producers, the VDC simplifies archiving, naming, and coordination with disseminators and end users. Through its implementation of the DDI specification, it ensures the information they used in creating the data is retained in the dissemination process and eventually delivered to the end users. In addition, the ability to attach persistent, unique iden-

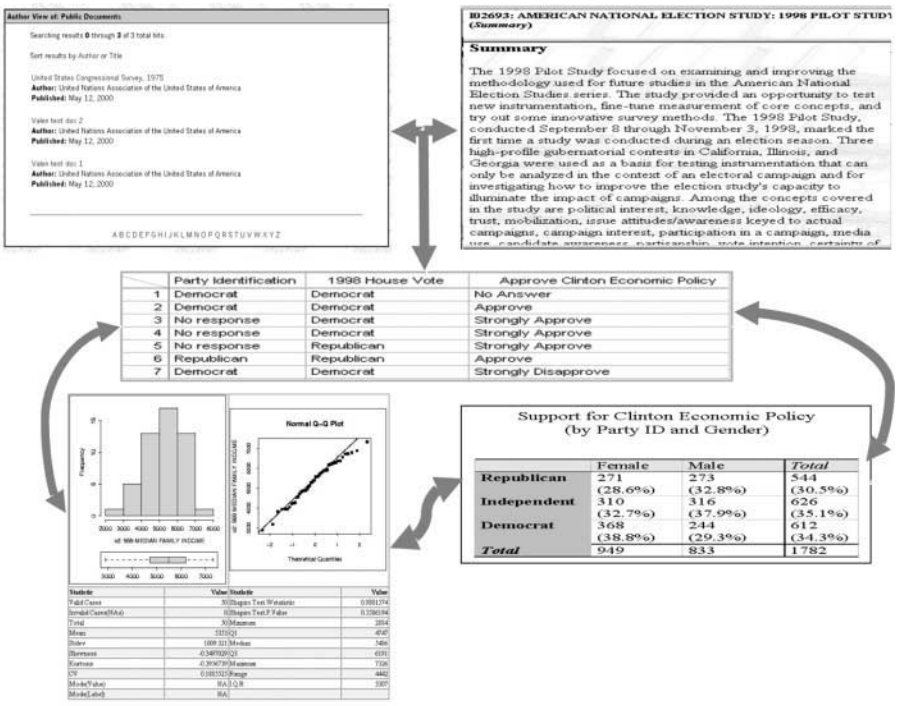

Figure 1. User Interaction With Prototype Virtual Data Center

tifiers to a particular dataset allows the producer to disseminate data through multiple archives while retaining citations that always point back to the original producer. When a producer serves objects to curators through the VDC, the system will ensure that the curator can access any objects for which they are authorized and that the curator always has an up-to-date copy of metadata and object. If multiple curators share collections, the VDC also guarantees that multiple copies of the same item are assigned the same identifier. That is, the system will "borrow" items from other curators (if authorized), reducing the load on the producer's server. Furthermore, multiple copies or the item are presented to the user as a single work with multiple locations, making it easier to find relevant works from that producer.

IMPLEMENTATION: BALANCING RESEARCH AND PRODUCTION REQUIREMENTS

Our implementation strategy emphasizes *open source* development and integration of the system into a production environment. The director of the Digital Library Initiative, Phase 2, of which the VDC is a part, noted the "unnatural separation" between the producers and consumers of digital libraries and called for a balance among research, application, content, and collections (Griffin, 1998). In keeping with this admonition, the VDC software system is not simply an isolated research project but is also a part of Harvard University's first generation production digital library system. VDC benefits from participation in an unusually large and decentralized library system, from cross-fertilization with Harvard's own digital library efforts (see Flecker, 2000), and from the heavy usage patterns of the Harvard research community.

The requirement that the system support production use in a decentralized environment has a number of implications. First, the architecture must be flexible enough to accommodate the administration of collections and their contents by multiple and independent curators. Second, the system must be accessible by standard Web browsers without special configuration. Third, the system must support the protocols and standards currently in use in library environments. Conversely, as a first-generation production system, there is much in the way of architecture, implementation strategy, and features that remains to be discovered.

VDC CORE ARCHITECTURE

The VDC digital library borrows core concepts from Arms, Blanchi, and Overly (1997) and from NCTSRL (Networked Computer Science Technical Reference Library) (Davis & Lagoze, 1999) and extends these in innovative ways to support services on digital objects, complex collections, distributed authentication authorization, and deep citations. Moreover, the objects stored in the VDC system differ in important ways from these earlier systems.

The basic object managed in the system is the study. Each study comprises a meta-data object and a set of associated data objects. The metadata object follows the DDI standard (Ryssevik, 1999) and contains all of the structural metadata for that study as well as the descriptive meta-data for the corresponding (abstract) intellectual work. The associated data objects consist of text files (usually for supplementary documentation), MIME (Multipurpose Internet Mail Extensions)–typed BLOBs (Binary Large Objects), and/or structured quantitative databases. The metadata object acts to document the study and to tie the associated data objects together.

These objects are managed with a set of cooperating services. The core of the architecture supports four services: the UIS (User Interface Service), the repository service, the name resolution service, and the index service. (See Figure 2. Core components are shown in white.)

The UIS is the gateway to the system and coordinates access to the other components. The UIS supports two user interfaces: one for the end users of the library and another for the curators who manage the collections. Both are accessed through a standard Web browser.

The UIS is implemented as a set of Java servlets, each of which encapsulates access to particular services and objects. Each object or service is itself described in XML, and XSL (Extensible Stylesheet Language) is used to render the object.

The repository stores and manages digital objects and the administrative metadata (such as the object's owner or last time of access) associated with them. A repository access protocol allows for maintenance and hiding the details of their storage (currently a SQL [Structured Query Language] database) from the rest of the system. The repository itself treats every object as a MIME–typed BLOB. All knowledge about complex objects (objects that cannot be rendered by a browser without preprocessing) is encapsulated inside the UIS.

The NRS (Name Resolution System) manages identifiers for each digital object. Each distinct intellectual work stored in the system is assigned a unique identifier. The NRS uses URN (Uniform Resource Name) methods (Daniel, 1997; Moats, 1997) to resolve each identifier to a repository (or set of repositories) that stores a copy of that work.

The IS (Index Server) manages indexing and searching (queries) of the descriptive metadata associated with each object. Index servers act with a large amount of independence—they are assigned sets of identifiers that they are responsible for indexing. In addition, the index servers asynchronously resolve the identifiers to a repository, retrieve the metadata component of these objects, and build indices based on this metadata.

Together, these four services provide the core of digital library functionality. To support specialized services on these objects and to support distributed operations, we introduced a

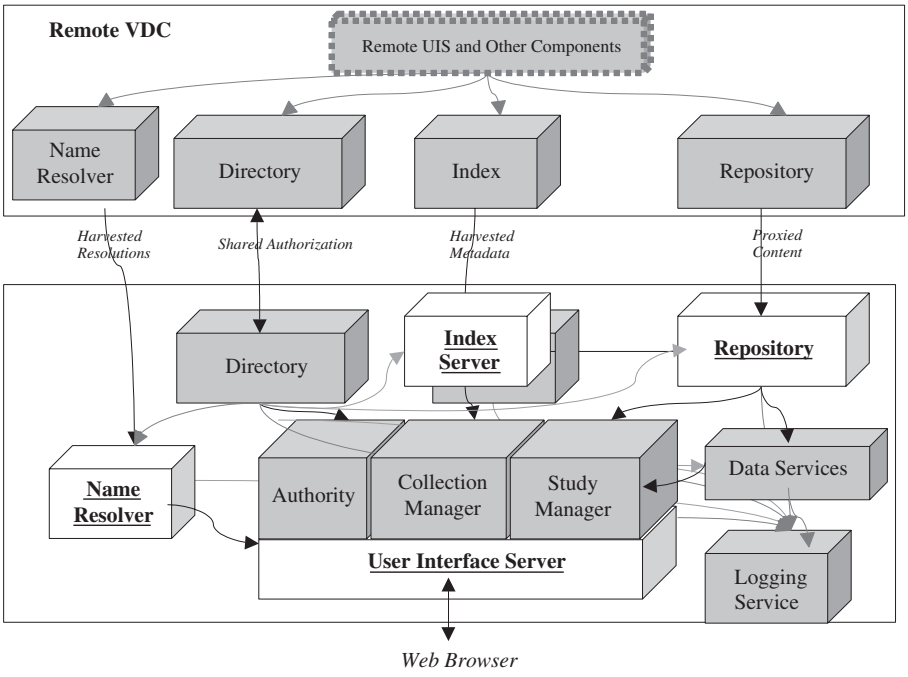

Figure 2: Simplified Representation of VDC Architecture
NOTE: VDC = Virtual Data Center.

number of other services (see Figure 2; new services are shown in gray). We discuss these new services in the next section.

MULTIPLE MODES OF DISTRIBUTION: DISTRIBUTED COMPONENTS AND FEDERATED LIBRARIES

One of the more innovative aspects of the VDC is its support for multiple models of distributed operation. Each VDC library comprises a set of interoperating components that can function independently and that can be distributed across systems. In addition, the VDC supports "virtual collections" within a VDC library that bring together studies indexed in multiple index servers and stored in multiple repositories. Moreover, unlike the NCSTRL system, independent VDC libraries can cooperate together in a federation to share individual studies or entire collections.

In the simplest form of distributed operation, components are distributed across multiple systems or networks, but only a single administrative unit is formed (see Figure 2, bottom portion). To support this scenario, the VDC architecture introduced three new components to the Arms–NCSTRL architecture.

1. A directory services component provides a central registry for all other components in the administrative unit. This service enables components to locate each other and allows components to be added or removed from the system dynamically.
2. A centralized logging facility provides an interface where each of the distributed components can record events that occur in the process of servicing a request. This supports debugging of the system and auditing of the system usage.

3. A data services broker coordinates reformatting, subsetting, aggregation, and analysis of study objects. This component supports the UIS in performing services on data objects for users. Because the UIS consults the broker about what services can be performed on each object, it is possible to add other types of services without modification to existing components.

(Because services are provided through a local broker, these services can be applied to studies copied from other parts of the federation. This is possible even where the federation member providing the study does not support these services locally. A discussion of this federated model follows.)

These three components, along with the repository, index server, name resolver, and UIS (described in the previous section) cooperate to form a single "library" service. Users see one point of presence, and one set of administrative rules is maintained there for the VDC library unit.

The second mode of distributed operation adds an additional dimension of flexibility: Multiple independent VDC systems can be "federated" together to share collections. In this mode, LDAP (Lightweight Directory Access Protocol) referrals are used in each local directory server to point to the directory server of cooperating VDC libraries. The harvester in each library then uses these referrals to locate remote indices. It harvests metadata from the indices to replicate indexing and name resolution information across local name resolvers, enabling any member of the federation to find copies of studies stored in a neighboring repository. When such studies are requested by local components, a proxy is then used to retrieve and cache copies from neighboring repositories.

Authentication and authorization is also federated. Each VDC library maintains authority over how studies in its repositories are accessed. Distributed authentication works as follows: A user can log in from any UIS in the system but must identify their home VDC in the process. The user is then redirected to their home VDC for authentication, which, if successful, supplies signed credentials that identify the user as a member of that institution. Finally, the user is redirected back to the UIS where they originally entered.

When an authenticated user makes a request of the system, this request must then be authorized. Authorization is role based—a user is mapped to multiple roles based on his or her localized profile of attributes (e.g., membership status). Each study is assigned to one or more logical access classes, as determined by its curator. To authorize an operation on an object, the system looks for a {role, class, operation} entry in the local VDC authority table.

In a federated context, access to an object is always determined by the VDC owning that object. The owning system authorizes each request by (a) identifying a remote user's home library from their credentials, (b) mapping that user's attributes to a set of roles at the home library, (c) mapping the home roles to a set of roles within the "owning" library, and then (d) searching for a {role, class, operation} entry in the local authority table.

This process is analogous to how brick-and-mortar libraries function: Guest borrowers can present library cards from a cooperating institution, signaling that they are authentic members of that institution. The local library then assigns them, for example, "guest faculty" status and authorizes access to its materials on this basis. The result is that content from a group of libraries is made available to the users of each library while each library maintains complete control over how its collections are accessed and over the authentication of its patrons.

VIRTUAL COLLECTIONS

In addition to federation, VDC also supports a complementary way of creating and organizing distributed collections—the "virtual" collection. Virtual collections give the curator an opportunity to directly mediate between users and sets of studies. By creating a virtual collection, the curator identifies a body of logical content and how that content should be represented to the user, regardless of whether that content is owned by the curator. Virtual collections are managed by the collection service component with the support of the metadata harvester service.

Figure 3 shows the architecture for virtual collections. Each virtual collection comprises a specification of the content of that collection and a set of views to be applied to that content. The content of a virtual collection comprises a set of queries that are run against a set of local index servers or other collections. Views are then overlaid on the content to provide navigation and display. In addition, the harvesting component runs asynchronously to gather metadata from remote index servers and remote collections, which is then replicated in a local index server and can be used as content for local collections.[1]

For example, consider a curator who wishes to create a virtual collection of studies about Argentina. The content selection rule would specify a set of index servers or other collections in which relevant data are likely to be found as well as a query for all studies in those servers that have Argentina listed in the coverage metadata attribute. Thus, the content of the virtual collection is not fixed—as new studies are added to repositories, they are indexed by the index servers and dynamically incorporated into the virtual collection.

The curator would also designate a set of views that should be applied to this content. He or she might create their own views or use views already supplied by the VDC system. Some examples of views include the following:

- A simple search interface, which allows the user to search Dublin Core fields.
- A recent additions list that filters the query results by creation time and shows the 100 most recently created studies.
- A thematic outline that shows the content as organized by LOC (Library of Congress) subject classification (if available) or other controlled vocabulary.
- An author index that determines the list of authors by analyzing the collection content itself (as opposed to using a controlled vocabulary, as above) and groups the studies in the collection by author.

These views are not static renderings of particular content but rather are logical descriptions of how the content should be searched, navigated, and displayed. So, curators can reuse views and apply them to different virtual collections with different content wherever the metadata attributes used by the view are present in the virtual collections.

Virtual collections are flexible and powerful because they make use of multiple layers of distributed services. Index servers can index items in distributed repositories, and harvesters use standard protocols to gather indices from remote index servers and remote collections. This means it is possible to create virtual collections that extend across, or even beyond, a federation.

ENABLING DEEP CITATIONS IN ACADEMIC JOURNALS

A fundamental goal of the VDC project is to increase the replicability of research by providing a foundation for "deep citation" of quantitative data. The principle that references to data and data analysis be specific enough to support replication of the research is widely

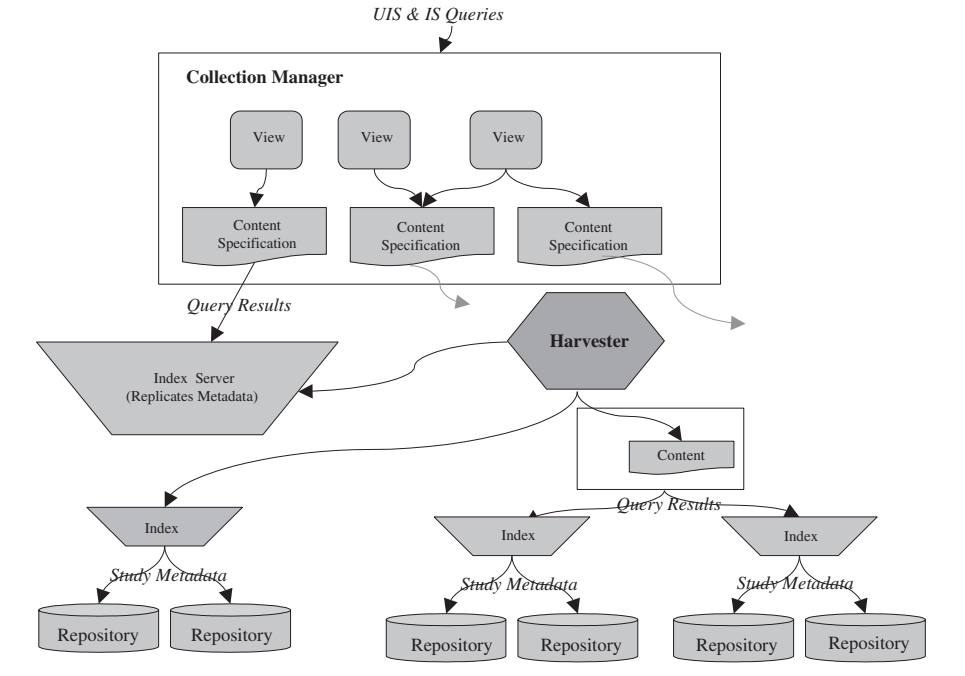

Figure 3: Virtual Collection
NOTE: UIS = user interface service, IS = index server.

accepted as a critical part of publishing (King, 1995). In many other fields, the premier journals, such as *Science*, *Nature*, *American Economic Review*, *Journal of the American Statistical Association*, *Journal of the American Medical Association*, *Social Science Quarterly*, *Lance*, *American Political Science Review*, and the various journals published by the American Psychological Association, explicitly document in submission guidelines that the provision of important data on which the article is based is a condition of publication.[2]

Currently, however, citations (and other references) to shared data are generally insufficient to uphold this principle. Citations to quantitative data as used today, however, suffer from three fundamental problems. First, in published articles, citations to data are ad hoc: They are far less specific and systematic than references to textual sources. No standard citation format exists to describe what data sets were used in a research article. Although data used in research articles are sometimes discussed in footnotes, they are almost never included as a formal citation in a bibliography of references.

Second, citations to data are typically fragile: They fail to answer the most basic questions needed for location and verification of the data over the long term. Which version of the public data set, exactly, did the author use? And, is the version of the data used accessible from a data archive (or library or publisher), although perhaps in a different format, still intellectually equivalent to the version used by the author? Moreover, whereas large public archives have been tremendously successful in obtaining large data collections, many research articles are based on small sets of data that are either collected by the researcher or derived from multiple sources in ways that are not rigorously documented.

Third, citations of data are usually shallow. Citations of data rarely contain precise information about the portion of the data set used and the manner in which it was manipulated. Even where it is possible to find the relevant data set, it is seldom possible for any reader to

easily and unambiguously traverse the chain of empirical evidence amassed to support the conclusions published: What are the full details of calculations on the data? How did the analyst move from publicly available data to numerical results? What coding rules were used for which cases? Were variables recoded, and if so, how? How were multiple data sets merged or subsets of the data selected? How were missing values treated in the analysis?

Moreover, datasets have a fundamentally different "functional granularity" (Paskin, 2000) than journal articles—it is essential for a citation of data to enable the researcher be refer exactly to particular data elements used and transformations made upon those elements so as to be able to reproduce published results (Williams, Bunn, Moore, & Poole, 1998).

Several brief examples, drawn from our experience as researchers and data librarians and from personal communication with administrators of data collection projects and archives, illustrate this point. As for ad hoc use of study numbers, many major data archives, such as those of the U.K. data archive, the ICPSR (Inter-University Consortium for Political and Social Research), and the ROPER (Roper Center for Public Opinion Research), use acquisition numbers to informally identify their systems. Often, researchers use these study numbers in an attempt to uniquely identify the data set used in an article. Unfortunately, even during the past 10 years, the limitations of this practice have been revealed.

- A major archive renumbered its acquisitions, invalidating, or rendering ambiguous, many previous references.
- In some cases, a dataset published by a third party (e.g., the U.S. Government or Gallup) is disseminated by multiple archives. When this happens, each archive typically assigns the data set a different acquisition number. Thus, references using these study numbers appear to refer to different data but actually implicate the same intellectual object.
- In some cases, a publisher withdraws data from the data archive. This invalidates the study number, although the data set may continue to be available from other sources.
- When a cumulative research study is extended (e.g., with another wave of data), the previous study number may be "deaccessioned" and a new one assigned. Again, although the relevant data continues to be available, the citation to it becomes invalid.
- Researchers distribute slightly (sometimes substantially) modified versions of data sets that also exist in archives but refer to these in publications as they would the original.

Examples of the fragility of references include the following:

One of the largest continuing data collection efforts in political science uses the CIESIN (Center for International Earth Science Information Networks) geographic correspondence database Web site (Geocorr) when aggregating their data for release to the public. Because neither the Geocorr database nor the methodology behind its creation has ever been explicitly published, when the Geocorr Web site is updated, it may become impossible to reproduce exactly the particular aggregation rules used in previous studies.

The Bureau of Labor Statistics' Current Population Survey, which is widely used by economists, often corrects or updates a particular survey after its initial release. Often, these (admittedly minor) updates are available only online, are not announced, are not tracked by other disseminators (such as ICPSR), and do not change the version number of the data set. It is almost a certainty that the data obtained by a researcher who tracks a reference in a published article back to the CPS (Current Population Survey) Web site will be at least marginally different from the data used by the original author(s) in the published article.

The Poole-Rosenthal congressional roll-call voting scores have been widely used in recent years in analyses of congressional member behavior, and ICPSR archives the data used in the original 1984 study of Poole and Rosenthal and several subsequent years (Poole & Rosenthal, 1984). However, data covering an additional decade, along with corrections to

the original voting data in ICPSR's collection, appear only on the Web site of one of the authors, and it has moved several times since it was created.

In regard to the shallowness of current references, consider the following:

In the course of our own research in only the past few years, we have found it productive to replicate more than a dozen research studies (Altman & McDonald, 1999; King, Honaker, Joseph, & Scheve, 2001; King, Tomz, & Wittenberg, 2000; King & Zeng, 2001). Because we had chosen articles in which the data were derived from public sources, the replication process should have been simple and straightforward. But our experience was that many barriers to replication persist. In practice, we find that current replication policies are not sufficient, in part because of the imprecision with which data are cited. Although some authors provided online related materials, typically, we were still unable to reproduce results without contacting authors about missing data management details or supplementary data.

In one replication we attempted recently (Altman & McDonald, 1999), the original author, who had preserved the data from his study, could not successfully reproduce that data from current, supposedly identical sources.

These examples illustrate the range of ad hoc, fragile, and shallow ways in which data are cited at present within the social sciences. Our findings are not isolated, however—similar problems have been reported in other fields (Feigenbaum & Levy, 1993; McCullough & Vinod, 1999). The reliable exchange of research data is vital to the advancement of the social sciences (Ceci & Walker, 1983; Sieber, 1991), and the requirement that another researcher be able to understand, evaluate, build on, or reproduce research without any additional information from an author (King, 1995) is a fundamental principle of scientific research.

In contrast to the system of ad hoc study numbers now in use, a citation to a dataset will enable a particular data set originally obtained in the holdings of one distributor to be found in the holdings of another. In addition, this standard would enable the researcher to precisely and uniformly describe the variables and observations extracted from the dataset to support each particular published result (e.g., tables and figures) appearing in the journal article. This standard would also, as much as possible, allow common transformations and recodings of data to be recorded in a standardized way. Moreover, this standard would specify protocols for versioning datasets. This will allow researchers to consistently recognize substantive changes to a dataset (e.g., when data is adjusted) and consistently ignore nonsubstantive changes (e.g., when data is converted from SPSS to SAS format), and citations referring to "old" datasets will be able to be correctly mapped to newer versions, where possible (e.g., when a new year of data is added to a dataset and the previous study number is abandoned.)

The VDC system takes a substantial first step toward improving citation to data. First, by providing every study in the system with a persistent, location-independent identifier, links to a study will remain valid when repositories are relocated[3]—they are more robust. Second, by converting every study into a canonicalized XML format, we also contribute to the robustness of the citations because the study then becomes insulated from changing statistical software. Third, by developing a syntax for specifying the subset of the study used in an analysis and embedding that in the URL (Uniform Resource Locator) with the persistent identifier, we support citations that are "deeper" than those currently used to refer to data. Fourth, by providing these functions with an open-source system (see as follows) that anyone can examine or use, we take steps toward standardizing the current ad hoc system.

For deep citations to become a reality in academic publishing, however, many technological and institutional mechanisms must still be developed. Naming conventions, registration methods, and citation protocols for publishing and citing datasets must be refined, standardized, and recognized. Moreover, citations to data must be integrated with the systems and databases currently used to manage citations to journal articles.

CREATING AN OPEN-SOURCE INFRASTRUCTURE FOR SCIENTIFIC RESEARCH

When Edison invented the electric light, he also had to design a system that would deliver electricity to his customers and found companies to manufacture it (Cowan, 1997). Even 10 years ago, digital library development faced similar challenges. Now, however, infrastructure and standards are developing rapidly, and we believe that open source now provides a strong base for scientific research and infrastructure.

In the VDC project we have built on the work of numerous other open-source projects instead of "reinventing the wheel." The *R* statistical language (Ihaka & Gentlemen, 1996) is used extensively in the data services component; the PostGres database system (Momjian, 2000) is the basis for our current repository component; the Apache Web server (Laurie, Laurie, & Denn, 1998) and Apache's Jakarta-Tomcat servlet engine provide a foundation for the user interface server (as well as other components); and OpenLDAP (`http://www.openldap.org/`) is used as the basis of the directory service and extensively in the distributed authentication and authorization components. Building on these existing projects enabled us to focus our implementation efforts on the innovative architectural aspects of the system, such as distributed virtual collections, rather than, for example, the nuts and bolts of protocol implementation or file storage.

An open-source development strategy has additional advantages. First, exposing the source code to a wide community of programmers makes it more likely that bugs will be spotted promptly and fixed. Second, the code can be adopted by those who find it useful and so will continue to progress after the project has ended (Raymond, 1999).

ACCESS TO THE VDC SOFTWARE AND ITS CONTENTS

It is our goal that the VDC be of use as part of the infrastructure for doing scientific research. Thus, to support the academic norms of openness and accessibility associated with research data, we are in keeping with Lessig's (1999) assertion that the code supporting the fundamental infrastructure for citations must be open. Our source code is open source and freely available to anyone for use, examination, and modification, forever (see `http://TheData.org`).

Although the code itself is freely available, some data served through the VDC system will not be so. We think there is great value in making the research data itself available freely, and we will make research data that we have produced freely available through the system and are encouraging our partner sites to do the same. We believe that there is even greater value in making the metadata available freely so that others can, at least, discover that a particular collection is in possession of something of interest, even if the data itself is restricted. We also realize, however, that there are often compelling reasons why some data cannot be disseminated publicly. Anything placed within the VDC software is subject to the licensing restrictions imposed on it by the producer and disseminator. Thus, our system supports flexible distributed authentication and authorization mechanisms to give the disseminator complete control over which users can access which parts of the data and what they can do with it within the system. For example, the curator of a data collection can make all data in the system publicly available or restrict access to certain data classes of local or federated users, such as users from ICPSR member institutions. The curator can also apply fine-grained restrictions on particular variables or even restrict how metadata is shared. Although, again, we encourage curators who use the VDC to make their collections and the metadata for them publicly available where possible.

CONCLUSION

Researchers, archives, and casual users all face many challenges when trying to find and use quantitative research data—and the replicability of research suffers as a result. Digital libraries can help to overcome these challenges. By providing a portable software product that makes the process of data sharing automatic and standardized, we believe that the VDC will help researchers and data archives to meet the challenges of sharing and using quantitative data and take the first step toward support for deep citations of quantitative data.

NOTES

1. The NCSTRL collections component (Davis & Lagoze, 2000) corresponds to the content specification in our architecture. The NCSTRL does not have functionality that directly corresponds to VDC views and does not support the harvesting of index servers.

2. We gathered information from the Contributors section of the 1997 edition of each journal. The exact terms of the requirement to share data vary both in principle and in practice.

3. Because URN (Uniform Resource Names) are not resolvable within standard Web browsers, we wrap the URN in a PURL (Persistent Uniform Resource Locator) (Shafer, Weibel, Jul, & Fausey, 1997) so that it can be accessed using widely available browser technology. Because PURL's are limited to a one-to-one resolution, one copy of the item (the copy contained in the VDC node that published the work originally) in the system is designated to be the canonical copy for purposes of PURL resolution.

REFERENCES

Altman, M., Andreev, L., Diggory, M., Krot, M., King, G., Kiskis, D., Kolster, E., Sone, A., & Verba, S. (2001). An introduction to the Virtual Data Center project and software. In, *Proceedings of the first ACM+IEEE Joint Conference on Digital Libraries*. New York: ACM.

Altman, M., & McDonald, M. (1999, July 17). *The robustness of statistical abstractions: A look under the hood*. Paper presented at the annual Society for Political Methodology meeting, College Station, Texas.

Arms, W. Y., Blanchi, C., Overly, E. A. (1997). An architecture for information in digital libraries. *D-Lib Magazine* [Online serial], *3*(2). Retrieved May 1, 2001, from the World Wide Web: `http://www.dlib.org/dlib/february97/cnri/02arms1`

Ceci, S., & Walker, E. (1983). Private archives and public needs. *American Psychologist*, *38*, 414-423.

Cowan, R. S. (1997). *A social history of American technology*. New York: Oxford University Press.

Daniel, R. (1997). *A trivial convention for using HTTP in URN resolution* [Online] (Internet Engineering Task Force, No. RFC 2169). Retrieved from the World Wide Web: `http://www.ietf.org/`.

Davis, J. R., & Lagoze, C. (2000). NCSTRL: Design and deployment of a globally distributed digital library. *JASIS*, *51*, 273-280.

Feigenbaum, S., & Levy, D. M. (1993). The market for (ir)reproducible econometrics. *Social Epistemology*, *7*, 215-292.

Flecker, D. (2000). Harvard's library digital initiative: Building a first generation digital library infrastructure. *D-LIB Magazine* [Online serial], *6*(11). Retrieved May 1, 2001, from the World Wide Web: `http://www.dlib.org/dlib/november00/flecker/11flecker.html`

Griffin, S. M. (1998). NSF/DARPA/NASA digital libraries initiative: A program manager's perspective. *D-Lib Magazine* [Online serial], *4*(7). Retrieved May 1, 2001, from the World Wide Web: `http://www.dlib.org/dlib/july98/07griffin.html`

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*, 299-314.

King, G. (1995). Replication, replication. *PS: Political Science and Politics*, *28*, 443-499.

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, *95*, 49-69.

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, *44*, 341-355.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*, 137-163.

Laurie, B., Laurie, P., & Denn, R. (Ed.). (1998). *Apache: The definitive guide*. Sebastapol, CA: O'Reilly and Associates.

Lessig, L. (1999). *Code, and other laws of cyberspace*. New York: Basic Books.
McCullough, B. D., & Vinod, H. D. (1999). The numerical reliability of econometric software. *Journal of Economic Literature*, *37*, 633-665.
Moats, R. (1997). *URN syntax* [Online] (Internet Engineering Task Force, No. RFC 2141). Retrieved from the World Wide Web: `http://www.ietf.org/`
Momjian, B. (2000). *PostgresSQL: Introduction and concepts*. Reading, MA: Addison-Wesley.
Paskin, N. (2000). E-citations: Actionable identifiers and scholarly referencing. *Learned Publishing*, *13*(3), 159-166.
Poole, K. T., & Rosenthal, H. (1984). The polarization of American politics. *Journal of Politics*, *46*, 1061-1079.
Raymond, E S. (1999). *The cathedral and the bazaar.* Sebastapol, CA: O'Reilly & Associates.
Ryssevik, J. (1999, September 22-24). *Providing global access to distributed data through metadata standardization. The parallel stories of NESSTAR and the DDI*. Paper presented at the UN/ECE work session on statistical metadata, Geneva, Switzerland.
Shafer, K., Weibel, S., Jul, E., & Fausey, J. (1997). *Introduction to persistent uniform resource locators*. Dublin, OH: OCLC Online Computer Library Center.
Sieber, J. E. (Ed.). (1991). *Sharing social science data. Sharing social science data*. Newbury Park, CA: Sage.
Williams, R., Bunn, J., Moore, R., & Pool, J.C.T. (1998). *Workshop on interfaces to scientific data archives* [Online] Pasadena, CA: California Institute of Technology, Center for Advanced Computing Research. Retrieved May 10, 1999, from the World Wide Web: `http://www.cacr.caltech.edu/isda`.

Micah Altman (Ph.D., California Institute of Technology) is the director of the Virtual Data Center project, which aims to promote the sharing of research data by building open-source software tools. He is also the associate director of the Harvard-MIT Data Center and a postdoctoral fellow in the Department of Government at Harvard University. Dr. Altman has published (or has in press) more than 10 articles, datasets, and reviews on the topics of redistricting numerical accuracy, and digital libraries.

Leonid Andreev is a systems programmer for the Harvard-MIT Data Center and a member of the Virtual Data Center project team.

Mark Diggory is a software engineer on the Virtual Data Center project team.

Gary King is a professor of government in the Department of Government at Harvard University and a founding member of the center for Basic Research in the Social Sciences. He has been a Guggenheim fellow, a visiting fellow at Nuffield College, Oxford University, and is currently a fellow of the American Academy of Arts and Sciences. He has authored and coauthored 50 journal articles and four books in political methodology and other fields of political science. He is currently president of the Society for Political Methodology and serves as director of the Harvard-MIT Data Center.

Akio Sone (Ph.D., Syracuse University) is a postdoctoral fellow at the Harvard-MIT Data Center and a member of the Virtual Data Center project team.

Sidney Verba is a Carl H. Pforzheimer University Professor. He is the author and coauthor of a number of books on American and comparative politics and of articles on those subjects. He is a member of the National Academy of Sciences and a fellow of the American Academy of Arts and Sciences and has been a Guggenheim fellow and a fellow of the Center for Advanced Study in the Behavioral Sciences. His current research interests involve the relationship of political and economic equality, mass and elite political ideologies, and mass political participation. He is also the director of the university library.

Daniel L. Kiskis (Ph.D., University of Michigan) is an assistant research scientist at the University of Michigan and a member of the Virtual Data Center project team.

Michael Krot is a graduate student in the University of Michigan School of Information Science and a member of the Virtual Data Center project team.

Crime Mapping and Its Extension to Social Science Analysis

IRVIN B. VANN
G. DAVID GARSON
North Carolina State University

The purpose of this article is to present the definitions of 21 geographic information system (GIS) functions used in crime mapping and to propose how these functions may be applied more generally to social science research. Most social data are spatial, but this fact has been largely ignored in sociological and social science research. More extensive use of both spatial statistics and spatial analysis in sociology seems likely, and hopefully this article will stimulate social scientist readers to explore using GIS in their research.

Keywords: crime mapping, geographic information systems, social science research methodology

The increased availability of digital data is opening new opportunities for detailed spatial analysis of social behavior, just as increased scrutiny of public expenditures suggests uses of spatial analyses of policy initiatives to target resources where they are most needed (Craglia, Haining, & Wiles, 2000). One policy area in which the use of geographic information systems (GIS) combined with spatial analysis tools has made significant progress is crime analysis. Whereas the uses of crime mapping are interesting in themselves, a typology of crime mapping reveals numerous actual and potential applications to more generic social science problems and topics.

In a recent study conducted on behalf of the North Carolina Governor's Crime Commission, the authors examined the implementation of GIS in five different police jurisdictions in North Carolina. In addition, the authors surveyed the nation's top 12 police departments using GIS in crime analysis. In the course of the study, the authors identified 20 distinct GIS applications of crime mapping, many of which are pertinent to social science research.

THE NATURE AND POTENTIAL OF GIS

GIS is usually defined as a constellation of hardware and software integrating computer graphics with a relational database for the purpose of managing data about geographic locations (Garson & Biggs, 1992). GIS users have benefited from advances made in the computer industry as the market has shifted from mainframe computers to desktop personal computers. This migration has greatly expanded the number of GIS users and the availability of GIS to social science researchers (Vine & Degnan, 1998). Although the graphical output of a GIS is a form of data visualization that supports analysis of multiple attribute layers (DeMers, 1997) and as such is closely aligned with conventional social science analyses of data tables,

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 471-479

most social scientists are unfamiliar with fundamental GIS concepts such as themes, spatial queries, and geodatabase data structures (DeLorenzo, 2000; Raftery, 2000).

GIS analysis involves one of the following four general types of operations: reclassification, overlay, distance and connectivity measurement, and neighborhood characterization.

1. Reclassification operations transform the attribute information of a single theme. Reclassification is seen on maps as color-ramping themes such as population density or income where the color represents a particular range in the database. GIS packages such as ArcView support various types of reclassification, such as reclassification by standard deviations or by natural breaks (setting cutting points that maximize within-groups homogeneity and between-groups heterogeneity).
2. Overlays use GIS layers to intersect the polygons associated with features in one layer with those of another such that one may visualize the intersection of data. For instance, overlays answer questions such as how homicide clusters relate to census tract with high unemployment (DeMers, 1997).
3. Distance and connectivity measurements are used jointly to measure point-to-point distances and the complexity of networks, which may be not only transportation systems but also any phenomenon involving flow. Built-in GIS functions solve routing problems but also service area, response-time, bottleneck, and other generic types of problems.
4. Neighborhood characterizations are functions that classify objects as part of a larger neighborhood of objects based on some shared attribute. An example of this operation might be classification by distance from a critical point or event (DeMers, 1997). In the raster method, features can be located within a grid and associated with attributes of features in grid cells within a researcher-determined radius.

Spatial analysis using a GIS is not synonymous with statistical analysis. The statistical analysis of event locations belongs to the specialized field of spatial statistics. Spatial statistics is a branch of statistics that includes measures of spatial distribution, spatial autocorrelation, and spatial association. Most commercial GIS software implement only the most fundamental of special statistical tools (Levine, 1996), but software extensions, such as the SpaceStats extension developed by Luc Anselm for ArcView, remedy this shortcoming. Employing spatial statistical analysis, for instance, a social researcher may discover a relationship between the racial characteristics of a census tract and the collocation of toxic waste sites (Levine, 1996).

GIS AND SOCIAL SCIENCE

A decade ago, it was evident that with widespread availability of spatial data, mapping was destined to become a far more important social science research tool than it had been in the past (Garson & Biggs, 1992). Today, however, although GIS data are now routinely used by city planners, environmental impact analysts, emergency management teams, policy analysts, market researchers, and governmental agencies, the use of such spatial data is rare in social scientific analysis (Ebaugh, O'Brien, & Saltzman Chafetz, 2000).

One of the main tools for collecting social science data is the opinion survey, a methodology that could serve as a basis for the strong involvement of social scientists in GIS. Surveys often collect spatial variables such as address, zip code, county, and so forth. Events social scientists study are also grouped into boundary areas such as census tracts or city, county, state, or national jurisdictions. In theory, social science data have a rich potential for both spatial analysis and spatial statistical analysis but have been used primarily for simple functions, such as creating choropleth (shaded) maps, for instance, of a variable such as educational attainment for a level of aggregations such as counties (Abbott & Argentati, 1995).

CRIME-MAPPING FUNCTIONS IN RELATION TO SOCIAL SCIENCE ANALYSIS

The authors' investigation of GIS implementation in law enforcement identified 20 functions used in crime mapping. All but one of these functions involve the use of spatial rather than statistical analysis (Garson & Vann, 2001). In the sections that follow, these crime-mapping functions are defined and presented with a focus on the use of such functions in social science research. The authors acknowledge that not every function is amenable to social science uses beyond police operations, but they find surprising versatility in these GIS techniques.

Pin Mapping

Pin mapping is perhaps the most elementary of GIS functions, involving merely the location of a dot or symbol at the coordinates of some feature such as locations of schools, polling places, police stations, or crime incidents. However, the dot or pin does not need to represent just the location of a feature; instead, it may represent a geographic center of concentration of many instances. For example, it may represent not a single crime incident but rather the geographic mean location of a set of like crimes such as auto thefts. Likewise, pins may represent the geometric mean location of Republican versus Democratic voters, male versus female voters, or voters in different age brackets. When pin mapping of this type is combined with time-series data, discussed as follows, the mean spatial locations of the value categories of any discrete variable may be animated over time in a form of data visualization, which illustrates spatial correlation of attributes in any developmental sequence.

Hot-Spot Mapping

In a law enforcement context, hot-spot mapping is the spatial representation of areas with high concentrations of crime. This GIS technique uses spatial clustering methods to draw contour lines around areas of concentration. The concentration contour forms a polygon, which can be used within GIS software to calculate the associated values of the polygon with respect to any or all other attribute layers. Thus, for instance, one might use crime incident data to define a contour hot spot, then use the capabilities of GIS software to easily calculate such correlates as per capita income, land use proportions, street lighting density, or any other variable thought germane. Comparing hot-spot contour polygons and their correlates across urban areas could provide insight into factors associated with crime incidence above a researcher-determined threshold. However, similar analysis might be performed on social science variables outside law enforcement. A researcher might conduct a similar analysis for mental-health-related hot spots, political campaign donation hot spots, or any other incident-based phenomenon.

Mapping Crime Density

Mapping crime density values through shaded contour maps involves interpolating crime levels by using spatial analysis to estimate crime levels between points of observed crime. A dot density map shows location, and through clustering of dots can show density, but its utility as a data visualization tool is hampered due to the problem of overlaying dots at the same location, which obscures the true density. Density surface mapping—a type of contour mapping with shaded contours linking points of approximately equal density—overcomes this

problem. Using a color gradient to display increasing density (e.g., darker reds) makes it is easier to see the true density of any area. Density mapping is appropriate wherever spatial estimates of a geographically continuous phenomena must be made based on a finite number of sampled observation points. For instance, a linguistic anthropologist might use density mapping to chart the concentration of speakers of a certain dialect. By developing several such maps for various dialects and using them as overlays, a visualization can be effected yielding hypotheses about cultural diffusion and cultural hegemony and, moreover, tested by creating such overlay sets for time-series data.

Creating Briefing Maps

Creating briefing maps at first seems like a police-specific function. However, social scientists might think of a briefing map as a tool for communicating with various audiences, including academia or the general public. In an article by Gaddie, Johnson, and Wildgen (1998), the authors powerfully illustrated the potential of briefing maps in an application dealing with the collocation of toxic waste dumps within the predominately Black parishes of Louisiana. The briefing maps combined the previous density types of mapping with one of the GIS overlays. Gaddie et al. first used density maps to illustrate the population density of minorities in Louisiana's parishes. Next, they used the overlay operation to place point coverages or themes of the toxic sites over the population density map, creating a powerful visual representation of the juxtaposition of a specific population with a specific event.

Mapping for Decision Making

In crime mapping, maps at the beat-, sector-, and citywide levels may be used routinely for strategic decision making by law enforcement officials. Likewise in social science, maps may be useful tools for policy decision making. For example, one application used GIS and remote-sensing technologies in redressing land distribution issues in post-Apartheid South Africa (Harris, Weiner, Warner, & Levin, 1995). Similarly, social scientists using a variety of data and remote-sensing techniques may gain a more integrated picture of environmental and urban growth variables. Urban planners routinely use remote-sensing information from LandSat and aerial photography to study the geography of cities. The database management system of GIS is useful for appending related attribute information such as political, environmental, or demographics information. Resulting maps are effective in policy analysis and decision making regarding regional planning issues, for example.

Pattern Detection

Pattern detection maps in crime analysis detect specific, repeating patterns of particular types of crime for purposes of allocating patrol resources. For social scientists, pattern detection maps may be used in relation to any sociological phenomenon characterized by repeating patterns (e.g., incident data on police racial profiling). Although GIS was not used in the American Civil Liberties Union study of racial profiling in the state of Maryland, GIS would be useful for detecting repeating patterns of social problems such as racial profiling. After the problem of racial profiling surfaced in the courts and in the media, several states instituted procedures for collecting information about traffic stops by their highway patrolmen. Making this type of data GIS compliant would only require adding a location variable supplied by an on-board global positioning system. The benefit to the social science researcher would be the ability to look for profiling patterns appearing in particular geographic areas

(Ramirez, McDevitt, & Farrell, 2000) and to relate detected patterns more effectively to various boundaries (e.g., patrol areas), point features (e.g., businesses with high minority employment), or line features (interstate, versus secondary state, versus county roads).

Another similar use involving pattern analysis is using GIS to monitor the fair-lending practices of various banking institutions. In this form of spatial analysis, the researcher can use the functions of GIS to select the parcels affected by one particular lending institution of interest and display the resulting pattern, perhaps in relation to the spatial distribution of minority residences based on census data. These combined operations enable a researcher to visually examine data for suspicious gaps in the lending pattern that may correlate significantly to race, religion, or other sociological variables.

Integrating Interagency Data

As crime mapping matures in an agency, data from other agencies may be integrated into the crime-mapping process (e.g., property ownership from tax databases, demographic information from the Census, street light information from public works, alcohol sales information from the zoning department, and so forth). Likewise, in a sociological study of religious congregations in Huston, Texas, researchers combined GIS overlays of the city with ethnographic data on immigrant congregations to analyze the sociological characteristics contributing to more parish or niche-like configurations in their spatial structure (Ebaugh et al., 2000). Social scientists can benefit too from database aspects of GIS.

Although joining datasets is commonplace in social science, GIS adds the possibility of spatial joins that are difficult outside the GIS context. For instance, one may overlay a polygon layer on top of a feature layer, select features within a polygon (e.g., within a flood plain), and create a new third layer. The GIS uses the x and y coordinates of the feature to match them to the polygon(s) (e.g., census tracts) and create a new merged table containing both the feature information and the polygon information. Rows will be the features (e.g., flood plain areas), but there will now be columns for census tract identification and population associated with these areas. Likewise, the researcher can sum features by tract and join the summary table to the tract table, where rows are tracts, and then divide the sum of features by population to create a new features-per-population variable. Also, one polygon variable may be overlaid on another polygon area such that the boundaries of the first layer are used to subdivide the boundaries of the second area, yielding a merged layer with all polygons formed by the intersection. In this way, the social science researcher can spatially aggregate data by multiple criteria represented by different spatially tagged data layers.

Time-Series Mapping

Time-series maps can be used to detect seasonality or other time-dependent trends. A series of chronological overlays can be compared and contrasted, or even animated, to show, for example, how patterns in crime are or are not related to sequences of vacation days in the public school calendar. A social science analog to time-series mapping is, of course, time-series analysis. Time-series mapping extends time-series analysis through analogous techniques of spatial autocorrelation and spatial regression.

Proximity Mapping

Proximity mapping is the creation of spatial buffer zones around an area of interest such as a school, the address of an offender, or even a crime incident. Often, a particular type of

arrest or crime within a certain distance from a school may carry a heavier penalty or warrant closer attention. All GIS software packages have the ability to establish buffers around spatially defined points of interest. Survey address information can be transformed into proximity variables, which might have explanatory value in social science analysis. Political scientists, for instance, might use voter registration addresses in polling locations to establish proximity variables in a study of the effects of polling location density on voter participation and nonvoting. Social workers, as another example, may be interested in establishing multiple distance zones around an area of interest to study the commuting patterns of low-income or minority workers in relation to a community-provided service or their employment (Queralt & Witte, 1998). Another example of using proximity mapping is to examine the characteristics of a surrounding neighborhood by buffering a point of interest. For example, one might analyze knowledge about a particular incident by buffering the incident and conducting random surveys in the concentric zones. The resulting information can be displayed as a density map highlighting the spread of awareness regarding the point event in question.

Mapping Spatial Displacements

Mapping spatial displacement in crime mapping is tracking the movement of crime patterns over time. For instance, parole release data may show spatial movement of concentrations of parolees by residence over time. Social scientists in the United States using readily available census data can monitor the movements of many aspects of the population using GIS beginning with the 1990 U.S. Census. These data are keyed to the Topologically Integrated Geographic Encoding and Referencing (TIGER) line files, which contain map layers for census geography, street networks, hydrology, railroads, and other man-made features (Vine & Degnan, 1998).

Spatial Tracking

Spatial tracking involves the identification of similar patterns that seem to migrate from one location to another. The authors' study cited an example of similar patterns of car break-ins in two different parts of the town of Cary, North Carolina. The break-ins were the work of one suspect who traveled from his home to his girlfriend's home, breaking into cars in both places (Garson & Vann, 2001). Similarly, social scientists using spatial tracking would look for clusters of behavior that seem to travel from one place to another in conjunction with an individual or group. Spatial tracking is broadly applicable to sociological studies of the diffusion of ideas, technology, and other phenomena. Pattern clusters may be represented as point coverages or themes in a GIS overlay operation.

Spatial Situational Analysis

Situational analysis is the analysis of types of crime by types of setting. For example, crime mapping may focus on analysis of crime in abandoned buildings, analysis of crime in establishments serving liquor, analysis of crime in high schools, or the overlay of setting maps to understand possible interaction of settings. Spatial situational analysis includes target profiling, as in comparing convenience store locations with actual convenience store robberies. Social scientists can generalize the concept of spatial situational analysis by analyzing types of events by types of settings. Types of school violence (e.g., fighting, bullying, and so forth), may be compared to the type of school (e.g., primary, middle, or high school) or other attributes (e.g., charter, private, or public status). Although this can be done statisti-

cally, situational analysis visualized through map overlays, with each context being a layer capable of being turned on and off in the course of data exploration, may be a tool some researchers will find more versatile and generative of data insights.

Spatial Historical Analysis

GIS may involve historical analysis of noncrime events in relation to crime events (e.g., producing spatial profiles of crime related to sporting events for purposes of allocating patrol resources for future scheduled sports events). Similarly, a political scientist or sociologist might study the spatial distribution of survey data for a public policy (e.g., support for the death penalty) in relation to historical data on related events (e.g., murder locations).

Multivariate Pattern Analysis

Multivariate pattern analysis is the comparison of multiple variables and their spatial patterns based on incident reporting with patterns determined by some other database, such as based on dispatch data, warrant data, or crime victimization survey data.

Orthophotographic Mapping

Orthophotography involves integrating aerial photographs as detail backdrops for map projections or overlay operations. Digital orthophotography often adds recognizable context to the visual display of spatial analysis. In crime mapping, overlaying semitransparent hot pots on orthophotographs of neighborhoods enhances the impact of mapping information for citizens, beat officers, and law enforcement decision makers alike. Whereas at one level "just" an aesthetic enhancement to mapping, the effectiveness of orthophotographic backdrops is more than trivial. The potential application of digital orthophotography as a backdrop for social science research is limited only by the imagination of the researcher. For example, a researcher might seek visual impact by using orthophotography to display poverty-stricken areas in conjunction with other data themes overlaid using dot-density or choropleth maps.

Spatial Modeling

Spatially modeling is modeling crime based on poverty and other risk factors using some form of statistical analysis (e.g., regression models, structural equation models, simulation models, or neural network models). Each of these statistical tools may be used in conjunction with GIS to create a predictive tool with visual maps suitable for comparison with actual data. For instance, neural network analysis in conjunction with GIS was successfully used in the early 1990s in Pittsburgh, Pennsylvania, to predict the flare up of drug hot-spot areas. The author of the study compared the forecasting capabilities of neural network analysis with other techniques such as regression, finding the former to be very robust compared with the latter in terms of forecasting migrating drug hot spots (Olligschlaeger, 1997).

OTHER CRIME-MAPPING FUNCTIONS

There are, of course, several crime-mapping functions that relate specifically to police operations, with few applications in mainstream social science. Real-time response mapping involves receiving the address of an incident as an input and then quickly generating a map

output, often as an aid to emergency dispatch. Mapping for resource allocation involves using spatial analysis to create service zones or patrol areas or using network analysis to establish emergency response routes or patrol routes. Mapping recovery operations is a very police-specific function that maps the initial location of property before it was stolen in relation to the location where police officers recovered the property. Tactical crime analysis using suspect databases uses suspect databases to understand the spatial distribution of suspects for a crime category in relation to the distribution of actual crimes in that category. Tactical analysis includes linkage analysis listing known suspects within a given distance from selected incidents, for instance. Mapping applications in community policing serve as an effective way of sharing information with Neighborhood Watch and other community groups.

METHODOLOGICAL LIMITATIONS OF THE FUNCTIONS

Because GIS maps are developed from layers of information, the final product is only as good as the data in the composite layers. Although final maps may look accurate because they contain familiar landmarks and boundaries, a combination of incompatible data layers could result in erroneous attribute information within the various boundaries (Vine & Degnan, 1998). Even when data layers are comparable, there are limits to the interpretations one can make regarding overlaid map layers. Often, researchers want to make individual-level inferences from group-level data. The underlying assumption is that the values within map layers reflect the characteristics of the individuals to whom they apply. This is the classic definition of an ecological fallacy in social science research methods (Vine & Degnan, 1998) and must be avoided in spatial analysis and elsewhere in social science research.

CONCLUSION

Despite these limitations, GIS is an analysis tool successfully used in many other disciplines but appears underused by social scientists. The underuse of GIS in social science research is possibly a function of the nature of research methods training in the social sciences—a training that generally provides no place for spatial analysis. In this article, the authors have attempted to make the case that GIS functions illustrated in criminology, where it is widely used, are directly applicable to more general social science research. Of 20 GIS functions studied, only 5 were found to be focused narrowly on police operations. Even with these, some creative researchers may find insightful social science applications

The methodological limitations of GIS functions are only those common to social science research in general. Data issues such as validity and generalizability are the same using a GIS in statistical analysis. The utility of GIS lies in the researcher first understanding the difference between a statistical and a spatial analysis and then having a framework in which to do the analysis. The crime-mapping functions outlined here are presented as a starting point for social science researchers who are considering incorporating these analytic tools into their research.

REFERENCES

Abbott, L. T., & Argentati, C. D. (1995). GIS: A new component of public services. *Journal of Academic Librarianship, 21*, 251-257.

Craglia, M., Haining, R., & Wiles, P. (2000). A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies*, *37*, 711-730.

DeMers, M. N. (1997). *Fundamentals of geographic information systems* (1st ed.). New York: John Wiley.

Ebaugh, H. R., O'Brien, J., & Saltzman Chafetz, J. (2000). The social ecology of residential patterns and membership in immigrant churches. *Journal for the Scientific Study of Religion*, *39*, 107-117.

Gaddie, R. K., Johnson, R. K., & Wildgen, J. K. (1998). Geographic information systems in social policy formation. In B. L. Neuby (Ed.), *Relevancy of the social sciences in the next millennium* (pp. 104-113). Carrollton: State University of West Georgia.

Garson, G. D., & Biggs, R. S. (1992). *Analytic mapping and geographic databases* (1st ed.). Newbury Park, CA: Sage.

Garson, G. D., & Vann, I. B. (2001). *Geographic information systems for small and medium law enforcement jurisdictions: Strategies and effective practices*. Raleigh: North Carolina Governor's Crime Commission.

Harris, T. M., Weiner, D., Warner, T. A., & Levin, R. (1995). *Pursuing social goals through participatory geographic information systems* (pp. 196-222). New York: Guilford.

Levine, N. (1996). Spatial statistics and GIS. *Journal of the American Planning Association*, *62*, 381-392.

Olligschlaeger, A. M. (1997). Artificial neural networks and crime mapping. In D. Weisburd & T. McEwen (Eds.), *Crime mapping & crime prevention* (pp. 313-348). Monsey, NY: Willow Tree.

Queralt, M., & Witte, A. D. (1998). A map for you? Geographic information systems in the social services. *Social Work*, *43*, 455-470.

Raftery, A. E. (2000). Statistics in sociology, 1950-2000. *Journal of the American Statistical Association*, *95*, 654-662.

Ramirez, D., McDevitt, J., & Farrell, A. (2000). *A resource guide on racial profiling data collection systems*. Washington, DC: U.S. Department of Justice.

Vine, M. F., & Degnan, D. (1998). Geographic information systems. *Journal of Environmental Health*, *61*, 7-17.

Wier, K. R., & Robertson, J. G. (1998). Teaching geographic information systems for social work applications. *Journal of Social Work Education*, *34*, 81-97.

Irvin B. Vann is a doctoral candidate in the public administration program at North Carolina State University.

G. David Garson is a professor of public administration at North Carolina State University.

Excel as a Teaching Platform for Managerial Economics

STEPHEN ERFLE
Dickinson College

Managerial economics is one of the most applied areas of economics. Managerial economics teaches students how to use the tools of economics to make decisions. As a result, this class is most effective if it simulates what students face when working in a business setting. Students should therefore be taught using the same toolkit on which business managers rely. Today, that means using Excel as the platform on which to teach the course. Excel is integrated into the course content not only during computer labs and in homework but also during class. Such integration requires more than a set of spreadsheet exercises to accompany a text; it requires Excel programs that faculty can use to interactively introduce concepts during lectures. This article describes a series of exercises and programs that teach students how to use Excel, and how to use Excel to analyze data needed to make reasoned decisions.

Keywords: Excel, managerial economics, applied microeconomics

One of the most important jobs of managers is to make decisions, and managers make better decisions if they employ the tools of economics in their analysis. Therefore, the goal of a managerial economics course should be to prepare the student to use the tools of economics to make decisions. Those decisions often will be based on at least some empirical information, and that information is no longer typically stored on pieces of paper. Rather, empirical information (data) is typically stored in some sort of spreadsheet, the most common of which is Excel. In my view, managerial economics is most appropriately taught using the same toolkit on which business managers rely. Today, that means using Excel as the platform to teach the course.

Most managerial economics texts assume students have had introductory microeconomics, but they do not require a background in calculus or statistics. Those elements of differential calculus and statistics used are explained in the body of the text. The course described in this article was designed for students who have had introductory microeconomics but who have not necessarily had an exposure to calculus, statistics, or Excel. Classes meet for 3 hours of lecture and 1 hour of computer lab per week. (I have found the 75 minutes twice-a-week format is preferable to the 50 minutes three-times-a-week format due to the complexity of this material.) The lecture room must have some method for projecting computer output onto a video screen. The computer lab should have a faculty machine with projection capabilities as well as one computer per student, although two students per computer would be adequate in all but exam situations. Students require individual access to computers in exam situations, as each exam contains both an in-class and a lab component. As a result, multiple lab sections may be necessary for a single lecture section depending on the computer lab capacity relative to the size of the lecture at your school. Given the nature of the material being covered and the synergy that develops between lab and lecture, the lab is essential to the course.

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 480-486

None of the texts on the market in the mid 1990s were integrated with general-purpose spreadsheet programs commonly used in business.[1] Therefore, I began to develop Excel-based lab assignments that would supplement a text. Subsequently, I began to develop Excel programs for use during lectures as well. I have tied these files to Mansfield's (1999) text, but most files would work well with other texts as most texts devote significant resources to estimation techniques that form the core of these materials. These lab and lecture files are listed in Figure 1 in the order in which they are used in the course.[2] I have also begun work on a more user friendly set of files based on the materials discussed in this article. These will be published as part of the ancillary package accompanying the new edition of that textbook (Allen, Doherty, Weigelt, & Mansfield, in press).[3]

Because students have not necessarily been exposed to calculus, statistics, or Excel, these concepts must be taught from the ground up. The course is broadly divided into three parts. The first examines the concept of maximization, the second provides an introduction to statistical analysis with the goal of understanding basic regression modeling, and the third essentially combines the ideas presented in the first two parts. By the end of the semester, students are able to answer complicated but realistic questions using the tools they have at their disposal. Excel is used throughout, in lab, for homework, and in most classes.

LAB CLASSES

Students are not required to have a background in Excel; therefore, the first three labs provide a necessary introduction to using Excel by attacking successively more complicated present value problems.[4] Students seem quite comfortable with the concept of present value; therefore, labs can focus on how Excel can be used to do comparative static analysis.

I do not use Excel's built-in financial functions during these labs because the focus is on learning Excel. Students quickly learn about entering formulas, relative versus absolute cell referencing, moving between worksheets, dragging cells, and setting up the general architecture of problems so that comparative static analysis is easily performed. The only advanced tool introduced in these introductory labs is Excel's Goal Seek function. The first lab midterm tests basic Excel techniques by requiring students to design a structure to answer a present value problem that is given as a word problem.

The first lab of the middle third of the semester is a statistics lab that provides students with the opportunity to examine statistical data using very real information—their own scores and the scores of previous classes on the first in-class and lab midterm.[5] Students use the Paste function menu to obtain basic statistical functions and apply these concepts to answer questions. I also ask students to predict their lab performance as a function of their in-class performance. This provides a nice introduction to univariate regression, and it allows them to explain what their residual means. Because they are explaining their own behavior, students seem particularly interested in being able to interpret this simple regression.

The next three labs introduce regression modeling. Regression Lab (RLab) 1 estimates a multivariate demand equation. Students must grapple with dependent versus independent variables and learn to interpret coefficient size and significance. RLab 2 provides a cross-sectional analysis of soft drink consumption in the United States. A goal of this lab is to have students focus on residuals to see if patterns emerge and then to model those patterns by creating dummy variables to test for regional differences in consumption. RLab 3 examines airline passenger miles flown over time. This lab provides the opportunity to transpose and manipulate data and to create monthly dummy variables to search for seasonality. A goal of this lab is to examine linear and nonlinear trends and to do simple forecasting.

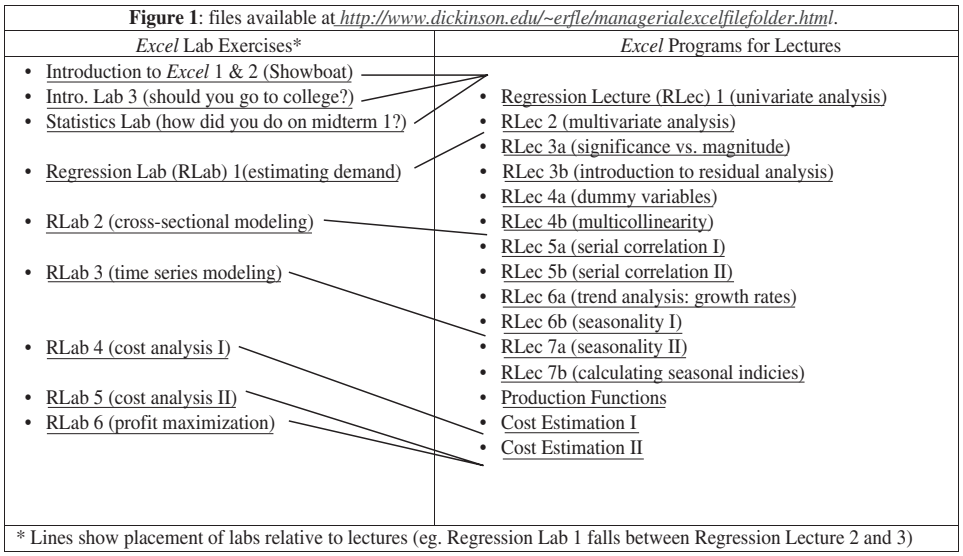

Figure 1.

The final three labs use cost information to examine how managers make decisions to maximize profits. RLab 4 and RLab 5 are both based on simulated data for variable cost of production for two plants of different size. In RLab 4, students use per-unit cost functions for both plants obtained from this information to figure out in which plant they would recommend investing, based on various expectations regarding demand per unit of time. In RLab 5, the reverse question is asked: Suppose you own both plants and wish to produce a given amount of output at minimum cost; how should you distribute production between plants? This lab also examines long-run versus short-run issues. This is the most challenging lab in the semester. Students must first solve a constrained minimization problem with pen and paper and then program their solution into Excel to answer the lab assignment. RLab 6 examines cost analysis in noncompetitive markets.

EXCEL DURING LECTURES

I use Excel sparingly in lectures during the first third of the course, as I want students to focus on derivatives as a concept rather than on how to program derivatives into Excel. Excel becomes a daily part of lectures once we begin to study empirical estimation.

The standard exposition of regression analysis proceeds from univariate to multivariate regression for good reason: Two-dimensional graphs are more readily understood than are higher dimensional graphs. Students are able to visualize how an independent variable x relates to the dependent variable y because they see that some lines fit the data better than others. Regression Lecture (RLec) 1 examines the univariate model by using a small data set Mansfield (1999) used to make his case for the method of least squares.[6] I tell the class to read the material on least squares but to try to just get the big picture rather than get bogged down in the mathematical detail, because Excel takes care of all of those details in a matter of seconds. It is useful to initially show a graph of the data and ask what the best-fit line should look like. One can then run the regression and show that our visual best guess is mirrored by the

results of the regression (both from the line-fit plot and from the estimated coefficients themselves). I hand out a copy of Excel's regression output so that students can take notes directly on the output regarding its various parts. I use this regression to explain the ANOVA part of the regression output that examines the concept of total variation in the dependent variable and how this is broken down into variation explained by the regression and unexplained variation. The coefficients are chosen so that unexplained variation is as small as possible. Finally, I relate these concepts to R^2.

In the next lecture, I move to multiple regression and discuss the interpretation of coefficients for each independent variable as simply slopes in different directions. With two independent variables, we are trying to find the best-fit plane, and Excel can do this just as easily as it found the best-fit line in the univariate case. From here, students seem able to make the leap to how a set of n independent variables $x_1, \ldots, x_n$ relate to the dependent variable y without having to try to visualize the best fitting n-dimensional hyperplane. RLec 2 provides the results of such a multivariate analysis and allows a preview of RLab 1. RLec 2 also allows further discussion of other parts of the regression output, such as the standard error of coefficients and standard error of estimate (unfortunately, both are titled "standard error" in Excel's output, so the difference must be explicitly pointed out to avoid confusion).

Simulations using Excel allow me to examine an array of issues that must be understood to correctly use regression analysis. The simulated data in RLec 3a examines the difference between magnitude and significance. This allows for an extended discussion of what we mean when we say a coefficient is significantly different from zero as well as how we might interpret the 95% confidence interval for the coefficient. I also begin an extended discussion of residual analysis in this lecture with two topics: heteroscedasticity using RLec 3a and nonlinearities using RLec 3b.

In the fourth regression lecture, I describe a variety of situations in which dummy variables can be used to help explain the dependent variable before going to Excel. The example I use in RLec 4a is based on bond sales in the time period surrounding World War II.[7] I also examine multicollinearity during this lecture, a topic I found difficult to explain until I created a simulation model, RLec 4b, which shows the effect of multicollinearity using peanut butter and jelly sandwiches. I posit that students like their "PB&Js" in the following ratios: 1 ounce of peanut butter, 1 ounce of jelly, and two slices of bread. People are not perfect at keeping to this ratio; some people might have a bit too much peanut butter, and others might have a bit too much jelly, but overall, this is what we would expect. Suppose we wish to predict bread consumption on the basis of peanut butter and jelly consumption in the dining hall. If we run a regression of bread consumption as a function of both peanut butter and jelly consumption, we have a multicollinearity problem. The model allows you to introduce varying amounts of noise between peanut butter, jelly, and bread. By varying the amount of error between peanut butter and jelly, one sees the effect of multicollinearity: Peanut butter's and jelly's coefficients and significance levels vary wildly, as you would expect given the multicollinearity problem. Nonetheless, the sum of peanut butter's and jelly's coefficients remains constant at 2 (you need two slices of bread for each sandwich).

The next class, on serial correlation, tends to be the most difficult of the semester. It is imperative that students understand serial correlation, although many of the techniques used to correct for it are beyond the scope of this course. The discussion is made easier to understand using RLec 5a, a model that simulates different degrees of positive or negative serial correlation. I show serial correlation visually using both a standard residual plot and by showing a scatter plot of residuals versus lagged residuals. I calculate a Durbin-Watson statistic from these residuals and work through how to apply the Durbin-Watson test. One of

Excel's shortcomings is that a Durbin-Watson statistic is not automatically calculated among the regression statistics, so one of the tasks that students must learn is how to calculate it from the residuals that are provided within Excel. I show students how to do this calculation using both this simulated data set and RLec 5b, which reexamines the bond data introduced in RLec 4a. RLec 5b is instructive because there is strong positive serial correlation when the dummy variable is not included, and it vanishes upon introduction of the dummy variable. This provides a nice lesson for why serial correlation sometimes exists—specification error that results from omitted variables.

The final two lectures in this section of the course examine linear and nonlinear trend analysis and seasonality. The growth rate simulation, RLec 6a, compares linear, quadratic, and logarithmic estimates based on a constant growth rate data set. The clear patterns that emerge from the residuals in the first two specifications suggest that these models are misspecified. By contrast, the logarithmic model is the correct specification if the data has a constant rate of growth; the lack of pattern from this model's residuals therefore comes as no surprise. The seasonality simulation, RLec 6b, examines three quarterly data sets, the first with linear trend and no seasonality, the second with linear trend and seasonality, and the third with nonlinear trend and seasonality. Inevitably, students seem hesitant to use logarithms, so it is important to explain carefully the differences that arise in interpreting coefficients of time as well as the quarterly dummy variables when using ln(Q) as opposed to Q as the dependent variable. The second lecture on this topic begins with RLec 7a, a file that reexamines the airline lab, RLab 3, because it allows an extended discussion of compounding and seasonality as well how to create seasonal indices from regression output. In RLec 7b, I begin by using seasonal indices for soft drink consumption presented in Mansfield (1999), then show how these could have been obtained from a regression model similar to the airline model, and conclude by using the airline results from RLab 3 to create seasonal indices for the airline industry.

The final part of the course combines the tools learned in the first two parts of the course. The central focus is cost analysis, but production functions and market structures round out the semester's discussion. To understand costs, it is useful to begin by discussing production theory. The Production Functions file examines returns to scale, relative factor intensity, and cost minimization.

Using simulated data based on a variable cost function used early in the semester, I work through the rationale for the standard cubic form for variable cost in Cost Estimation 1. Residual plots for lower order forms exhibit patterns that are eliminated with the cubic form. Similarly, multicollinearity issues may be examined if the form is actually quadratic but is estimated as cubic. The estimated coefficients of the variable cost function form the basis for the per-unit cost curves used for decision making. I recommend a format within Excel for creating per-unit cost curves that minimizes algebraic errors and allows a visual check of the curves in Cost Estimation 2.

Estimated cost functions form the basis for many of the topics examined at the end of the semester. The long-run–short-run distinction is easily put into perspective, as are issues such as when the firm should shut down and profit contribution analysis. Profit maximization is examined using cost and demand information for competitive, monopolistically competitive, and monopolistic market structures, but time constraints preclude discussion of oligopoly markets.

CONCLUSION

The course described omits a number of topics covered in many managerial economics texts. Instead, it focuses on two of the most important tools in the managerial economist's arsenal: constrained optimization and econometric estimation. Because our students are not required to take calculus or statistics, these concepts must be taught from the ground up. Similarly, our students are not required to know Excel, so it must also be taught from scratch. One might question whether the additional burden imposed by learning Excel is worth the effort given the above focus. My sense is that although there is an initial cost imposed in learning Excel, the value added far outweighs the cost. The regression material would be much more difficult to teach were it not for the ease with which Excel allows the student to focus on the various issues involved in running a regression as well as interpreting the results of that regression. Excel acts in an enabling rather than a limiting capacity in this course.

Given the constraint imposed by that lack of mathematical and statistical preparation of students in this class, it is more important to cover these basic topics carefully rather than cover a wider array of topics at a more hurried pace. My sense from the five times I have taught this course is that this trade-off is very reasonable. I covered more pages of the text in my first attempt at this course, but the current course teaches much more by focusing on fewer topics while giving students time to reflect on what they have learned.

If I were teaching managerial economics at a school that had formal calculus and statistics prerequisites, I would pursue a more broad-based strategy. In these circumstances, Excel would remain my choice as an analytical platform due to its prominence and versatility.

NOTES

1. Managerial textbooks have subsequently begun to show up that provide some integration, for example, Edwin Mansfield's (1999) text *Managerial Economics: Theory, Applications, and Cases* has a spreadsheet exercises supplement using Excel.

2. Copies of these Excel programs and labs, together with a class syllabus, student guide, and day-by-day outline, are available from my Web site: `http://www.dickinson.edu/~erfle/managerialexcelfilefolder.html`.

3. The new Excel supplement to the textbook will be composed of 40 files with at least one file for each chapter of the text. Most of the files on my Web site will be included in modified form, but there will be approximately 20 new files as well. These files are being annotated with notes, comments, questions, and answers, and they are being write protected so that students can work directly with these files. Alternatively, faculty members could assign a file as a homework assignment and have the student fill in answers. These files will be released with the textbook early in 2002.

4. The first lab provides a break-even analysis of an investment project discussed in the text, the revival of the Broadway musical *Showboat* (Mansfield, 1999). The second lab extends *Showboat* (Mansfield, 1999) by examining the more realistic case of facing a probability distribution of when shutdown will occur. The third lab provides students with wage and college cost assumptions and asks them to provide their own architecture to answer the question: Is college a good deal on a present-value basis?

5. Labs and lectures refer to Excel files noted in Figure 1 and are available at my Web site: `http://www.dickinson.edu/~erfle/managerialexcelfilefolder.html`.

6. Most other texts have similar sample data sets that are used for the univariate exposition. I find it helpful to use the same data set in class as in the text, as this material is very difficult to grasp given all the summation signs at this point in the text.

7. This example is borrowed from Wonnacott and Wonnacott (1970) because it provides such a crisp example of a situation that requires the use of a dummy variable. When I plot the data and ask if anyone can explain what they see, inevitably, someone will say that this is three separate data sets, one pre–World War II, World War II, and post–World War II. They also see that World War II differs from the other two, so creating a dummy variable called *War* takes care the problem that exists when you regress bond sales as a function of national income.

REFERENCES

Allen, W. B., Doherty, N., Weigelt, K., & Mansfield, E. (in press). *Managerial economics* (5th ed.). New York: Norton.

Mansfield, E. (1999). *Managerial economics: Theory, applications, and cases* (4th ed.). New York: Norton.

Wonnacott, R. J., & Wonnacott, T. H. (1970). *Econometrics*. New York: John Wiley.

Stephen Erfle is an associate professor and the chair of the Department of International Studies, Business & Management at Dickinson College in Carlisle, Pennsylvania.

NEWS AND NOTES

G. DAVID GARSON
North Carolina State University

Call for contributions and reviewers: the *Social Science Computer Review*. Please contact the appropriate associate editor listed on the inside back cover or write to the editor, G. David Garson, NCSU Box 8102, Raleigh, NC 27695; (919) 515-3067. Regarding book reviews, contact the book review editor, Carl Grafton, Department of Political Science and Public Administration, Auburn University at Montgomery, Montgomery, AL 36193; grafton@tango.aum.edu.

All prices cited are list prices at time of review. Street prices and introductory offers may be available for substantially less: Check with manufacturer and vendors.

The World Wide Web address for the *Social Science Computer Review* is http://hcl.chass.ncsu.edu/sscore/sscore.htm.

SOCIAL SCIENCE COMPUTING ASSOCIATION

The Social Science Computing Association (SSCA) is the professional organization associated with the *Social Science Computer Review*. The SSCA sponsors an annual conference, most recently CSS 2000, the Web site for which is http://www2.chass.ncsu.edu/css2000/. The next conference is being coordinated by Jeffrey Roy, who welcomes inquiries at roy@admin.uottawa.ca.

ANTHROPOLOGY AND ARCHAEOLOGY

Archives of European Archaeology (AREA) is a Web site located at http://www.inha.fr/area-archives/. AREA seeks to be "a research network dedicated to the history of archaeology, with particular emphasis on the archives of the discipline, their promotion and preservation."

The California Academy of Sciences, Department of Anthropology, has placed its entire collection database of more than 18,000 records and 6,000 digital images online; it is updated monthly. The URL is http://www.calacademy.org/research/anthropology. For further information, contact Dinah Crawford, Curatorial Assistant, dcrawford@calacademy.org.

DATA RESOURCES

Ameristat is located at http://www.ameristat.org/. AmeriStat is developed by the Population Reference Bureau in partnership with demographer Bill Frey and a team of his colleagues from the University of Michigan and the State University of New York at Albany. It is a one-stop Web site for graphic and text summaries of U. S. demographic characteristics.

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 487-494

Bartleby.com has relaunched its electronic version of the Columbia Encyclopedia at http://www.bartleby.com/65/. Searchable by keyword, this sixth edition has numerous updates, including improved search and navigation features. Visitors can browse the close to 51,000 entries, including 17,000 biographies, with more than 80,000 hypertext cross-references and links to other resources, alphabetically or by one of the 140 categories.

Boxmind's online library at http://www.boxmind.com/ covers 27 subject areas.

Census data. The Bureau of the Census has released *Population Change and Distribution: 1990-2000* at http://www.census.gov/population/cen2000/c2kbr01-2.pdf.

The Civic Involvement Survey, 1997—American Religion Data Archive is available in SPSS and other formats at http://www.TheARDA.com/archive/CIVIC.html. A codebook is also available at the site.

The Social Science Data Analysis Network (SSDAN) is located at http://www.ssdan.net/.

ECONOMICS

EcoSim is located at http://ecedweb.unomaha.edu/ecosim.htm. EcoSim is an interactive simulation in which each student plays the role of a company and a group of people or a legislator in the government and/or central bank. EcoSim establishes among the players the same incentives and general restrictions that exist in the real world. As the simulation progresses, players create the same market forces within EcoSim that exist in the real world. Players determine all market prices and interest rates, production and consumption levels, and so forth. EcoSim operates on a VAX/VMS system (a Windows-based version for a personal computer network is in the works). The software is available at no charge to all interested colleges and universities. All players are requested to purchase a manual and spreadsheet disk (the spreadsheet aids in decision making in the simulation) at a nominal fee. Bookstores can be supplied directly. To obtain a copy of EcoSim, contact Anthony Davies, adavies@altavista.net, A.J. Palumbo School of Business, Duquesne University, Pittsburgh, PA 15282.

The Principles of Economics Web site at http://www.mhhe.com/economics/frankbernanke compliments the textbook by Robert H. Frank and Ben S. Bernanke. Included at the site are "learning sessions," interactive graphs, online testing, current news, exercises, and video.

SIMUL8 is business simulation software for solving queuing and other time-dependent business process decisions through graphical simulation modeling. Cost is $995 from SIMUL8 Corporation. For information, e-mail info@SIMUL8.com. Student versions are free with university site license to the full version. SIMUL8 Corporation, 2214 Rock Hill Road, Suite 501, Herndon, VA 20170; tel. 800-547-6024; fax 800-547-6389.

The South Asia Network of Economic Research Institutes is at http://www.saneinetwork.org/index.asp.

SWIEE is a Swarm Web interface for experimental economics, with a page at http://swiee.econ.unito.it/project.htm. SWIEE is a normal Swarm simulation, with little changes to manage destroy notification, synchronization of time steps between virtual agents and human players (timer class), and RMI communication tunnel (Remote* classes). Simulators can prepare a daily (or weekly, or every 2 days, and so forth) time step or can use a Boolean variable to make a step only when all users chose their behavior. A sample prisoner's dilemma game is provided.

Urban Economics: Simulation Study is a Web site at http://vega.icu.ac.kr/~yskwon/urban/urbanh.htm. It provides a simulation study program for a monocentric city where students can alter parameter values and observe how characteristics of cities change in response. For example, students can see how rental rates for land change by altering commuting cost, income, and so forth. The site also provides simulation programs for an open urban model and a closed urban model based on the assumptions that households in a city have the same income and preference. Originally, all programs at the site were based on John Yinger's Excel programs.

EDUCATION

The Encyclopaedia of Philosophy of Education is located at http://www.educacao.pro.br/.

National Education is a *The New York Times* site at http://www.nytimes.com/library/national/040801ed-index.html. This site focuses on education news and *The New York Times* coverage of education, including book reviews, issue analyses, and related links.

HISTORY

The digital classroom of the National Archives and Records Administration is located at http://www.nara.gov/education/. It provides reproducible primary documents, educational units correlated to national academic standards, and cross-curricular connections.

The European Integration History Index is located at http://www.iue.it/LIB/SISSCO/VL/hist-eur-integration/Index.html.

The History Project is at http://historyproject.ucdavis.edu/. It provides Grade K through 12 history lesson plans on California, United States, and world history.

The *Journal of American History* has created an instructional site at http://www.indiana.edu/~jah/teaching/index.shtml. Digital resources focus on classroom teaching, with primary documents, suggested further reading, and related sites.

Peopling North America: Population Movements & Migration, at http://www.ucalgary.ca/HIST/tutor/migrations/, is an educational Web site from the Applied History Research Group at the University of Calgary. The site presents "an historical overview of migratory movements to and within Canada, the United States, Mexico, and the Caribbean from Europe, Asia, and Africa." The site spans from the archaeological evidence concerning the first Americans to the changes in migration after World War II.

LAW

The *Guide to Law Online* is at http://www.loc.gov/law/guide/index.html.

POLITICAL SCIENCE

The British Labour Party platform for the Blair administration, titled "Ambitions for Britain–Labour's General Election Manifesto," is found at http://www.labour.Org.uk/lp/new/labour/labour.wwv_main.main?p_full=1&p_language=us&p_cornerid=364783.

Civil liberties. The story of the 1947 House on Un-American Activities Committee (HUAC) hearings, which targeted nine Hollywood screenwriters and one director as Communists, is found at http://www.hollywood10.com. In addition, a brief biographical sketch and a filmography are given for most figures.

Critique is a semiannual journal for student writings on politics from Illinois State University, with a Web site at http://www.lilt.ilstu.edu/critique. For more information, contact Susan Craig, Editor, at sucraig@ilstu.edu.

Foreign policy. The 40th anniversary of the Bay of Pigs has lead to the release of numerous documents related to the ill-fated expedition. Hundreds of documents have been placed online by the CIA at http://www.foia.ucia.gov/popdocs/bayofpigs.htm. There is also an online database "containing the full text of English translations of speeches, interviews, and press conferences by Fidel Castro, based upon the public domain records of the Foreign Broadcast Information Service (FBIS)" at http://www.lanic.utexas.edu/la/cb/cuba/castro.html.

Interest groups. At the Chemical Industry Archives at http://www.chemicalindustryarchives.org/, there is a large selection of the internal industrial documents revealing a campaign to limit regulation of toxic chemicals. Created by the Environmental Working Group, the database is searchable by keyword. This site is a companion site to the PBS program on the subject, which is located at http://www.pbs.org/tradesecrets/.

Internet voting. Sponsored by the National Science Foundation, the Report of the National Workshop on Internet Voting: Issues and Research can be found at http://www.netvoting.org/Resources/E-Voting%20Report_3_05.pdf. Conducted by the Internet Policy Institute, this report concludes that online "voting from home" is currently not feasible. The report finds that "remote Internet voting systems pose significant risk to the integrity of the voting process and should not be fielded for use in public elections until substantial technical and social science issues are addressed." However, the report does recommend that trials be conducted at poll sites to determine the feasibility, reliability, and accuracy of Internet voting systems.

National Security Action Memoranda of the John F. Kennedy administration have been posted to the Web by the John F. Kennedy Library and Museum at http://www.cs.umb.edu/jfklibrary/nsam.htm. Topics cover memos from John F. Kennedy, National Security Advisor McGeorge Bundy, and others on Cuba, Vietnam, South Africa, NATO, and Russia. Some portions have been censored.

Researching U.S. Treaties and Agreements is a Web site located at http://www.llrx.com/features/ustreaty.htm#international, sponsored by the Law Library Research X-Change (LLRX). The document at the site, written by Marci Hoffman (a law librarian at the E. B. Williams Law Library) discusses research and access issues pertaining to the study of treaties and agreements, including print and online publications, research guides, online full-text sources, indexes, and more.

U.S. News Classroom is at http://www.usnewsclassroom.com/.

PSYCHOLOGY

Great Ideas in Personality is a Web site at http://www.personalityresearch.org/. Hosted by G. Scott Acton of Northwestern University, it provides overview essays, paper listings, researcher contacts, related links, and more related to personality theory.

MiniPsychdiagnoser is software for diagnosis of mental health disorders. Patient responses to given probes branch to appropriate follow-on questions. There is automated scoring of responses. Cost is $250 from Tall Tree Software at http://www.psychdiagnoser.com; phone: (714) 773-0301. There is a more extensive version, PsychDiagnoser, for $500, also covering family history, developmental history, and rug response history.

Psychological Research on the Net is a Web site of the American Psychological Society, located at http://psych.hanover.edu/APS/exponnet.html. It provides an annotated list of psychological research currently being conducted on the Web.

The Science of Emotions Web site at http://www.news.wisc.edu/packages/emotion/ is from the Health Emotions Research Institute at the University of Wisconsin–Madison School of Medicine. Topics cover the connections between brain chemistry and human emotional experience, including research project overviews, researcher biographies, and related news.

PUBLIC ADMINISTRATION

The 2001 annual report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds is available at http://www.ssa.gov/OACT/TR/TR01/index.html. The current status of the Social Security and Medicare trust fund is discussed.

The University at Albany's Center for Technology in Government has a Web-based "Insider's Guide to Using Information in Government," with topical essays on strategy, policy, data, skills, cost, and technology along with eight case studies of state and local government projects and links. The site is at http://www.ctg.albany.edu/guides/usinginfo/. A printed eight-page overview is also available. To obtain a copy, send a request to info@ctg.albany.edu or contact Sharon S. Dawes, Director, Center for Technology in Government, sdawes@ctg.albany.edu; http://www.ctg.albany.edu.

SIMULATION

Otterville is a free online community development simulation based on Stella, a simulation language which has been widely used in economics and other social sciences. The Web site is at http://www2.hps-inc.com/otterville/. What are the causes of unemployment in an urban area? Can a city "grow its way" out of urban problems? Is it possible for an urban area to have it all—low unemployment, vibrant businesses and services, and high quality of life for all its citizens? Students try their hand at being a mayor, setting policies that will affect the evolution of the community, and testing a range of urban policies via simulation. Stella itself is now in version 7, featuring a better user interface and easier to interpret output. The uniform resource locator is http://www.hps-inc.com/Education/new_Stella.htm. Cost is $299 for the basic version and $549 for the research version. Demo versions available. Also available is NetSim Creator, a product that allows Stella simulations to be deployed interactively over the Web. The textbook *Stella Introduction to Systems Thinking* (cost $30) is also available. For more information, contact High Performance Systems, sales@hps-inc.com; phone: 800-332-1202.

Pasion, whose Web site is at http://www.raczynski.com/pn/pn.htm, is simulation software for discrete events, continuous models, animation, bond graphs, signal flow graphs, queuing models, and more. Cost is $70. A demo version and low-cost student ver-

sions are available. Contact: Stanislaw Raczynski, P.O. Box 22-783, 14000 Mexico D.F. Mexico; stanracz@netservice.com.mx.

Simulation software for the social sciences is linked at http://www.uni-koblenz.de/~kgt/Software.html.

Edmund Chattoe, Department of Sociology, University of Oxford, provides this bibliography of simulations of innovation diffusion, which includes social networks. Additional bibliography is at http://www.soc.surrey.ac.uk/~scs1ec/innovate.html.

Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization Science*, *8*, 289-309.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, *4*, 203-226. [Simulation of cultural exchange on a lattice where sites interact on the basis of probabilities with nearby sites with similar culture]

Buskens, V. W. (1999). *Social networks and trust* [Online] (Available: http://www.fss.uu.nl/soc/iscore/staff/buskens.htm

Buskens, V. W., & Yamaguchi, K. (1999). A new model for information diffusion in heterogeneous social networks. In M. Becker & M. Sobel (Eds.), *Sociological methodology 1999* (pp. 281-325). Oxford, UK: Basil Blackwell. Available: http://www.fss.uu.nl/soc/iscore/staff/buskens.htm-paper070.pdf

Carley, K. M. (1991). Growing up: The development and acquisition of social knowledge. In J. A. Howard & P. L. Callero (Eds.), *The self-society dynamic: Cognition, emotion, and action* (pp. 75-102). Cambridge, UK: Cambridge University Press. [The CONSTRUCT software package that has been used for innovation diffusion studies]

Carley, K. M. (1991). A theory of group stability. *American Sociological Review*, *56*, 331-354. Available: http://www.jstor.ac.uk

Castlefranci, C. (2001). Towards a cognitive memetics: Socio-cognitive mechanisms for memes selection and spreading. *Journal of Memetics—Evolutionary Models of Information Transmission* [Online], *5*. Available: http://www.cpm.mmu.ac.uk/jom-emit/2001/vol5/castelfranchi_c.html

Chattoe, E. & Gilbert, N. (1997, September 8-11). *Modelling the adoption of agri-environmental measures as an innovation diffusion process* [Online]. Paper presented at the IMAGES meeting (M1), Dolomieu, France. Available: http://www.soc.surrey.ac.uk/~scs1ec/publications.html

Chattoe, E., & Gilbert, N. (1998, March 16-18). A basic simulation of information diffusion [On-line]. Paper presented at the IMAGES meeting, Saint-Sauves D'Auvergne (M6). Available: http://www.soc.surrey.ac.uk/~scs1ec/publications.html

Chattoe, E., & Gilbert, N. (1998, May 25-29). *Modelling the adoption of agri-environmental measures using decision plan nets* [Online]. Paper presented at the IMAGES meeting, Clermont-Ferrand (M2). Available: http://www.soc.surrey.ac.uk/~scs1ec/publications.html

Chattoe, E., & Gilbert, N. (1998, October 19-20). *A simulation specification for innovation diffusion through social networks with boundedly rational agents* [Online]. Paper presented at the IMAGES meeting, Paris. Available: http://www.soc.surrey.ac.uk/~scs1ec/publications.html

Chattoe, E., & Gilbert, N. (1999, April 8-10). *Specification of University of Surrey simulation Version 1.1 and proposed developments for Version 1.2* [Online]. Paper presented at the IMAGES meeting, Miramare (M3). Available: http://www.soc.surrey.ac.uk/~scs1ec/publications.html

Contractor, N. S., & Grant, S. J. (1996) *The emergence of shared interpretations in organisations*. In J. Watt, & A. Van Lear (Eds.), *Dynamic patterns in communication processes* (pp. 215-230). Thousand Oaks, CA: Sage.

Hedström, P. (1994). Contagious collectivities: On the spatial diffusion of Swedish trade unions, 1890-1940. *American Journal of Sociology*, *99*(5), 1157-1179.

Hedström, P., Sandell, R, & Stern, C. Mesolevel networks and the diffusion of social movements: The case of the swedish social democratic party. *American Journal of Sociology*, *106*, 145-172.

Jager, W., & Janssen, M. A. (2001, June 26-29). *Multi agent simulation of the behavioural dynamics behind the innovation diffusion of "green" products*. Paper presented at Eurosim 2001: Shaping the Future with Simulation, the 4th International EUROSIM Congress, Delft, the Netherlands. Available: http://www.bdk.rug.nl/medewerkers/w.jager/

Krackhardt, D. (1997). Organizational viscosity and the diffusion of controversial innovations. *Journal of Mathematical Sociology*, *22*, 177-200. [Contrasted with Carley's CONSTRUCT]

Plouraboue, F., Steyer, A., & Zimmermann, J. B. (1998). Learning Induced criticality in consumers adoption pattern: A neural network approach. *Economic Innovation and New Technology*, *6*, 73-90.

Steyer, A., Zimmermann, J. B. (1997). *On the frontier: Structural effects in a diffusion model based on influence matrixes* (GREQAM Working Paper Series No. 97A14). Abstract available: http://netec.wustl.edu/BibEc/data/Papers/fthaixmeq97A14.html

Strang, D., & Macy, M. W. (in press). In search of excellence: Fads, success stories, and adaptive emulation. *American Journal of Sociology.* Available: http://www.people.cornell.edu/pages/mwm14/

Valente, T. W. (1995). *Network models of the diffusion of innovations* (Hampton Press Communication Series). Cresskill, NJ: Hampton.

SOCIAL WORK

IMPACT–Increasing the Impact of Assistive Technology is instructional software on using technology in health and human services professions, with a focus on older service users and possible causes of disability. The software is funded by the European Commission and available from Fontys University, the Netherlands at http://www.fontys.nl/impact/.

SOCIOLOGY

Criminology. The Department of Justice has released two crime data sets on the Web: "Violent Victimization and Race, 1993-98" is at http://www.ojp.usdoj.gov/bjs/abstract/vvr98.htm and "Contacts Between Police and the Public: Findings From the 1999 National Survey" is at http://www.ojp.usdoj.gov/bjs/abstract/cpp99.htm. Each is available in pdf and ASCII formats with zip spreadsheets.

Social, Economic and Political Change is a Web site at http://sites.netscape.net/gsociology, developed by Gene Shackman, with links to sites with data, theory, research, course syllabi, and other useful information including the site author's review of theories of change.

The Timeline of Sociology is located at http://www.ac.wwu.edu/~stephan/timeline.html. The site is published by Ed Stephan, a Western Washington University professor emeritus, and covers the history of sociology from 1600 through 1995, featuring many links to the works and biography of noted sociologists and related researchers.

Women's history. An annotated directory of Web sites devoted to women's labor history is located at http://www.afscme.org/otherlnk/whlinks.htm to honor Women's History Month. Sponsored by the American Federation of State, County, and Municipal Employees (AFSCME), it includes a number of sites on famous women agitators and labor advocates as well as historical sites dedicated to key periods in women's labor history and a section of general women's labor history links.

Women's history. Documents from the women's liberation movement of the 1960s and 1970s are available from the Duke University special collections department at http://scriptorium.lib.duke.edu/wlm/.

Women's history. The papers of Elizabeth Cady Stanton and Susan B. Anthony are now available at http://adh.csd.sc.edu/sa/sa-table.html. The site covers letters, pamphlets, reform organization documents, leaflets, and more. There are some 14,000 images. The site has keyword searching.

STATISTICS AND RESEARCH METHODOLOGY

Adaptive Conjoint Analysis for the Web (ACA/Web) is new software from Sawtooth Solutions, 530 West fir St., Sequim, WA 98392; phone: (360) 681-2300; fax: (360) 681-2400; http://www.sawtoothsoftware.com; info@sawtoothsoftware.com. Also available is CiW, Sawtooth's general survey administration tool for online surveys. It has advanced skip logic, randomization of items, and other complex options. No programming is needed. Cost for ACA/Web is $3,000 for the 10-attribute version, $6,000 for the 30-attribute version. Cost for CiW ranges from $3,000 for 50-item surveys to $9,000 for 500 item surveys.

Gene Shackman has a sociologically oriented research methods site at http://sites.netscape.net/gsociology/methods, listing links to free resources for methods in evaluation and social research. It lists links to manual, articles, and papers, and to resources on how to do surveys, interviews, observations, and some statistics and software and how to display statistics.

The Internet Rogator Survey Tool can be found at http://www.internet-rogator.com. Here, by using this multifunctional instrument, users can build their own interactive online surveys, market research, and opinion polls. The system is programmed in UNIX and controlled through hypertext markup language. A free demo version is available.

BOOK REVIEW

Intellectual Property in the Information Age: The Politics of Expanding Ownership Rights. Debora J. Halbert. Westport, CT: Quorum Books, 1999. 208pp., $59.95 (cloth)

For a brief period in the mid-1960s, conservative, liberal, and radical social scientists imagined a "postscarcity" future. Extending the then-current decrease in poverty, expansion of the middle class, and growth in industrial production into the not-very-distant future promised a humanity freed from material need. Daniel Bell's (1973) vision of a postindustrial society a decade later retreated slightly from this utopian possibility, but nevertheless, it suggested a withering away of the conflict between capital and labor in favor of a less unequal and more meritocratic social order based on educational resources. Perhaps with a gradual waning of optimism, this orthodoxy has been repeated by others too numerous to mention. Robert Reich (1991), in the *The Work of Nations*, for example, argued that in a global economy, "symbolic workers" become key to economic well-being such that wealth now follows the educated.

The promise of the "information age" has not been realized. The data are in, and the last quarter century has been characterized by growing inequality. Although educational "resources" have arguably grown more important, the power of capital has not withered. Many have accounted for this inequality by arguing that in a global system, the advantages of uneven development are accrued by capital, but perhaps even more important, changing political and legal definitions of information ownership have created new techniques for accumulation.

As the information age has shifted from being a projective description of an immediate future to a present reality, so too analysis has shifted from general claims to detailed descriptions. A fully developed political and economic analysis of information control and proprietary practice has not yet matured, but a spate of recent work on intellectual property has moved analysis beyond broad claims and false promises.

Debora J. Halbert's *Intellectual Property in the Information Age: The Politics of Expanding Ownership Rights* is a welcome addition to this literature. As intellectual property has become more commodified, efforts to establish ownership have been aggressively pursued in a variety of political, social, and legal sites. Such efforts, Halbert argues, not only have implications for how resources are distributed but also create a corresponding cultural change. In particular, new heroes, villains, and notions of legitimate practice are constructed to define and enforce the new property boundaries.

Her work centers on a conceptual tension suggested by Michael Shapiro (1991) between sovereignty systems and exchange systems. Sovereignty systems emphasize ownership and maintenance of control; exchange systems emphasize reciprocity and the relaxation of control to promote circulation. Copyright battles, she argues, represent a struggle to define "boundaries" around cultural goods by fixing respective rights and privileges along the continuum between sovereignty and exchange. For most intellectual production, expressions and exchange are the source of value. Moreover, democracy, political liberalism, public education, and science all depend on the free circulation of ideas. Copyrights, however, may

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 495-498

constrain reciprocity by limiting access, maintaining control, and strengthening property rights.

Halbert argues that new technologies, especially the Internet, encourage free and open discussion and exchange. Many authors find greater social value through circulation rather than ownership, and Halbert argues for the benefits of noncommodified authorship. At the same time, this increase in circulation disrupts traditional forms of ownership and control and has yielded a countervailing force that advocates for stricter protection. Such advocacy often is presented not by directly emphasizing the virtues of ownership but by asserting the need to create incentives and reward creative, intellectual work.

Halbert does not assume a neutral political stance in this debate. She identifies some of the hidden interests and describes some of the unrecognized social, political, and economic costs associated with an enhanced sovereignty system. To accomplish this, she analyzes the narratives that emerge in the context of political and legal struggles. Narratives are embedded in a normative process that allows citizens to evaluate practice, and thus, she argues, new notions of right and wrong are being actively constructed to undermine the greater public good.

The book begins with a discussion of the historical origins of copyright and some brief comparisons between the French, English, German, and American legal cases. She argues that booksellers and publishing houses, especially in the later part of the 18th century (the roots of copyright law extend earlier), promoted copyright law to protect their economic interests, yet they did so by emphasizing the need to offer incentives for authors. The missing piece in this story is that the initial commodification of intellectual work resulted in a transference of property rights from the author to the publisher.

This theme is developed in a later, more theoretical chapter on the nature of authorship that borrows from Foucault and cultural studies. She maintains that the traditional copyright narrative yielded the author as sovereign, as source, as genius, and as owner of the idea and the text. This notion of authorship is enhanced further by a narrative that heightens the contrast between a romanticized depiction of the author and alienated labor and between the unity and wholeness of artistic production and the fragmentation in the industrial division of labor. The creation of the author as sovereign is itself a mythic and often illusory creation of the traditional copyright narrative, which is itself deeply embedded in the formation of the modern, bourgeois self. By contrast, Halbert asserts that intellectual production is always a collaboration that occurs as people enter into an ongoing discourse within an established institutional context.

In the empirical chapters, Halbert focuses on the political and legal narratives that accompanied recent initiatives and changes in the copyright law. The ease with which software can be copied raises unique difficulties in maintaining sovereignty, yet as Halbert demonstrates, the courts have relied on traditional notions of authorship to view software as literary creations. The subsequent rulings not only limit the legal rights of consumers to use, copy, and share the program but also prevent other software developers from using lines of code. Such rulings, she maintains, have resulted in a continuous erosion of the value of free exchange of ideas and has amplified the idea of information ownership. She argues that the articulation of assumptions and provisions in several recent government policy papers concerning intellectual property have amounted to a "virtual land grab," leading to a more pervasive control over ideas and expressions and a retreat from sharing and free exchange.

The discussion of these legal and political narratives is coupled with a description of how "the notion of the proprietary author developed on the backs of pirates." That is, legal and political claims to ownership require a compelling story that specifically identifies how one's rights or privileges are damaged by the immoral or illegal action of others. Software companies, the film industry, and especially Disney have been particularly vocal concerning international "pirates" who illegally copy their work and thereby deprive them of revenue (i.e., the incentive for the creative artists). Domestically, "hackers" perform much the same function by violating standards of information ownership and control.

Halbert concludes with a discussion of the future of intellectual property law in which she pessimistically forecasts further expansions of copyright protections, significant increases in "cybercops" to patrol cyberspace in search of pirates and hackers, and a greater erosion of human freedom to facilitate the accumulation of private profit. Against this pessimism, she asserts a resistant possibility in which information sharing leads to an alternative narrative that effectively counters the domination of significant cultural symbols by a remarkably small set of giant producers. Throughout the text, however, the strength of her argument rests on the imaginative reconstructions of the concealed interests behind an emergent normative reality that has already granted extensive rights to information owners behind the backs of an unsuspecting public. The efforts at resistance are only occasionally addressed and treated as peripheral to the making of new norms, so that by the end of the book, the alternative narrative seems more like a faint hope.

Although Halbert's work makes broad claims about intellectual property, the empirical materials focus exclusively on issues of copyright. Patents, trade secrets, and trademarks establish broader propriety rights than copyright. Halbert acknowledges this limitation, but a wider net would come closer to meeting the challenge implied by the title. More important, much of the public narrative regarding copyright is not included in her case materials. For example, absent from her discussions are the MP3 and Napster controversies, the sharing of academic materials and the use of the photocopier, the efforts to establish copyright protection over DNA sequences, the evolution of rulings that followed the introduction of the VCR, or the development of new technologies to enforce copyright protections (i.e. CDs, DVDs, scramblers, and terminal seeds). Upon completion of her book, the reader does not emerge with a well-stocked set of examples and a comprehensive understanding of the recent evolution in copyright law. Instead, examples primarily focus on computer software that provide anecdotal evidence for the conceptual distinctions, but this emphasis comes at the expense of a thick description of current legal practice and contemporary normative expectations regarding copyright infringements. The protection of intellectual property rights should not be limited to a discussion of the interests of media industries and software companies, because such protections are rapidly becoming the legal foundation that support new forms of capital accumulation in agribusiness, biotechnology, pharmaceuticals, financial services, advertising and marketing, and all those (still surviving) dot.coms.

Modern liberalism is based on the creation of a public sphere that permitted the free exchange of ideas. Is that space being closed now not by religious elites seeking a monopoly on truth but by corporate interests that want to get paid? Marxism is grounded in the relations of production because it was at the point of production that value was produced and appropriated. In an economy that rests on intellectual production, initial designs may be labor intensive but use values can be infinitely reproduced with remarkably little labor. Is the protection of intellectual property nothing more than a political construction of scarcity?

If the information age is something quite different than modern industrial capitalism, then a fully developed political economic analysis still needs to be produced. This work by Halbert provides an important contribution to that project. It is well written and accessible for advanced undergraduates and offers interesting conceptual discussions to open up the area of intellectual property to further inquiry.

Stephen Adair
Central Connecticut State University

REFERENCES

Bell, D. (1973). *The Coming of Post-Industrial Society: A Venture in Social Forecasting.* New York: Basic Books.
Reich, R. (1991). *The Work of Nations.* New York: Vintage Books.
Shapiro, M. (1991). Sovereignty and exchange in the orders of modernity. *Alternatives* 16, 447-477.

INDEX

to

SOCIAL SCIENCE COMPUTER REVIEW

Volume 19

Number 1 (Spring 2001), pp. 1-128
Number 2 (Summer 2001), pp. 129-240
Number 3 (Fall 2001), pp. 241-392
Number 4 (Winter 2001), pp. 393-504

Authors:

Social Science Computer Review, Vol. 19 No. 4, Winter 2001 499-503

Articles:

Reports and Communications:

Software Reviews:

Book Reviews: