2nd Pre-ICIS Annual Workshop on

HCI Research in MIS

(HCI/MIS '03)

Organized by AIS SIGHCI
http://melody.syr.edu/hci

PROCEEDINGS

http://melody.syr.edu/hci/pre_icis03_wksp

December 12-13, 2003, Seattle, Washington

Workshop Co-Chairs
Ping Zhang
Jonathan Lazar
Scott McCoy

Program Chair
Fiona Fui-Hoon Nah

Local Organizing Committee
Traci Hess
Jeff Kim

Sponsored by

SYRACUSE UNIVERSITY
School of
Information Studies

PROCEEDINGS TABLE OF CONTENTS

* Best paper nominees

Gender and Personality in Media Rich Interfaces: Do Birds of a Feather Flock Together?

Traci J. Hess
Washington State University
thess@cbe.wsu.edu

Mark A. Fuller
Washington State University
mark@wsu.edu

John Mathew
Washington State University
jmathew@cbe.wsu.edu

ABSTRACT

This research explores how user and interface characteristics can interact to influence decision performance. Specifically, this research examines the effects of gender, personality similarity, and increased levels of information cues on user involvement with a computer-based decision aid. In addition, this research explores the downstream effects of user involvement on decision time, effort, satisfaction, confidence, and quality. Findings indicate that gender has a significant influence on user involvement, and that involvement and the level of information cues provided by the decision aid have a direct influence on decision performance.

Keywords

Gender, Involvement, Decision-Making, Decision Performance, Personality, Information Cues

INTRODUCTION

Advancements in computing technology, particularly in the area of interface design, have provided end-users with rich, highly interactive online environments. Various forms of multimedia (e.g., voice, animated graphics) have been leveraged to create innovative software applications that facilitate user interaction, as evidenced by the emergence of animated interface agents in both professsional and commercial environments. Yet research into the underlying behavioral impacts of these rich interaction environments has not kept pace with these advancements in interface design. The effect of these richer, more engaging interfaces on user behavior and decision-making performance warrants more focused exploration.

Understanding how humans and computers interact has been the subject of research from a variety of perspectives. Early research on the presentation of information focused on how information was represented to the user, and how a particular representation 'fit' with the requirements of the task (Vessey, 1991; Vessey and Galletta, 1991). This research was extended into the task technology fit (TTF) framework that not only considered characteristics of the task and the technology, but also incorporated characteristics of the user (Goodhue, 1995; Goodhue and Thompson, 1995). This model inherently implies that user characteristics (such as personality and gender) can influence human computer interaction.

Further research on human computer interactions (HCI) has shown that users will respond in a social manner to interfaces that exhibit social cues, either through text, voice, or animations. Similarly, communication research has focused on how the user interface design may influence communication processes and related constructs. Researchers in the HCI and communications areas have begun investigating how advanced interfaces interact with characteristics of the user. Given the recent advances in our ability to design these types of complex interfaces, this research is necessarily in its early stages.

The purpose of this research is to investigate how gender, personality similarity, and the multiplicity of information cues provided in the interface affect the user's involvement with the system and subsequent decision-making performance. Research on consumer information-processing and involvement provide the theoretical foundation for the study. An experiment that manipulates computer-based personality and interface information cues has been designed and carried out to assess the impact of increased involvement on decision-making outcomes in a computer-based decision support environment.

The paper is organized in the following sections. First, the theoretical framework and hypotheses are presented. The research design, a 2x2 between subjects experiment, is then described, and results from the completed experiment are briefly summarized. A full discussion of the experimental findings will be presented at the workshop.

THEORETICAL FRAMEWORK

The theoretical framework for this paper focuses on the concept of involvement in the context of consumer information processing and decision-making. First, the involvement construct is reviewed and relevant IS research is identified. Determinants of involvement in the context of computer-based decision support are then discussed. Gender, personality similarity of the decision aid and the decision-maker, and media richness are reviewed for the potential influence on involvement. Hypotheses for each determinant of involvement are presented as the relevant literature is reviewed. The impact of involvement and communication levels on

decision-making performance is then described and related hypotheses are presented. The hypotheses supported below are visually represented in the research model shown in figure 1.

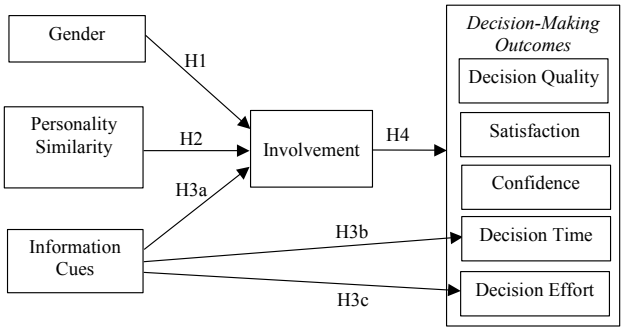


Figure 1. Research Model: the Impact of Gender, Personality Similarity, and Cues on Performance.

Involvement

Researchers in consumer information processing have long recognized the importance of involvement, or focused attention, on decision-making performance (Celci and Olson 1988; Payne, Bettman, Johnson 1993; Mishra, Umesh, and Stem 1993). Involvement affects information processing at a fundamental level as greater levels of involvement can lead to greater information acquisition, improved comprehension/understanding, increased effort, and overall decision performance. In this context, a person's level of involvement with a task has been defined as the degree to which the person finds the task to be personally relevant and is motivated to complete the task (Celci and Olson 1988). Involvement is believed to come from two broad sources: 1) intrinsic or stable sources due to individual differences and 2) situational sources, those that may be manipulated within the immediate environment (Celci & Olson 1988). Increased information cues in an interface (richer media) are believed to impact the situational form of involvement.

Recent MIS research has investigated involvement-related constructs in the context of the technology acceptance model (TAM). Agarwal and Karahanna note that current IS often employ richer media that provide an " increasingly riveting and engaging experience" (2000, p. 667). Focused attention/immersion is one of the dimensions of the cognitive absorption (CA) construct that they developed and studied in the context of TAM (Agarwal and Karahanna 2000). This notion of focused attention, where a user's attention is completely focused on an activity, is closely related to the construct of involvement advanced in the consumer information processing literature. Similarly, Koufaris has applied the concept of flow from the psychology literature to online consumer behavior in the context of TAM (2002). A state of flow occurs when an individual is absorbed in a task and acts with complete involvement (Csikszentmihalyi 1988).

The MIS research community has thus recently recognized the importance of focused involvement with an IS on user acceptance of an IS. From a decision-making perspective, however, the impact of greater involvement on performance outcomes has received little attention. One study in the communications literature, with limited sample size, investigated involvement, some decision-making outcomes, along with other communication measures in an experimental communication task (Burgoon, Bonito, Bengtsson, Ramirez, Dunbar, and Miczo 2000). Additional research is needed to identify the relevant influences on involvement and the subsequent effect on decision-making performance in the context of computer-based decision aids and information processing.

Gender and Technology

Gender researchers in social behavior (Skitka and Maslach 1996), communication (Dennis, Kinney, and Hung 1999; Spangler 1995), and IS acceptance (Gefen and Straub 1997; Venkatesh and Morris 2000) have noted differences in how men and women interact with each other and technology. Women are perceived to be more socially focused than men are as they are more aware of other's feelings and concerned with group harmony, consensus building, and interrelationships. Men, on the other hand, are viewed as being more independent, assertive, and unemotional.

In the context of technology acceptance, this more socially focused view of women has been empirically supported. Gefen and Straub found that women perceived a higher level of social presence in email than did men (2000). Dennis et al. found support for the premise that women were more sensitive, or aware, of non-verbal social cues in computer-mediated conditions (1999). Greater awareness of non-verbal social cues and perceptions of greater social presence suggest that women may be more involved or attentive in social interactions. While gender research has found differences between men and women in communication patterns and in initial beliefs or expectations with regard to technology, there has been less support for gender differences in actual performance with technology.

Based upon these findings, women appear to be more socially focused than men, and more observant of social cues in general. In addition, women have been found to perceive a greater social presence in electronic communication (Geffen and Straub 1997). Therefore, in the context of a computer-based decision aid,

Hypothesis 1. Women will be more involved than men.

Personality Similarity and Computers

Researchers in communication and HCI have demonstrated that users respond in a human-like manner to social cues exhibited by computing applications (Nass and Lee 2001; Nass and Moon 2000; Burgoon et al. 2000; Moon and Nass 2000). This application of social rules to computing applications is referred to as the Computers as

Social Actors (CSA) paradigm. This paradigm asserts that users respond to social cues from computers with social behaviors, but that this conditioned response occurs despite the user knowing that the computer is not human. One important finding of this research is that users can accurately assess personality traits in computing applications and respond differently to the application depending upon their own personality. In this manner, a computing application that exhibits personality traits is providing additional information cues. Personality theories offer explanations for these responses.

Similarity-attraction theory (Byrne & Griffitt 1969) states that individuals will be more attracted to individuals that exhibit similar characteristics. It has been applied to interactions with friends, business colleagues, partners, and computing applications. The theory asserts that people are more comfortable with people that exhibit personality traits that are similar to their own traits, especially in the early stages of a relationship. In human-computer interactions, the theory predicts that users will be more comfortable with computer-based personalities that exhibit personality traits that are similar to their own personality traits. Several studies that examined the personality traits exhibited by a computing application found support for similarity-attraction theory (Nass and Lee 2001; Burgoon et al 2000; Nass & Moon 2000).

The psychology literature and prior HCI studies provide support for the relationship between personality similarity and user perceptions of a computer-based decision aid that exhibits personality traits. Therefore, in the context of a decision-aid that exhibits personality traits and the known personality traits of the user,

Hypothesis 2. The similarity of personality traits will increase involvement.

Communication Levels and Media Richness

As noted previously, the development of more engaging, media-rich interfaces has been viewed as an improvement over simple, text interfaces. Theoretical support for this assumption, however, has been lacking. Media richness theory (Daft & Lengel 1986) is an obvious choice for investigating the effects of richer interfaces on user involvement and decision-making performance, but the empirical tests of this theory have provided disappointing results.

According to media richness theory, richer media should enable users to more quickly communicate and better understand equivocal information. Since media vary in their ability to support communication, richer media are thought to convey information better through a greater multiplicity of cues (visual information, tone of voice, etc.), by allowing greater personalization of the message, and by providing faster feedback.

Most of the early studies on media richness focused on media choice, however, not media use, and found only limited, if any, support for the theory (Daft, Lengel and Trevino 1987; Trevino, Lengel, and Daft 1987). Later studies focused on the performance effects of media use, but still did not find support for the theory (Dennis and Kinney 1998; Dennis et al 1999). The later studies did, however, find results that are applicable to general studies of richer versus leaner media. The use of leaner versus richer media resulted in slower performance overall, regardless of task equivocality (Dennis & Kinney 1998; Dennis et al 1999). Richer media supported the communication of more information cues, and thus reduced the time required to communicate information and make decisions. Similarly, richer media could be assumed to reduce decision-making effort, as the additional information cues should result in better comprehension from the delivery of the same message. The additional levels of information, or multiplicity of information cues, could also provide a means to alter the situational involvement of the user. The increased level of information cues should focus more of the user's attention on the interface. Therefore,

Hypothesis 3a. The multiplicity of information cues will increase involvement.

Hypothesis 3b. The multiplicity of information cues will decrease decision time.

Hypothesis 3c. The multiplicity of information cues will decrease effort.

Involvement and Decision Performance

The consumer information processing literature (Payne et al 1993; Celci and Olson 1988; Mishra et al 1993) provides theoretical support for the influence of involvement on various measures of decision-making performance. Multiple aspects of decision performance were investigated in this study to provide a more rich understanding of decision aid involvement and increased information cues. An individual, who is more involved in a decision-making task, is more committed to completing the task and is thus more likely to devote increased effort and time to the task. This involvement or motivation to complete a decision-making task may also translate into improved decision quality. A more committed, involved individual, who feels that a decision-making task is more personally relevant, would also feel higher levels of satisfaction with the task and be more confident in the accurate completion of the task. Therefore, in the context of a computer-based decision aid,

Hypothesis 4. Higher levels of involvement will increase decision-making outcomes (decision time, effort, decision quality, satisfaction, and decision confidence).

RESEARCH METHODOLOGY

A 2x2, between subjects research design was used, varying the communication levels (text only - T, text and voice -TV) and the personality of the decision aid (dominant, submission). Participants were 184 undergraduate students recruited from a sophomore-level

business course with a research study participation requirement. The average age of the students was 20.8, with 121 males and 67 females participating.

The subjects performed an apartment selection task similar to that employed by Todd & Benbasat (1999) and Payne et al. (1993) in prior decision-making studies. This task was chosen as it is a personally relevant choice problem for most college students. The subjects were presented with ten apartment alternatives that varied by eight attributes (rent, size, laundry, distance, high speed Internet access, facility age, parking, and noise.

Treatment Conditions

The communication levels (T, TV) were developed using the Microsoft® Agent Technology. In the T treatment, the decision aid provides subjects with instructions through text displayed in text balloons. In the TV treatment, the text balloons along with a computer-generated voice that reads the text in the balloons is provided.

The dominant-submissive dimension from the five-factor personality model (Trapnell & Wiggins 1990) was used to assess the impact of personality similarity. This dimension represents the degree to which an individual is assertive and willing to exercise control over others, and was represented in the treatments by varying word choice and voice characteristics in keeping with the personality literature and similar experiments on personality traits (Nass & Moon 2000, Burgoon et. al. 1999). The information content was kept the same in all treatments, but the script used in the dominant treatment included more assertive, commanding statements, while the submissive treatment script used more timid, unassuming statements. The voice used in the TV treatments was also varied to represent dominant and submissive traits. Based upon the personality literature and previous studies, the dominant voice was given a higher overall frequency, a larger range of pitch during speech, and greater speed than the submissive voice.

Procedure

A pre-experiment survey was first administered to ascertain the subjects' perceptions of their own personality traits (dominant or submission). The subjects were then randomly assigned to one of the four treatment conditions (dominant text, dominant text and voice, submissive text, submissive text and voice).

The computer-based decision aid provided the user with instructions on how to use the tool and then guided the user through the actual use of the tool. The delivery of the instructions was in keeping with the subjects' assigned treatment condition. The decision aid first requested the subjects to specify their preference for each apartment attribute by allocating 100 points among the eight attributes. The decision aid then provided a spreadsheet-based interface with several functions to facilitate the subject's selection of an apartment. These functions include hiding/showing apartment alternatives (rows) and features (columns), changing the order of the apartments and features, and sorting by one or two of the apartment features. The subjects were instructed to rank order the apartments according to their preference and then select their preferred apartment. After the students selected an apartment, a post-experiment survey was administered.

Measures

The measurement of the subject's personality on the dominant-submissive dimension of extraversion and the manipulation check on the dominant-submissive nature of the decision aid was obtained using a 16-item adjective scale (Trapnell & Wiggins 1990). Personality difference scores were calculated from the subject's mean response to the dominant-submissive scale items and the dominance of the treatment assigned. These raw scores were then converted to z scores, to facilitate comparison among the different treatments.

The involvement scale was developed from an existing 5-item scale that measures a user's focused attention (immersion) with an information system (Agarwal & Karahanna 2000). This scale was comparable to marketing scales used to measure involvement or attention in non-IS settings and more reliable than comparable communication scales (Burgoon et al. 2000).

Decision-making performance was measured by several common decision-making outcomes. Subject effort was evaluated by the number of decision aid features used during the experiment, as tracked by the experimental application. Decision time was measured in minutes and seconds by the experimental application. Satisfaction with the decision aid was measured with a 4-item scale adapted from other IS satisfaction scales (Doll & Torkzadeh 1988). The subjects' confidence in their decision choice was measured with a 4-item scale adapted from the decision support literature (Ghosh and Ray 1997). Decision quality was measured by comparing the subjects' final selections to their normative choice using a weighted-additive calculation.

RESULTS

Analysis of the experimental results has been completed and a summary of the hypotheses testing is shown in Table 1. Due to space limitations, statistical analysis and discussion of the results is not included in the proceedings, but will be presented at the workshop.

Hypotheses	Findings
H1. Gender ⇒ Involvement, Women > Men	√
H2. Personality similarity ⇑ Involvement	x
H3a. Communication levels ⇑ Involvement	x
H3b. Communication levels ⇓ Decision time	√
H3c. Communication levels ⇓ Decision effort	√

H4a. Involvement ⇑ Decision time	√
H4b. Involvement ⇑ Decision effort	√
H4c. Involvement ⇑ Decision quality	x
H4d. Involvement ⇑ satisfaction	√
H4e. Involvement ⇑ Decision confidence	√

Table 1. Result Summary

REFERENCES

1. Agarwal, R. & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24,4, 665-694.
2. Burgoon, J., Bonito, J., Bengtsson, B., Ramirez, A., Dunbar, N. & Miczo, N. (1999). Testing the interactivity model: Communication processes, partner assessments, and the quality of collaborative work. *J. of Management Information Systems*, 16, 3, 33-56.
3. Byrne, D. & Griffitt, W. (1969). Similarity and awareness of similarity of personality characteristics as determinants of attraction. *Journal of Experimental Research in Personality*, 3, 179-186.
4. Celci, R. & Olson, J. (1988). The role of involvement in attention and comprehension processes. *Journal of Consumer Research*, 15, 1, 210-224.
5. Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience, Harper and Row, New York.
6. Daft, R.L., & Lengel, R.H. (1986). Organizational information requirements, media richness and structural design. *Man. Science*, 32, 5, 554-571.
7. Daft, R.L., Lengel, R.H., & Trevino, L.K. (1987). Message equivocality, media selection and manager performance: Implications for information systems. *MIS Quarterly*, 11, 3, 355-366.
8. Dennis, A., Kinney, S., and Hung, Y. (1999). Gender differences in the effects of media richness. *Small Group Research*, 30, 4, p. 405-438.
9. Dennis, A. & Kinney, S. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research*, 9, 3, 256-274.
10. Doll, W. & Torkzadeh, G. (1988). The measurement of end user computing satisfaction. *MIS Quarterly*, 12, 2, 259-274.
11. Gefen, D. & Straub, D. (1997). Gender differences in the perception & use of e-mail: Extension to the technology acceptance model. *MIS Quarterly* 21, 4, 389-400.
12. Ghosh, D. and Ray, M. (1997). Risk, ambiguity, and decision choice: Some additional evidence. *Decision Sciences*, 28, 1, 81-105.
13. Goodhue, D. (1995). Understanding user evaluations of IS. *Management Science*, 41, 12, 1827-1844.
14. Goodhue, D.L. & Thompson, R.L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19, 2, 213-236.
15. Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, 13, 2, 205-223.
16. Mishra, S., Umesh, U., & Stem, D. (1993). Antecedents of the attraction effect: An information processsing approach. *J. of Marketing Research*, 30, 331-49.
17. Moon, Y. & Nass, C. (1998). Are computers scapegoats? *Int. J Human-Computer Studies*, 49, 79-94.
18. Nass, C. & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied* 7, 3, 171-181.
19. Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56, 1, 81-103.
20. Payne, J., Bettman, J., & Johnson, E. (1993). The Adaptive Decision Maker, Cambridge University Press, New York, 1993.
21. Skitka, L.J. & Maslach, C. (1990). Gender roles and the categorization of gender-relevant behavior. *Sex Roles*, 22, 3-4, 133.
22. Spangler, L. (1995). Gender-specific nonverbal communication: Impact for speaker effectiveness. *Human Resource Development Quarterly*, 6, 4; 409-419.
23. Todd, P. & Benbasat, I. (1999). Evaluating the impact of DSS, cognitive effort, and incentives on strategy selection. *Information Systems Research*, 10, 4, 356-374.
24. Trapnell, P. & Wiggins, J. (1990). Extension of the interpersonal adjective scales to include the big five dimensions of personality. *Journal of Personality and Social Psychology*, 59, 4, 781-790.
25. Trevino, L., Lengel, R.H., & Daft, R.L. (1987). Media symbolism, media richness, and media choice in organizations. *Comm. Research*, 15, 5, 553-574.
26. Venkatesh, V. & Morris, M.G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly*, 24, 1, 115-139.
27. Vessey, I. (1991). Cognitive fit: a theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22, 2, 219-241.
28. Vessey, I. & Galletta, D. (1991). Cognitive fit: an empirical study of information acquisition. *Information Systems Research*, 2, 1, 63-84.

A Test of the Theory of DSS Design for User Calibration: The Effects of Expressiveness and Visibility on User Calibration

Brian M. Ashford
U.S. Army Logistics Management College
ashfordb@Lee.Army.mil

George M. Kasper
Virginia Commonwealth University
gmkasper@VCU.edu

ABSTRACT

This paper reports a test of the theory of decision support systems design for user calibration that compares the efficiency of the visual computing paradigm with that of the conventional text paradigm over varied levels of problem novelty. Perfect user calibration exists when a user's confidence in a decision equals the quality of the decision. The laboratory study reported here compared the effects on user calibration of problems depicted either using a text paradigm or visual computing paradigm. The results support the theory. When problems are new and novel, visual depiction improves user calibration. As problems became more familiar and problem novelty decreases, no difference was found in user calibration between subjects exposed to visibility diagrams and those exposed to a traditional text paradigm.

INTRODUCTION

One's belief in the quality of a decision influences the decision selection process (Russo & Schoemaker 1992). Failure to appreciate the quality of a decision can mean that good decisions are not implemented or poor decisions are not properly hedged. Although confidence, as discussed herein, is a subjective prediction, in many situations its accuracy can be objectively assessed. The best-known measure of the accuracy of one's confidence in a decision is calibration, the correspondence between one's prediction of the quality of a decision and the actual quality of the decision (Lichtenstein et al. 1982, Clemen & Murphy 1990, Keren 1991). When this correspondence is equal, and one's decision confidence equals the quality of the decision, calibration is said to be perfect. Perfect calibration is indispensable when selecting a decision from among competing alternatives (Russo & Schoemaker 1992).

The theory of decision support systems (DSS) design for user calibration prescribes requisite DSS design properties needed for users to realize the performance goal of perfect calibration (Kasper 1996). Reviewed below, the theory asserts that a DSS can engender perfect calibration to the extent that it contains requisite properties of **Expressiveness** (expression of words, phases, and audio ranging from, e.g., cryptic to anthropomorphic), **Visibility** (visual icons, images, and animation ranging from, e.g., realistic to abstract), and **Inquirability** (investigative tools and styles ranging from, e.g., data-oriented servile to dialectic contrarian inquiry). The theory further asserts that as problem novelty increases the effective mix of three properties varies from expressiveness to visibility to inquirability.

This paper reports a partial test of the theory of DSS design for user calibration. The effects on user calibration of expressiveness in the form of text and visibility in the form of diagrams were investigated at two levels of problem novelty. Specifically, a laboratory study was conducted in which subjects were exposed to logically identical sets of problems displayed using either expressiveness text or visibility diagrams, and user calibration was computed and compared for higher and lower levels of problem novelty. The results show that the effects of the instantiations of expressiveness and visibility on user calibration varied as prescribed by the theory: visibility resulted in significantly better user calibration when problem novelty was higher, but there was no difference in user calibration between visibility and expressiveness when problem novelty was lower. In other words, visual computing had its greatest impact on user calibration when problems were new and novel.

BACKGROUND

Differentiating confidence, trust, predictability, and decision accuracy, Muir (1994, p. 1915, parenthetics added) states,

Predictability is a basis for trust (and confidence), which in turn, is the basis for an operator (user/decision maker) to make a *prediction* about the future behaviour of a referent. The *accuracy* of that prediction may be assessed by comparing it with the actual behavioural outcome. In addition, an individual who makes a prediction may associate a particular level of *confidence* with the prediction. Thus, *confidence* is a qualifier which is associated with a particular prediction; it is not synonymous with trust.

Realism in confidence is essential for good decisions; the ruinous consequences of unrealistic confidence litter the business decision-making landscape (Russo & Schoemaker 1992). Because action precedes outcome, confidence plays an essential role in both selecting and implementing a decision (Russo & Schoemaker 1992). The confidence ascribed to a predicted outcome when compared to the accuracy of that prediction measures the decision maker's ability to calibrate his or her ascribed confidence

Since its beginning, the primary goal of DSS has been to improve decision quality (Keen & Scott Morton 1978). Unfortunately, evidence suggests that existing DSS can produce "illusory benefits" (Aldag & Powers 1986, Davis et al. 1991), resulting in miscalibration, thereby distorting the decision selection process. Thirty years ago, Chervany and Dickson (1974, p. 1342, parenthetics added) recognized this when they wrote, "Even though the . . .(decision aided) subjects (in their study) did better, their increased average time and reduced average confidence lead to the tentative conclusion that they did not have a 'handle' on the problem." By now, almost everyone can recount from personal experience a situation where computer-generated output produced an aura of exactness and reliance bordering on blind acceptance, even in the presence of

compelling evidence to the contrary. In these cases, user calibration may be distorted by the design of the DSS.

Based on and paralleling human problem solving, memory representation, and multiple intelligence theories (Kaufmann 1985; Helstrup 1987; Gardner 1993), Kasper (1996) proposed the notion and detailed a theory of DSS design for user calibration. His design theory prescribes requisite properties of a DSS so the user/decision maker can achieve the goal of perfect calibration. The theory asserts that a user/decision maker can achieve the goal of perfect calibration to the extent that the DSS possesses requisite properties of expressiveness, visibility, and inquirability, and that the effective mix of these properties varies with problem novelty.

The theory of DSS design for user calibration is a design theory (Walls, et al. 1992). It posits a goal, perfect calibration; properties, expressiveness, visibility, and inquirability; and the interaction of these properties to achieve the goal, a mix of expressiveness, visibility, and inquirability that varies systematically with problem novelty.

Expressiveness recognizes that the tone and delivery of words and phrases (written and audio) used in a human-computer interface dialogue (ranging from cryptic to anthropomorphic, from monotone and monotonous to melodic and overly melodramatic) can affect people's beliefs, perceptions, opinions, and predictions. **Visibility** encompasses the icons, symbols, and animation that promote discovery, comprehension, problem solving and engender feelings (Card, MacKinlay, & Shneiderman 1999, Gonzalez & Kasper 1997). **Inquirability** captures the affects produced by actions and interactions with the inquiring system, including scope and nature of dialectics (Churchman 1971) and the restrictiveness and decisional guidance of the system (Silver 1990).

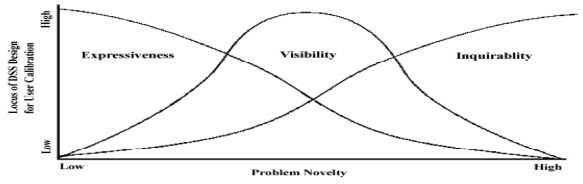


Figure 1 Locus of DSS Design for User Calibration in Relation to Problem Novelty (Kasper 1996)

Depicted in Figure 1, the theory of DSS design for user calibration posits that when problems are somewhat novel and unfamiliar, **Visibility** is the primary contributor to perfect calibration and **Expressiveness** and **Inquirability** play important but lesser, supporting roles. As problems become more familiar and problem novelty decreases, the theory posits that the contribution of **Expressiveness** increases, equals, and eventually exceeds **Visibility** as the primary contributor to user calibration. Stated in the null form, it is hypothesized that:

H_0: There is no difference in user calibration between subjects exposed to Expressiveness and those exposed to Visibility at higher and lower levels of problem novelty.

Larkin and Simon (1987) posited a beneficial role for visibility in search, recognition, and inference processing, and, in response, Bauer and Johnson-Laird (1993) studied the effects of diagrams on inference and found that the use of diagrams improved decision quality. Commenting on their findings, Bauer and Johnson-Laird suggested that in unfamiliar, novel situations, diagrams have a beneficial effect on decision making. Recently, Speier & Morris (2003) found that visual interface users performed better when task complexity was high and their subjective mental workload was less compared to users of a text-based interface. Extending these findings, the study reported here considers the effect of visibility on user calibration and whether this effect, if observed, varies with problem novelty.

EXPERIMENTAL DESIGN, RESEARCH METHOD AND MEASURES

To investigate the hypothesis, a laboratory experiment was conducted. The main effect studied was properties of DSS dialogue design and the dependent variable was user calibration. Specifically, the differential effect of expressiveness and visibility on user calibration was investigated. The experimental design included two different problems to increase the generalizability of the findings and to build upon earlier related research, in particular, that of Bauer and Johnson-Laird. Two calculations of problem novelty, Higher and Lower, were defined by dividing each subject's responses into earlier and later decisions, again based on the work of Bauer and Johnson-Laird. The treatments, measures, formula used to calculate user calibration, and procedures used in the experiment are discussed in detail below.

Treatments

The treatment combinations used in this study were borrowed directly from those developed by Bauer and Johnson-Laird to study deductive reasoning and inference. They developed two logically identical problems presented either as text, a form of expressiveness, or diagrams, a form of visibility. In the interest of space, the reader is directed to Bauer and Johnson-Laird (1993) for detailed descriptions of these treatment conditions. To investigating the hypothesis posited here, subjects also recorded their decision confidence in their selection.

Measuring User Calibration

To measure user calibration requires selecting a method and means for recording both decision quality and the subject's belief in the quality of each decision, a scoring rule and procedure that discourage gaming so that subjects are encouraged to honestly report their beliefs, and a formula for calculating calibration. Each of these requirements is discussed in the next sections.

Recording Beliefs And Decisions

Following convention in calibration research, subjects in this study answered a series of multiple-choice questions by reporting both their decision and confidence in the correctness of each decision. Each subject answered a total of ten multiple-choice questions. The ten questions consisted of the four questions used in the Bauer and Johnson-Laird (1993) study plus six additional questions generated using the same truth table. For each of these ten questions, the subject selected one alternative as his or her choice as the correct alternative and then assigned a confidence value to that alternative and other alternatives as desired. Analysis of pilot study data showed that assigning confidence values to multiple alternatives improved user calibration; a finding consistent with that of Sniezek et al. (1990).

Recording Sales

Confidence is typically recorded on a scale ranging from 0 to 1 or some subset. In this study, this range was divided into increments of five-hundredths (i.e., 0.0, 0.05, 0.10, 0.15,..., 1.0) because research suggests that this is consistent with the respondent's "natural scaling" of decision confidence (Winkler 1971).

Scoring Rules

The purpose of a scoring rule is to encourage respondents to honestly report their confidence in each decision by eliciting values that reflect the respondent's actual belief in the quality of his or her selection. For this to occur, a scoring rule must (1) be understood by the subject so that its implications and the correspondence between beliefs and numerical values can be fully appreciated, and (2) maximize the subject's expected total score only when the subject reports values that correspond to his or her actual beliefs (Stael von Holstein, 1970).

Assume that a subject's true decision confidence is expressed by probability vector $P = (p_1, p_2, ..., p_n)$ for a mutually exclusive and collectively exhaustive set of events, $\{E_1, E_2, ..., E_n\}$. Assume further that the confidence values an assessor reports are represented by $R = (r_1, r_2, ..., r_n)$. A proper scoring rule S exists if S is maximized only when $r = p$. This requirement is satisfied by only a very few somewhat complex scoring rules that require the respondent to perform high level operations such as exponential, root, or log calculations (Murphy & Winkler 1970). These complex operations make it almost impossible for subjects to quickly compute and fully appreciate the implications of their decisions and the correspondence between their actual beliefs and the values they report. In other words, these scoring rules confuse and may actually interfere with the subject's reporting values reflecting his or her actual beliefs.

A scoring rule that meets the criterion of understandability is the well-known simple linear scoring rule $S_k(r) = r_k$, where k refers to the event that actually occurred and r_k is the confidence probability assigned by the subject to the kth response. Unfortunately, in its simplest form, this scoring rule is not strictly proper because $S(r,p) = \sum p_k r_k$ is maximized by setting one r_i (i.e., the r_i corresponding to the largest p_i) equal to 1.0 and the other r_is equal to 0.0. If $r_{i=k}$, then the subject appears to have complete confidence in the answer that turns out to be correct. On the other hand, if $r_{i\neq k}$, the subject appears totally wrong, but losses nothing because the scoring rule imposes no penalty for being wrong. In other words, a subject maximizes his or her score by assigning a confidence of 1.0 to one answer despite his or her true belief in the quality of any answer.

Despite this limitation, most calibration research has used some variation of this simple linear scoring rule. In fact, comparing three complex proper scoring rules to the simple linear scoring rule, Rippey (1970) reported that the simple linear scoring rule actually produced more reliable results. Likewise, reviewing a number of these studies, Phillips (1970) concluded that the complex proper scoring rules did not yield significantly different values than those collected using a simple linear scoring rule, but, as expected, subjects found simple linear scoring rules more realistic and easier to understand.

Considering these tradeoffs, this study used a variant of the simple linear scoring rule that discouraged gaming and guessing by penalizing wrong answers. The scoring rule used here was:

$$S = r_k - [(\text{largest } r_{i\neq k})/2$$

where S is the score, k refers to the correct alternative, r_k is the confidence probability assigned to that alternative, and $r_{i\neq k}$ are the confidence probabilities assigned to the alternatives that turn out to be incorrect. This variant of the simple scoring rule is easily understood because its implications can be more readily appreciated and the respondent can better understand the correspondence between her beliefs and numerical values she reports. Yet, subjects are encouraged to report numerical values that correspond to their actual beliefs because of the penalty of one-half the largest confidence value assigned to an alternative that is wrong.

Computing Calibration

The most popular calculation for calibration is:

$$calibration = \frac{1}{N}\sum_{t=1}^{T} n_t \left(r_t - c_t \right)^2$$

where N is the total number of responses, n_t is the number of times the confidence value r_t is used, c_t is the proportion correct for all items assigned confidence value r_t, and T is the total number of different response categories used (Lichtenstein & Fischhoff 1977, Clemen & Murphy 1990). Using this formula, perfect calibration is a score of 0.0. The worst possible score, 1.0, can only be obtained when the responses are completely and consistently wrong; that is, $r_t = 1.0$ is always assigned to the wrong answer and $r_t = 0.0$ is always assigned to the answer that turns out to be correct.

Procedures

Subjects were recruited from students enrolled in upper-division, undergraduate courses in information systems and psychology. All participants volunteered for the study and were rewarded course credit as required by American Psychological Association guidelines (1992).

Upon arrival, each subject was randomly assigned to one combination of the two treatment levels, expressiveness or visibility, and the two problems, so as to balance the number of subjects in each cell of the experimental design. The subject then read a two-page handout of instructions that included an example of the expressiveness or visibility display, depending upon the treatment condition assigned, and a description of the navigation procedures and operations the subject would be using to answer the multiple-choice questions. The instructions also included a detailed discussion of the scoring rule, including a table of all possible outcomes that could be referred to throughout the study. The subject was then guided through a demonstration, and questions regarding the procedures and objectives of the study were answered. Each subject then completed a consent form and a short, 11-item questionnaire designed to collect descriptive demographic and background data. To describe the groups' visual acuity, the 16-question Vividness of Visual Imagery Questionnaire (Marks 1972, 1973) was also administered. The subject then began answering the ten questions presented as either visibility diagrams or expressiveness text.

To minimize any question ordering effect, the ten questions in each treatment combination were counterbalanced by order with each question presented in each order position once. This resulted in ten different primary orderings of the ten questions in each treatment. Each question was displayed and data collected

using Dell II machines with 17" monitors. The display used in the study was written in ToolBook 5.0 by Asymetrix.

DATA ANALYSIS AND RESULTS

A total of 54 students participated as subjects in the study. Forty subjects, 10 in each group, completed all aspects of the experiment, followed all the instructions and answered all the questions. Although subjects were not given a specific time restriction, on average, they took about 35 minutes to complete all aspects of the study.

Seventy percent of the subjects in the study were information systems majors and the remainder were psychology majors. Most subjects were adult, non-traditional students reporting an average age of 30.3 years. Forty-seven percent of the subjects were female and 80 percent reported that English was their native language. As a group, subjects also reported average to above average (mean = 32.4; s.d. = 9.62) visual acuity as measured by the Vividness of Visual Imagery Questionnaire and self-reported "average" facility with logic and math problems.

Recall that each subject in each treatment answered ten counterbalanced questions. The ten responses from each subject were divided into the first four and the last six responses, again, based on Bauer and Johnson-Laird's research. A calibration score was then computed for each of these two subsets for each subject. These subsets defined the two levels of problem novelty. Calibration based on the first four responses defined the Higher category of Problem Novelty and calibration computed on the subject's last six responses defined the Lower category of Problem Novelty.

Analysis of this data shows that the content of the problem, electrical circuit or people and places, had no effect on either percentage correct (questions 1-4, $F_{(1,36)} = .08$, $p = 0.7$ and questions 5-10, $F_{(1,36)} = .01$, $p = 0.9$) or user calibration (questions 1-4, $F_{(1,36)} = .07$, $p = 0.7$ and questions 5-10, $F_{(1,36)} = .01$, $p = 0.9$), so the data was collapsed over the problem content scenarios. In terms of percentage correct, these results are identical to those found by Bauer and Johnson-Laird who also collapsed the data over the same people-and-places and electric circuit scenarios. The mean of user calibration of this pooled data is shown for expressiveness and visibility for the two Problem Novelty categories in Figure 4.

Focusing on the higher category of the Problem Novelty axis shows that subjects using the visibility (V) treatment were much better calibrated, had calibration scores closer to zero, than were those assigned to the expressiveness (E) treatment. Conversely, at the Lower category of Problem Novelty there seems to be little difference between the average calibration of those exposed to expressiveness (E) and those exposed to visibility (V). In other words, over the last six questions, when decisions were more familiar and less novel, exposure to visibility or expressiveness did not differentially affected user calibration.

Figure 4 also shows that average user calibration for the visibility (V) treatment was overall the best, closest to zero, at the Higher category of Problem Novelty (.078). The next best level of user calibration was at the visibility (V) Lower category of Problem Novelty (.100). Comparing these results, the data suggest that the same subjects exposed to the visibility (V) treatment produced better user calibration in the first four tries (.078), when problem novelty was highest, than they did over the last six tries (.100) when problem novelty was lower.

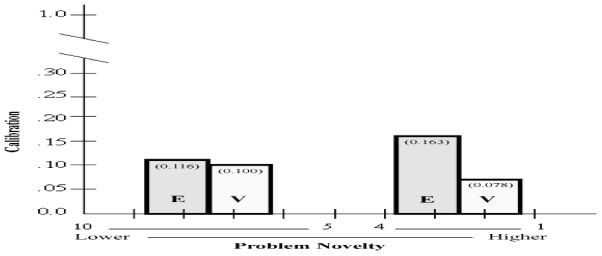


Figure 4. Mean Calibration of Expressiveness (E) and Visibility (V) for Higher and Lower Problem Novelty.

For expressiveness (E), the results in Figure 4 show that subjects exposed to the expressiveness (E) treatment had on average poorer calibration than did those exposed to visibility. Expressiveness produced the poorest average user calibration at both the higher and lower category of Problem Novelty (.163 & .116). However, comparing the two expressiveness (E) bars shows that there was a marked improvement in user calibration from the Higher to the Lower category of Problem Novelty for those exposed to expressiveness (.163 to .116). In this regard, the change in user calibration for those exposed to the expressiveness treatment was as might be expected, user calibration improved as problem novelty decreased.

To assess the statistical significance of the differences in user calibration suggested by the means depicted in Figure 4, a multivariate analysis of variance (MANOVA) was computed using the two dependent variables, user calibration at the Higher and Lower groupings of Problem Novelty, and the independent variable of DSS Locus of Design, either expressiveness or visibility, for each subject. This model produced a Wilks Lambda treatment effect of $F_{(2,37)} = 2.8$, $p = 0.07$. Although insignificant at the $\alpha = 0.05$ level, this result does not preclude significant univariate effects. Indeed, in the case of strong positive correlation between the dependent variables ($r = 0.45$, $p = 0.0031$), and interaction consistent with that hypothesized in Figure 1, the multivariate test is less powerful than it would be if the data were negatively correlated (Bray & Maxwell 1988, pp. 31-32). In other words, the Wilks Lambda *F*-value may be confounded by the nature of the interaction between dependent variables.

1.a. ANOVA Results of User Calibration by Expressiveness and Visibility for Higher Problem Novelty (questions 1-4).

Source	df	Type III SS	F-Value	P-Value
E/V	1	.073	5.232	.028*
Error	38	.528		
Total	39	.601		

$R^2 = 0.121$; * $p < .05$

1.b. ANOVA Results of User Calibration by Expressiveness and Visibility for Lower Problem Novelty (questions 5-10).

Source	df	Type III SS	F-Value	P-Value
E/V	1	.002	.229	.635
Error	38	.405		
Total	39	.407		

$R^2 = 0.006$

Table 1: Analysis of Variance of User Calibration for Higher (questions 1-4) and Lower Problem Novelty (questions 5-10).

To clarify the MANOVA results, analysis of variance (ANOVA) was computed for the Higher and Lower groupings of Problem Novelty separately. The results of these analyses are presented in Table 1.

The first ANOVA, Table 1a, shows results for data from the higher category of Problem Novelty. These data show that subjects exposed to visibility produced user calibration that was significantly better than those subjects exposed to expressiveness ($F_{(2,37)} = 5.23$, $p = 0.028$). The Bonferroni minimum significant difference of 0.0755 confirms that the difference between 0.163 and 0.078 is significant at the $\alpha = 0.05$ level. For this data, H_0 can be rejected. The evidence shows that for the higher category of Problem Novelty (i.e., when the problems were the most novel), the average calibration of subjects using visibility diagrams was significantly better than it was for those subjects using expressiveness text.

In contrast, results in Table 1b show no significant difference in user calibration as a result of visibility and expressiveness treatment levels ($F_{(2,37)} = .229$, $p = 0.635$). The Bonferroni minimum significant difference of 0.0661 exceeds the 0.016 difference in means (0.116 - 0.100). In this case, H_0 cannot be rejected. The data indicate that when problem novelty was Lower and problems were more familiar and less novel, there was no difference in user calibration between subjects using visibility diagrams and those using expressiveness text.

Though not related to the hypothesis, comparisons of visibility (V) or expressiveness (E) across Higher and Lower levels of Problem Novelty resulted in no significant differences. Likewise, comparing visibility (V) at the Higher level of Problem Novelty to visibility and expressiveness at the Lower level of Problem Novelty resulted in no significant differences. Analyses also showed no significant difference in user calibration due to VVIQ subject differences (questions 1-4, $F_{(1,37)} = 2.57$, $p = 0.12$; questions 5-10, $F_{(1,37)} = .14$, $p = 0.71$) or decision time. These results add to the generalizibility of the main finding that visibility improves user calibration when problems are new and somewhat novel.

SUMMARY AND CONCLUSIONS

The results of a partial test of the theory of DSS design for user calibration are reported. Specifically, a laboratory study was conducted to compare the effects of expressiveness and visibility on user calibration at two levels of problem novelty. The results of this study support the theory. When problems were new and novel, visibility diagrams significantly improved user calibration compared to expressiveness text. Later, when problems became more familiar, less novel, there was no difference in user calibration between visibility and expressiveness.

Bauer and Johnson-Laird (1993) and Speier and Morris (2003) report that diagrams improved decision quality. The results reported here demonstrate that visibility diagrams also improve user calibration. Together, these studies suggest that visibility results in better decisions ***and*** decision makers are better calibrated about their decisions. Specifically, when problems are new and somewhat novel, visibility can both improve performance and help decision makers assess their decision performance, combining to improve user calibration and better outcomes.

For researchers, these findings bode well for the continued development of the DSS design theory for user calibration. To the extent that DSS are applied in novel, one-shot situations, this study demonstrates the importance of visibility in DSS design for user calibration. This study also encourages more research into the effects of different forms of expressiveness, visibility, and inquirability on user calibration at different levels of problem novelty.

For builders and designers of DSS, these results clearly highlight the importance of visibility to decision-making and user performance, especially in new, novel decision environments.

This research also highlights the effects of interface design on user calibration. In particular, the results of this research establish the importance of visibility in DSS design, especially for new and relatively novel decision situations.

REFERENCES (See authors for a more detailed list)

APA, "Ethical Principles of Psychologists and Code of Conduct," *Am Psy*, 47 (1992), 1597-1611.

Bauer, M.I. & P.N. Johnson-Laird, "How Diagrams Can Improve Reasoning," *Psy Sci*, 4, 6 (1993), 372-378.

Bray, J.H. & S.E. Maxwell, *Multivariate Analysis of Variance.* Sage Publications: Beverly Hills, 1985.

Card, S.K., J.D. MacKinlay, & B. Shneiderman, *Readings in Information Visualization: Using Vision to Think.* Morgan Kaufmann Publishers: San Francisco,1999.

Clemen, R.T. & A.H. Murphy, "The Expected Value of Frequency Calibration," *OBHDP*, 46, 1 (1990), 102-117.

Gardner, H. Frames of Mind: *The Theory of Multiple Intelligences.* 10th Ed. Basic Books: New York, 1993.

Gonzalez, C. & G.M. Kasper, "Animation in User Interfaces Designed for Decision Support Systems: The Effects of Image Abstraction, Transition, and Interactivity on Decision Quality," *Dec Sci*, 28, 4 (1997), 793-823.

Helstrup, T., "One, Two, or Three Memories? A Problem-solving Approach to Memory for Performed Acts," *Acta Psy*, 66 (1987), 37-68.

Kasper, G. M., "A Theory of Decision Support Systems Design for User Calibration," *ISR*, 7, 2(1996), 215-232.

Kaufmann, G., "A Theory of Symbolic Representation in Problem Solving" *J Mental Imagery*, 9, 2 (1985), 51-70.

Keren, G., "Calibration and Probability Judgments: Conceptual and Methodological Issues," *Acta Psy*, 77 (1991), 217-273.

Larkin, J. and H. Simon, "Why a Diagram is (sometimes) Worth 10,000 Words," *Cog Sci*, 11 (1987), 65-99.

Lichtenstein, S., B. Fischhoff, & L. Phillips, "Calibration of Probabilities: The State of the Art to 1980," in D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, England, 1982.

Marks, D.F., "Visual Imagery Differences in the Recall of Pictures," *Brit J Psy*, 64, 1 (1973), 17-24.

Muir, B.M., "Trust Between Humans and Machines, and the Design of Decision Aids," *IJMMS*, 27, 5 (1987), 527-539.

Murphy, A.H. & R.L. Winkler, "Scoring Rules in Probability Assessment and Evaluation," *Acta Psy*, 34 (1970), 273-286.

Phillips, L. D., "The 'True Probability' Problem," *Acta Psy*, 34 (1970), 254-264.

Rippey, R.M., "A Comparison of Five Different Scoring Functions for Confidence Tests," *J Ed. Measure*, 7 (1970), 165-170.

Russo, J.E. & P. J. Schoemaker, " Managing Overconfidence," *Sloan Mgnt. Rev*, (Winter 1992), 7-17.

Shneiderman, B., *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. (Second Edition) Addison-Wesley: Reading, MA, 1992.

Silver, M.S., "Decision Support Systems: Directed and Nondirected Change," *ISR*, 1, 1 (1991), 47-70.

Speier, C & M.G. Morris, " The Influence of Query Interface Design on Decision-Making Performance," *MISQ*, 27, 3 (2003) 397-423.

Stael von Holstein, C.A., "Measurement of Subjective Probability," *Acta Psy.*, 34 (1970), 146-159.

Winkler, R.L., "Probabilistic Predictions: Some Experimental Results," *JASA*, 66, 336 (1971), 675-685.

When Information Technology Design Favors Form over Function: Where is the Value-Added "Tipping Point"?

Rita M. Vick
University of Hawaii
vick@hawaii.edu

Brent Auernheimer
California State University Fresno
brent@csufresno.edu

ABSTRACT

Performing usability analysis early in the design process results in lower overall development, deployment, and maintenance costs. Pre-development user and task analysis through questionnaires, observation, low-fidelity prototyping, and usability testing enables productive interactive testing of subsequent operable system prototypes. This helps assure a positive return on investment in information technology. When user-centered design assessment is supplanted by assumptions about user, task, and work environment, the result is often production of applications embellished with functionality unrelated to the user's task. Surveys were administered to elicit user perception of system usability and usefulness and of satisfaction with intra-team interaction. This was the first step in determining the relationship between form and function for users of a Synchronous Distributed-Decision Support System (SD-DSS). It was anticipated that the teamwork process would be most troublesome while the SD-DSS would be perceived as easy to use and functional. The reverse proved to be the case.

Keywords

Distributed teamwork, decision modeling, usability, cognitive fit, task analysis.

INTRODUCTION

A frequent assumption of end-user system designers is that creating systems that are usable and easy to use will help guarantee satisfied users and provide organizations with commensurate returns on investments in information technology (IT). This is true provided that designers do not make additional assumptions about what constitutes usability and ease of use in a given work context. Lack of careful user, task, context, and coordination analysis often leads to systems that provide more form than function. In their quest to present products that make a task easy and promise to reward business investment in the technology with competitive advantage, designers may inadvertently embed required functionality behind an interface that is overly simple, adorned with exotic features, or does not fit the cognitive requirements of the task. In either case, the result is a disappointed user and a disappointing return on investment. In the quest for usability, function has been sacrificed for form resulting in inadequate attention being devoted to providing cognitive fit between the user and the task and to assuring that the user receives support for understanding the task. The described study was conducted in order to find out more about how human-computer interaction (HCI) design principles can be leveraged to counter this design trend.

GDSS FUNCTIONALITY

An important aspect of solving complex multicriteria problems is the use of software support to structure the decision process and assist groups of decision makers with assigning preferences and weights to criteria. Decision Support Systems (DSSs) tend to be complex, and the use of software that provides modeling capability can result in a perceived increase in the complexity of the task (Limayem & DeSanctis, 2000). Attempts have been made to design explanation and automated decision guidance into Group Decision Support Systems (GDSSs) to enhance decision models building by providing cues to direct decision makers toward correct structuring and implementation of model components.

Dennis, Haley, and Vandenberg (1996) and Benbasat & Lim (1993) found that, although groups benefit from decision modeling as evidenced by improved decision quality, model building is time consuming and difficult. According to Limayem & DeSanctis (2000), both of those studies determined that use of GDSS technology tended to reduce consensus, decision confidence, and overall satisfaction despite the fact that decision quality improved. Decision makers tend to avoid decision aids because they reveal conflict and place a cognitive load on the user.

In accord with the findings of the present study, Tuttle and Stocks (1997) believe that most software puts too much emphasis on ease of use and too little emphasis on decision maker understanding of the models they are building. The suggested solutions include embedding explanations that require little cognitive effort and provide more problem-structuring support for group cognition. The suggestion has been that cognitive feedback could provide information about preferences and model structure (Te'eni, 1991) by, for example, calling attention to inconsistencies in decision-maker judgments. Bjorkman (1972) suggested that cognitive feedforward might provide explanation before each step of the model-

building process. The premise is that "feedforward" assistance may "attenuate cognitive strain by providing decision makers with information that otherwise would have been learned through feedback." (Limayem & DeSanctis, 2000, p. 388).

Beyond Interface Design

Adaptive Structuration Theory (AST) (DeSanctis & Poole, 1994) maintains that the productive potential of a GDSS is only partly determined by the features designed into the system. More importantly, adoption and continued use of a collaborative technology depends on how the features of the system are applied by users in the work setting. Disengagement between intended and actual use can be caused by the way the system is introduced into the organization, inadequate knowledge of the system's purpose and functionality, or use of the system for unintended (e.g., political) purposes. If this happens, the system will fail to gather a sufficient number of users and will not become an integral part of the organization.

AST was deployed in a case study in a natural work setting to study the technology adaptation process in virtual teams engaged in new product development (Majchrzak, Rice, Malhotra, King, & Ba, 2000). It was found that changes in the alignment of work structures as initially set by the team, flexibility of structures, and occurrence of discrepant events serve to mediate the pre-existing structure/appropriation relationship (Majchrzak et al., 2000, p. 595). This study extended AST to include these mediating factors between existing structures and appropriation of the technology by users.

In order to produce useable systems, designers need to model users' task knowledge and represent this knowledge in a way that provides a good cognitive fit between the user's problem-solving strategies and the nature of the tools provided by the technology. The problem representation and tools must match the characteristics of the task (Umanath & Vessey, 1994; Vessey & Galetta, 1991). Multiple converging design techniques need to be deployed to develop a working understanding of the individual field of practice in order to model the cognitive and interactive nuances that account for what constitutes expert knowledge of a given domain (Potter, Roth, Woods, & Elm, 2000). If the user is not a domain expert, then additional system and training support is required to avoid high rates of error or low usage levels. Usability is a concept that is often misunderstood and so is often oversimplified. It is a complex, multi-faceted concept that represents individual elements of user capability and task demand that impact one another and take on emergent properties in complex work environments. "Usability" must be understood to be inclusive of multiple independent concepts including user satisfaction, system effectiveness, context of use, applied task knowledge domain, and the level of expertise of the user (Frøkjær, E., Hertzum, M., & Hornbæk, K., 2000).

THE STUDY CONTEXT

The author designed and taught a course titled "Computer Supported Collaborative Work (CSCW) in Practice" for four semesters to a total of 74 senior-level undergraduate computer science and engineering students. Class sizes ranged from 18 to 27 students randomly assigned to self-directed teams of three to five participants for the virtual teamwork part of each class meeting. Teams were dispersed throughout a computer lab and communicated only through NetMeeting chat. Lab sessions were 60 to 80 minutes in duration. Teams remained intact throughout the semester and were free to assign members to particular tasks or to work on the task as a group through application sharing, sending files to other team members via file transfer, and accessing information as required from e-mail, the course Web site, or the Internet. Because the nature of the task scenario was complex and did not have a single "right" answer, teams were also free to exercise creativity and critical thinking in pursuit of appropriate responses to the series of variations on the overall task that were presented to them at the beginning of each lab session.

The four courses ran from 10 to 15 weeks in length. Observational, experimental, and survey methods were used to assess group process and outcome. Surveys were administered at the midpoint and at the end of each course to elicit information from participants regarding their satisfaction with their teams and with the software used for communication support (NetMeeting) and the software used for decision modeling (TeamEC™).

Participants generated ideas and determined their relevance, planned the problem solution, determined which criteria were of prime concern, developed alternative ways of meeting the criteria, and assigned weights to each model element using NetMeeting chat. This enabled capture of time-stamped transcripts of team interaction. Participants used the NetMeeting whiteboard to share information, visualize solutions, and as a form of team memory to capture the progress of decision model development.

During the lab sessions, each team assumed a real-life role within an assigned scenario. Roles were rotated so that each team was exposed to each role. The teams were expected to complete a decision model within the timeframe of the lab period. Each class worked consecutively with two scenarios. The first scenario was designed to familiarize the students with working as a team in a simulated distributed Group Decision Support System (GDSS) environment devised by using NetMeeting connectivity to support TeamEC™ as a shared application. This first scenario dealt with evaluation of alternative solutions for an ill-defined policy issue ("How to Revive Hawaii's Economy"). Participants assumed the perspectives of government, business, education, and organized labor. In the second scenario, participants assumed the roles of employees of a "tech startup" company where teams worked as a "task force"

responsible for the design, development and marketing of a collaborative system and assumed corresponding roles.

The first scenario presented a broad policy-based decision problem that was designed specifically to be removed from the technical computer science and engineering learning domain of the participants. This was done to focus students' learning on decision-making as a process and as a particular type of problem solving. Removing learning to a domain in which students were not expected to be expert also enabled experiential learning unencumbered by the need to excel. In this context, students were free to move forward and backward within the problem context as they experimented with learning to think critically about decisions as unique problems. The problem for Scenario 1 was one that is common to all locations and cultures (revitalizing the local economy) so that it could be readily understood by all participants regardless of individual demographic differences. The second scenario focused on a collaborative system design, development, and deployment decision problem specific to the participants' domain of expertise.

The primary measure of group performance was decision model quality. The decision modeling software (Expert Choice, Inc., http://www.expertchoice.com/) is designed for analyzing, synthesizing, evaluating, and justifying complex decisions in a group setting. The software brings structure, organization, and coherence to the decision-making process and supports a multi-objective decision making method based on the Analytic Hierarchy Process (AHP) methodology (Saaty, 1980) in which elements in a non-binary tree structure are subjected to a series of pairwise comparisons to assess their relative value, likelihood, or desirability. It is one of several optimization methods that decision makers can use to reconcile problems having multiple conflicting objectives.

At the end of each scenario, teams competed in a "Face-Off" to determine which team could produce the best decision model. During the lecture portion of the class session immediately following each Face-Off, all the models were discussed, the team with the best model was proclaimed the winner of the competition, and the members of winning team were awarded a small prize. At this time, students completed the surveys that revealed their perceptions of the usefulness and ease of use of TeamEC™ and NetMeeting. A separate survey focused on individual team members' perceptions of and satisfaction with teamwork. The surveys served as benchmarks to gauge team progress (team perceptions of the software support and the effectiveness of their work as a team in solving the assigned problems) and to provide feedback to the instructor. They also provided information intended to help understand more about how the design of interactive collaborative systems helps or hinders the user.

EFFECTIVENESS OF THE GDSS USED IN THE STUDY

It is important to discuss the extent to which TeamEC™, as a type of GDSS software, performed in terms of the above considerations. Although the interface is clear and relatively easy to use, it resulted in a perceived increase in complexity for participants in all four courses. TeamEC™ has embedded explanation into the design of the interface. This is accomplished through online help, explanatory comments that identify functions, and model element definitions. The only automated decision guidance, the "inconsistency ratio" indicator, appears at the end of the pairwise comparison process. This index assists structuring and implementation of model components by indicating whether inconsistent judgments have been entered into the model during the pairwise comparison process. Teams benefited from the guidance provided by these elements as well as the decision modeling process, itself, and all teams' models improved over time.

In the present study, the instructor compensated for lack of built-in feedforward or feedback support in TeamEC™ by (a) being constantly available during lab sessions for consultation and (b) using e-mail to make comments and suggestions to assist teams' understanding of the model-building process. E-mail feedback was sent after each lab session to each team. Since students referenced this e-mail feedback during subsequent lab sessions, the effect was to provide problem-structuring support for group cognition by providing team-specific help with model structuring and content problems as teams progressed. This form of feedback became feedforward assistance since it provided explanation for each subsequent model-building session and so alleviated cognitive strain on teams. Three general types of e-mail were sent to teams: (1) maintenance (file naming/saving, crash recovery), (2) structuring (tree structure validity), and (3) content (tree content validity).

Microsoft NetMeeting was used to simulate synchronous distributed teamwork in a computer laboratory. It provided communication support via text chat and enabled application sharing so that teams members could work simultaneously on collaboratively building decision models. Students communicated only via text chat. Other features of NetMeeting available for use by the teams were the whiteboard, shared clipboard, and file transfer. Teams also had access to course notes on the class Web site and to e-mail for referencing instructor feedback. Web access enabled searches for external information that might assist problem solution. This use of NetMeeting resulted in multiple windows open simultaneously on each participant's desktop.

PARTICIPANT SATISFACTION SURVEY RESULTS

Surveys were administered at the benchmark points (immediately following each Face-Off) of each of the four courses yielding two sets of survey results per course. Students were asked their opinions of TeamEC™,

NetMeeting, and their experiences of working in a team. Results of the surveys are shown in Table 1. IT system use has been found to be strongly correlated to perceived usefulness and perceived ease of use (Mahmood, Hall, and Swanberg, 2001). The survey instrument used for NetMeeting and TeamEC™ was "Measurement Scales for Perceived Usefulness and Perceived Ease of Use" (Davis, 1989), a 7-point Likert scale ranging from "Agree" to "Disagree" or "Satisfied" to "Dissatisfied" where lower scores indicate greater satisfaction. A modified version of an instrument for measuring meeting success (Davison, 1997) was used to assess participants' perceptions of their team's effectiveness.

It had been anticipated that the participants in this study would assign preference for the three major survey subjects in the following order: (1) TeamEC™, (2) NetMeeting, and (3) Teamwork. The assumption was that computer science and engineering students would find the structure and elegance provided by TeamEC™ cognitively compatible with their technical domain of expertise. NetMeeting would also be compatible with their skills and interests, but to a lesser extent because its communication support functions were less straightforward. NetMeeting's whiteboard, used for brainstorming, required original thinking. NetMeeting's chat feature contained the usual inefficiencies associated with text-based communication – time delays causing confusion in sequencing of chat entries and the read-think-respond requirement for expression of thoughts. It was predicted that the least liked aspect of the assignment would be having to work in groups. The participants were unaccustomed to this form of teamwork. Additionally, there were cross-cultural and other demographic differences that teams had to manage. Personality conflicts were a constant concern for the team members.

Survey	NetMeeting	Group Work	TeamEC™
1-1	27.4	25.4	35.4
1-2	23.1	24.4	34.9
2-1	26.4	30.7	40.1
2-2	23.7	25.0	39.6
3-1	22.6	26.5	31.6
3-2	21.9	25.9	31.9
4-1	26.3	25.3	35.8
4-2	21.7	24.4	34.1
Total	193.1	207.4	283.4

Table 1. Survey Results

As can be seen in the Table 1 Totals line, results were contrary to expectations. The most striking aspect of these results is the remarkable similarity across all four classes for all eight survey dates. Without exception, the order of preference was the same: (1) NetMeeting, (2) Group Work, and (3) TeamEC™. The difference between NetMeeting and group work was small, but consistent. The difference between TeamEC™ and both NetMeeting and group work was notably large and also consistent. While participants' opinions of the software support tools were more favorable at the time of the second administration of the surveys (with the exception of the third class, which held an even less favorable opinion of TeamEC™ in the second survey), the parallel results maintained.

On the surface, these survey results do not seem favorable to TeamEC™. A guiding HCI principle is that software must be useful as well as easy to use. An additional often-imposed requirement is that software should make the task more intelligible to the user. If software fails to meet these requirements, it is not likely to be accepted. TeamEC™ suffered from some of these drawbacks. However, there are at least two more specific explanations that mitigate unfavorable response to use of TeamEC™ in this study.

First, the network through which the students shared the TeamEC™ application was frequently unstable. Each team required complex multitasking support to concurrently use external NetMeeting server connectivity, chat, whiteboard, application sharing for TeamEC™, Internet, and e-mail access. The load on the support system was compounded because six teams were working simultaneously from the same lab. Although the network often crashed, and the TeamEC™ client-server application often crashed, NetMeeting was robust. Therefore, chat transcripts and whiteboard records did not fall victim to these regularly occurring episodes. The output of TeamEC™, an independent application, was not automatically archived. While students blamed TeamEC™ when they lost their decision models during a crash, they had repeatedly been advised to frequently save their models. Some teams followed that advice while others forgot. They usually remembered after they experienced their first crash and lost their models.

Second, TeamEC's strongest feature was, from the participants' point of view, the most troublesome. The TeamEC™ software allows fairly wide latitude as to what is inserted into the decision tree. However, the software does give the user an indication of whether all pairwise assessments were made consistently. When there are inconsistencies, the software suggests that model elements be reassessed, although, in some cases, the inconsistencies are not important to the overall outcome. Therefore, reassessment is left to the discretion of the user. In general, participants in this study were unaccustomed to dealing with the level of precision required by the software. The software places high value on fine-grained analysis based on critical thinking skills at a conceptual level that was foreign to most participants.

Results from other studies of GDSS technology use found that GDSSs tended to reduce consensus, decision

confidence, and overall satisfaction despite the fact that decision quality improved. It was earlier noted that decision makers tend to avoid decision aids because they reveal conflict and place a cognitive load on the user. In addition to these negative effects, software often emphasizes ease of use to the detriment of decision makers' understanding of the models they are building.

CONCLUSION

In the present study, the software's design emphasized understanding the task but users did not have a clear conceptual model of how to structure decisions. A mismatch occurred between the software's form and functionality and the users' ability to bridge their own knowledge gap. It is necessary to provide functionality for timely access to information and to present that information in a format that is easy to find and use. It is even more important to assure that task-specific domain knowledge is represented in a way that matches user understanding of a task to avoid misleading the user into believing that the functionality of the system supplants the need for the user to think critically about the task.

It is essential to identify which stakeholders are to benefit from a specific usability analysis (Mayhew and Mantei, 1994). The present study provides evidence that the value-added "tipping point," where form confounds function, may be reached for multiple reasons. An excess of features can lead to confusion. A paucity of features may result in insufficient guidance for novice users. Regardless of cause, if negative outcomes result from deployment of a technology, the lack of return on the investment in the technology will be highly detrimental to the strategic capabilities of the organization.

The study described in this paper was situated in an academic context. In order to carry this work forward, it will be necessary to further analyze distributed team decision making through case analysis of virtual teams in a real-world setting within and across organizations.

REFERENCES

1. Benbasat, I., & Lim, L.-H. (1993). The effects of group, task, context, and technology variables on the usefulness of group support systems: A meta-analysis of experimental studies. *Small Group Research, 24* (4), 430-462.
2. Bjorkman, M. (1972). Feedforward and feedback as determiners of knowledge and policy: Notes on a neglected issue. *Scandinavian Journal of Psychology, 13*, 152-158.
3. Davis, Fred D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly, 13* (3), pp. 319-340.
4. Davison, R. M. (1997). An instrument for measuring meeting success. *Information and Management, 32* (4), 163-176.
5. Dennis, A. R., Haley, B., & Vandenberg, R. J. (1996). *A meta-analysis of effectiveness, efficiency, and participant satisfaction in group support systems research.* 17th International Conference on Information Systems, Dec. 16-18, Cleveland, OH.
6. DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*(2), 121-147.
7. Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proceedings of SIGCHI.* (The Hague, The Netherlands, April 1-6), pp. 345-352. New York: ACM Press.
8. Limayem, M., & DeSanctis, G. (2000). Providing decisional guidance for multicriteria decision making in groups. *Information Systems Research, 11* (4), 386 - 401.
9. Majchrzak, A., Rice, R. E., Malhotra, A., King, N., & Ba, S. (2000). Technology adaptation: The case of a computer-supported inter-organizational virtual team. *MIS Quarterly, 24* (4), 569-600.
10. Mahmood, M. A., Hall, L., & Swanberg, D. L. (2001). Factors affecting information technology usage: A meta-analysis of the empirical literature. *Journal of Organizational Computing and Electronic Commerce*, 11 (2), pp. 107-130.
11. Mayhew, D. J. and Mantei, M. (1994). A basic framework for cost-justifying usability engineering. In R. G. Bias and D. J. Mayhew (Eds.) *Cost-Justifying Usability* (pp. 9-44). Boston: Harcourt Brace & Co.
12. Potter, S. S., Roth, E. M., Woods, D. D., & Elm, W. C. (2000). Bootstrapping multiple converging cognitive task analysis techniques for system design. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 317-349). Mahwah, NJ: Lawrence Erlbaum Associates.
13. Saaty, T. L. (1980). *The analytic hierarchy process*. New York: McGraw-Hill.
14. Te'eni, D. (1991). Feedback in DSS: Experiments with the timing of feedback. *Decision Science, 22*, 644-655.
15. Tuttle, B., & Stocks, M. H. (1997). The effects of task information and outcome feedback on individual insight into their decision models. *Decision Science, 28* (2), 421-442.
16. Umanath, N. S. & Vessey, I. (1994). Multiattribute data presentation and human judgment: A cognitive fit perspective. *Decision Sciences, 25* (5/6), pp. 795-824.
17. Vessey, I. & Galetta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research, 2* (1), pp. 63-84.

A Communication Goals Model of Online Persuasion

E. Vance Wilson
University of Wisconsin-Milwaukee
wilsonv@uwm.edu

Ying Lu
University of Wisconsin-Milwaukee
yinglu@uwm.edu

ABSTRACT

Online communication media are being used increasingly for attempts to persuade message receivers. This paper presents a theoretical model that predicts outcomes of online persuasion based on the structure of primary and secondary goals message receivers hold toward the communication.

Keywords

Computer-mediated communication; influence goals; uses and gratifications theory; interpersonal influence

INTRODUCTION

Persuasion is defined as shaping, reinforcing, or changing the responses of message receivers (Miller, 1980). As asynchronous online communication technologies proliferate across organizations, these are increasingly being used in attempts to persuade.

From the message receiver's viewpoint, online communication blurs the traditional distinction between mass communication, such as magazines and television, where the communication is broadcast to multiple receivers, and interpersonal communication, such as a phone call, where one is directly communicating with another human. In online communication, messages that are broadcast to multiple receivers can be customized to simulate the appearance and interactivity of interpersonal communication. Where receivers perceive these messages to be interpersonal in nature, we anticipate the literature of interpersonal communication will be more relevant than mass communication in predicting persuasiveness, regardless of whether a human sender actually is involved in the manner represented by the messages.

GOALS AND PERSUASION

Goals are cognitive representations of desired results, ranging from biological set points, such as body temperature, to complex, long-term outcomes, such as career success (Austin and Vancouver, 1996). In this paper, we center our focus on two lines of goal research that are particularly relevant to persuasion in an online context: Influence goals and uses and gratifications theory.

The study of influence goals derives from a theoretical observation by Clark and Delia (1979) that three goal types typically are present in interpersonal influence attempts. *Instrumental goals* are directly related to the sender's task. For example, in the case of a message requesting some action on the part of the receiver, the instrumental goal is to gain the receiver's compliance. *Interpersonal goals* are directed toward establishing or maintaining a relationship between the message sender and receiver. *Identity goals* relate to the sender's self-concept, including moral standards, principles, and other internal standards.

Empirical research has supported the conceptual structure of these three influence goals and shown that they significantly predict senders' actions in producing messages (Cody, Canary, and Smith, 1990; Hample and Dallinger, 1987). Dillard (1990) subsequently expanded and refined the goals into a bi-level structure called the Goals-Planning-Action (GPA) model. In the GPA model, primary goals serve to define and drive communication (Schrader and Dillard, 1998), thereby instantiating the instrumental goal type proposed by Clark and Delia (1979). Other goals, referred to as secondary goals, "derive directly from more general motivations that are recurrent in a person's life" (Dillard, Segrin, and Harden, 1989). Secondary goals serve to shape and constrain aspects of the communication. In an empirical test of the relationship between senders' influence goals and production of persuasive messages, Dillard et al. (1989) validated the presence of influence goals and the following secondary goals. *Identity goals* are as originally specified by Clark and Delia (1979). *Interaction goals* concern social appropriateness, including the desire to manage others' impressions of oneself, to avoid threatening or embarrassing others, and to appear relevant and coherent. *Relational resource goals* encompass the personal rewards, emotional support, and other gratifications resulting from participation in the communication and incorporate the interpersonal goals proposed by Clark and Delia (1979). *Arousal management goals* arise from the sender's desire to maintain his or her state of arousal and apprehension about the interaction within tolerable limits and avoid conditions of very high arousal, such as panic or rage.

The initial GPA model validation study by Dillard et al. (1989) finds clear distinctions between the function of primary and secondary goals as well as important unique predictions of several aspects of message production. Influence goals proved to be key predictors of planning

and cognitive effort by subjects in creating persuasive messages. Other aspects of message production predicted by the goals were directness, positivity, and logic of messages.

A subsequent study (Wilson and Zigurs, 2001) suggests that influence goals in the GPA model also provide distinctive predictions of how message senders apply various features of online communication technology to create persuasive messages. Subjects participated in a group planning exercise in which they used a custom online communication system to persuade undecided group members to adopt a negotiating position. Subjects who had the strongest influence goals used fewer special features, i.e., visuals and text formatting functions, suggesting that message content was the critical dimension they considered important to achieving their objective. Subjects with strong identity goals added emphasis to text more frequently than others, using bold, italics, and font controls to highlight and organize their work, suggesting that accurately representing their position, i.e., self-identity, was critical to these subjects. Subjects who had high arousal management goals produced terse, error-ridden messages.

The GPA model has proven useful in understanding and predicting how people approach message creation and what features they decide to use in producing online communication. However, the GPA model does not address message receivers nor does it apply to the context of mass communication.

Applications of goals in modeling persuasion of message receivers are addressed by a second literature stream known as uses and gratifications (U&G) theory (Katz, Blumler, and Gurevitch, 1974). U&G theory addresses the psychological needs and motives of an audience viewing mass-media communication, such as television ads. U&G research takes the approach that message receivers assume an active, goal-oriented role rather than a simple, stimulus-response role in evaluating media. The needs served by communication are considered to be part of the wider ranges of human needs that derive from social and psychological origins. Goals generated by such needs govern expectations of the media, which, in turn, lead people to select specific patterns of media exposure in order to gratify their needs

A substantial literature supports the basic tenets of U&G regarding goal-directed behavior of message receivers, most recently in the context of web surfing (e.g., Eighmey and McCord, 1998; Lin, 1999; Korgaonkar and Wolin, 1999). Numerous U&G typologies have been proposed to categorize receivers' needs and goals. Although these are necessarily biased toward the mass-media contexts in which U&G research is conducted, most of the typologies include categories relating to costs and benefits, and some include categories that substantially overlap goals of message senders described by the GPA model. For example, McQuail, Blumler, and Brown (1972) propose a four-dimension typology that includes categories of diversion (including emotional pleasure and escape from one's regular routine and personal burdens); personal relationships (including substitute companionship as well as social utility); personal identity (including personal reference, reality exploration, and value reinforcement); and surveillance. Categories of personal identity and personal relationships in this typology correspond closely with secondary goals of identity and relational resource presented by Dillard (1990) in the GPA model.

A Communication Goals Model of Online Persuasion

Both U&G theory and the GPA model point to goals as important predictors of persuasion. U&G theory suggests that goals are important to message recipients in determining which messages to view and how to interpret messages. The GPA model proposes that both primary and secondary goals will be present in the context of interpersonal communication. Primary goals motivate both the receiver's decision for viewing (or avoiding) the message and the cognitive effort that the receiver will expend in interpreting and understanding it. U&G research (e.g., McQuail et al., 1972) suggests that message receivers are motivated, generally, by goals of obtaining benefits and avoiding costs related to the message (for brevity, these goals are referenced hereafter as *benefit goals* and *cost goals*).

It also is likely that receivers will share at least some secondary goals from the GPA model where they perceive messages to be interpersonal in nature. Schrader and Dillard write, "Although the GPA model was developed for the purpose of illuminating influence attempts, it can be applied to virtually any sort of interaction" (1998, p. 279). U&G research has shown that goals closely related to the identity and relational resource secondary goals in the GPA model are important to message receivers in mass-media contexts (McQuail et al., 1972). Identity goals are considered to be "explicitly or implicitly present for overt or tacit negotiation in every communicative transaction" (Clark and Delia, 1979, p. 200). Thus, it seems likely that identity will assume an important role in online contexts as well. However, goals involving development of relationships are more problematic in online contexts. Walther and Burgoon (1992) found that relationships are slow to develop in online communication, and Wilson and Zigurs (2001) failed to find any significant effects relating to relational resource goals in their study of online message senders. This suggests that relational resource goals may not be important in situations where there is not an existing relationship and where online messaging utilizes only static presentation (i.e., text and static visuals, vs. audio and video). It is logical that two other secondary goals in the GPA model also may be important to message receivers, although we did not find empirical evidence directly supporting this proposition. The goal of interaction relates to socially appropriate behavior. It is likely that message receivers will have specific interaction goals relating to message content (e.g., regarding

controversial issues), and that these goals will be associated with such factors as perceived message credibility. Similarly, the goal of arousal management also may be anticipated to exist in message receivers (e.g., when requested to do something uncomfortable).

We propose the research model shown in Figure 1. In this model, messages act upon the receiver through a set of cognitive processes in which the receiver's primary and secondary goals toward the communication jointly predict persuasion outcomes. We anticipate that primary goals related to obtaining benefits from the message and avoiding costs of the message will be instrumental determinants of persuasiveness (i.e., persuasion outcomes), and secondary goals of identity, interaction, and arousal management will provide additional distinctive predictions.

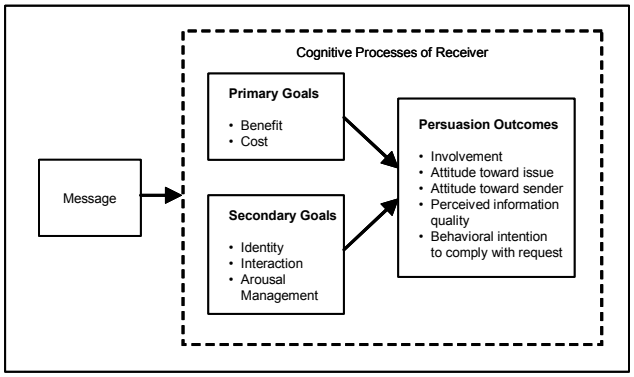


Figure 1. Communication goals research model.

Our first hypothesis tests the construct validity of the goal structure presented in the research model. We propose that message receivers have discrete communication goals regarding incoming messages similar to what has been previously reported in studies of influence goals and message production (e.g., Dillard et al., 1989; Schrader and Dillard, 1998; Wilson and Zigurs, 2001). Construct validation requires the goal measurements to exhibit both discriminant validity, in which measurements of different constructs discriminate among the constructs as predicted by the model, and convergent validity, in which measurements of similar constructs show substantial common association.

H1: Message receivers' benefit, cost, identity, interaction, and arousal management goals will form distinct dimensions.

Our second set of hypotheses relate to primary goals of benefit and cost. The GPA model (Dillard, 1990) proposes that primary goals will be instrumental in determining behavior toward the request. However, the criteria for instrumentality in prior research have been conceptual and qualitative rather than quantitative (e.g., Dillard et al., 1989; Schrader and Dillard, 1998; Wilson and Zigurs, 2001). We propose to test instrumentality using two quantitative criteria. First, we expect primary goals to be *universal* within their domain, (i.e., they will have significant effects on all measures within the area to which they pertain). It is logical that the relationships to persuasion outcomes will be positive for benefit goals and negative for cost goals, leading to two hypotheses:

H2a: Higher levels of benefit goals will increase persuasiveness of the message on all measures.

H2b: Higher levels of cost goals will decrease persuasiveness of the message on all measures.

Second, we expect primary goals to be *prominent*, (i.e., predicting more variance than any secondary goal across measures in their domain). This property is assessed by two additional hypotheses:

H2c: Benefit goals will account for more variance than any secondary goal on all measures.

H2d: Cost goals will account for more variance than any secondary goal on all measures.

Our final hypothesis addresses the role of secondary goals in predicting persuasion outcomes. Secondary goals act to shape and constrain interaction in communication, in effect providing specific, unique predictions of persuasion outcomes beyond predictions provided by primary goals. We did not find prior research that tests the relationship of secondary goals to persuasion outcomes. Due to lack of precedence, we present a single exploratory hypothesis regarding secondary goals, rather than attempting to predict effects relating to specific persuasion outcomes:

H3: Secondary goals of interaction, identity, and arousal management will provide distinctive predictions of persuasion outcomes.

RESEARCH METHOD

The research was conducted as a correlational study using online administration. A custom web application was developed that allowed subjects to log in to the study and view a request asking them to volunteer their time. The web application then collected subjects' responses to a set of open-ended and multiple-choice questions. Subjects were 119 students enrolled in an undergraduate business communication course at a university in the Midwest U.S.

Goals Instrument

Part of the online questionnaire assessed subjects' goals regarding the message. We hypothesized that subjects' primary goals would center on obtaining benefits and avoiding costs associated with the message. New items were written for each primary goal construct. Benefit goal items centered on positive and beneficial perceptions of the message, and cost goal items centered on perceptions of downsides and costs.

Secondary goals of identity, interaction, and arousal management also were assessed. Items for each of these constructs were drawn from a previous questionnaire that was validated initially by Dillard et al. (1989) and subsequently tested in an online communication context by Wilson and Zigurs (2001).

Persuasion Measures

Persuasion outcomes were assessed through measurement of involvement with the communication, attitude toward the sender, attitude toward the message issue, perceptions of information quality of the message, and behavioral intention to comply with the request. Items in the majority of measures were drawn from previously-validated scales: involvement was implemented using a personal involvement instrument (Zaichkowsky, 1994); attitude toward the sender was implemented using a source credibility instrument (McCrosky, 1966); attitude toward the issue implemented related items drawn from Bruner, James, and Hensel (2001); and perceived information quality utilized a scale developed by Moon (1999). New items were written for the behavioral intention measure, centering on self-assessed likelihood that the subject would volunteer to work or donate money toward the cause.

RESULTS

In order to test the structure of goals in the study and establish construct validity of the scales, reliability analysis was conducted on the questionnaire items to assess convergent validity within each of the underlying constructs. Reliability (Chronbach's alpha) of the constructs ranged from .67 to .92. An unconstrained factor analysis then was conducted, and the five-factor structure that emerged clearly supports Hypothesis 1.

To assess the relationships between primary and secondary goals and persuasion outcomes, structural equation models (SEM) were constructed using AMOS 4 software (Arbuckle and Wothke, 1999) to assess which goals significantly predict each persuasion measure. Goals with non-significant predictions were then pruned from each model. Results of SEM analysis, including model fit statistics, are summarized in Table 1.

Benefits goals showed a strong positive relationship with all persuasion measure, predicting 45% or more of the total variance for each of the measures. The effect size of this relationship is much larger than that of any of the other goals, supporting both Hypotheses H2a and H2c. Effects related to cost goals were weaker. Cost goals significantly predicted only involvement, attitude toward the issue, and perceived information quality measures, although for these measures the relationship was stronger than all others except benefit goals. Neither Hypothesis H2b nor H2d is supported by the results.

Secondary goals provided several distinctive predictions beyond those of the primary goals, supporting the exploratory propositions of Hypothesis 3. Arousal management goals predicted involvement and source credibility. Interaction goals predicted attitude toward the issue. Identity goals did not provide unique predictions of any persuasion measure, and none of the secondary goals uniquely predicted perceived information quality or behavioral intention to comply.

Table 1. Distinctive predictions of persuasion measures.

Persuasion Measure	Significant Predictor Constructs	*Total* R^2	Fit Statistics of Pruned Model
Involvement	Benefit goals (β = .75) Cost goals (β = -.25) Arousal management goals (β = -.14)	.69	GFI = .794 AGFI = .731 TLI = .892 RMSEA = .100
Attitude toward issue	Benefit goals (β = .70) Cost goals (β = -.28) Interaction goals (β = .18)	.68	GFI = .848 AGFI = .788 TLI = .914 RMSEA = .090
Attitude toward sender	Benefit goals (β = .67) Arousal management goals (β = -.31)	.55	GFI = .824 AGFI = .770 TLI = .929 RMSEA = .085
Perceived information quality	Benefit goals (β = .713) Cost goals (β = -.31)	.60	GFI = .906 AGFI = .844 TLI = .881 RMSEA = .097
Behavioral intention to comply	Benefit goals (β = .70)	.49	GFI = .971 AGFI = .893 TLI = .951 RMSEA = .101

CONCLUSION

The results generally support our opening proposition that communication goals can provide the basis for effective modeling of persuasion in online communication. Our research model predicted approximately half to two-thirds of the measured variance across an array of persuasion measures, suggesting that the research model represents key cognitive factors within the online persuasion process. It will be important for future research to study aspects of the topic that could not be addressed in the presents study, including investigation of the relationship of communication goals to characteristics of the message, medium, and source, assessment of the stability of communication goals, and articulation of communication goals with other models of persuasion.

The design of this research emphasized exploration and construct validation rather than establishing specific linkages between goals and external factors that are important to practice (e.g., emotional vs. logical appeals). For this reason, implications for practice are necessarily general, although these could become important. The findings suggest that having your message perceived as offering real benefits is critical. False advertising and come-ons are not a substitute for value, as goals are activated to some extent during the communication rather than in advance. Goals related to avoiding costs took a distant second place to obtaining benefits in our findings. Similarly, concerns for social appropriateness (interaction goals) and personal comfort (arousal management goals) suggest that fear appeals will not be particularly effective in online communication. Finally, our findings suggest that moderate changes in content and format of online messages have little effect on persuasion outcomes.

REFERENCES

1. Arbuckle, J., and Wothke, W. (1999). *AMOS 4.0 user's guide.* Chicago, IL: Smallwaters Corporation.
2. Austin, J. T., and Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin, 120*(3), 338-375.
3. Bruner, G., James, K. E., and Hensel, P. J. (2001). *Marketing scales handbook.* Chicago: American Marketing Association
4. Clark, R. A., and Delia, J. G. (1979). Topoi and rhetorical competence. *The Quarterly Journal of Speech, 65,* 187-206.
5. Cody, M. J., Canary, D. J., and Smith, S. W. (1990). Compliance-gaining goals and episodes. *Communicating Strategically.* Hillsdale, NJ: Lawrence Erlbaum.
6. Dillard, J. P. (1990). A goal-driven model of interpersonal influence. In J. P. Dillard (Ed.), *Seeking compliance: The production of interpersonal influence messages* (pp. 41-56). Scottsdale, AZ: Gorsuch Scarisbrick.
7. Dillard, J. P., Segrin, C., and Harden, J. M. (1989). Primary and secondary goals in the production of interpersonal influence messages. *Communication Monographs, 56,* 19-38.
8. Eighmey, J., and 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
9. Hample, D., and Dallinger, J. M. (1987). Individual differences in cognitive editing standards. *Human Communication Research, 14*, 123-44.
10. Katz, E., Blumler, J. G., and Gurevitch, M. (1974). Utilization of mass communication by the individual. In J. G. Blumler and E. Katz (Eds.), *The uses of mass communications: Current perspectives on gratifications research* (pp. 19-32). Beverly Hills, CA: Sage.
11. Korgaonkar, P., and Wolin, L. (1999). A multivariate analysis of web usage. *Journal of Advertising Research, 39*(2), 53-68.
12. Lin, C. A. (1999). Online-Service adoption likelihood. *Journal of Advertising Research, 39*(2), 79-89.
13. McCrosky, J. C. (1966). Scales for the measurement of ethos. *Speech Monographs, 3,* 65-72.
14. McQuail, D., Blumler, J. G., and Brown, J. R. (1972). The television audience: A revised perspective. In D. McQuail (Ed.), *Sociology of mass communications*. Hammondsworth: Penguin.
15. Miller, G. R. (1980). On being persuaded: Some basic distinctions. In B. E.Rolloff and G. R Miller (Eds.), *Persuasion: New directions in theory and research* (pp. 105-152). New York: Harper and Row.
16. Moon, Y. (1999). The effects of physical distance and response latency on persuasion in computer-mediated communication and human-computer communication. *Journal of Experimental Psychology: Applied, 5*(4), 379-392.
17. Schrader, D. C., and Dillard, J. P. (1998). Goal structures and interpersonal influence. *Communication Studies, 49*(4), 276-293.
18. Short, J., Williams, E., and Christie, B. (1976). *The social psychology of telecommunications.* London: Wiley.
19. Sproull, L., and Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science 32*(11), 1492-1512.
20. Walther, J. B., and Burgoon, J. K. (1992). Relational communication in computer-mediated interaction. *Human Communication Research, 19* (1), 50-88.
21. Wilson, E. V., and Zigurs, I. (2001). Interpersonal influence goals and computer-mediated communication. *Journal of Organizational Computing and Electronic Commerce, 11*(1), 59-76.
22. Zaichkowsky, J. L. (1994). The personal involvement inventory: Reduction, revision, and application to advertising. *Journal of Advertising, 23*(4), 59-70.

The "Voice Effect" in Groups

Tom L. Roberts
University of Kansas
troberts@ku.edu

Paul Benjamin Lowry
Brigham Young University
Paul.Lowry@byu.edu

ABSTRACT

This study looks at how collaborative technology, proximity choices, and group size can affect voicing in groups. Results of the study, involving two experiments with 550 participants, show that collaborative technology can improve an individual's desire to voice, instrumental motives to voice, non-instrumental motives to voice, and the opportunity to voice in face-to-face groups. The results also show that the use of collaborative technology can lesson individual voice losses as groups increase in size especially in distributed environments. These findings have important implications in group interactions using technology.

Keywords

Human Computer Interaction , voice effect, collaboration

INTRODUCTION

For collaboration to be successful, effective communication between group members is crucial. The importance of communication increases when the exchange of information in verbal or electronic discussions is imperfect. Members must first decide to contribute the information and then have the opportunity to contribute it. Individuals' motivations to voice their opinions may greatly vary within the context of a given work situation or environment (Dennis, Hilmer, and Taylor, 1998). Although businesses have begun to use collaborative technology to improve communication, the HCI impact of collaborative technology on information sharing activities is unclear. Some studies find collaborative technology to enhance information sharing within groups (Dennis, 1996A); others find no effects (Mennecke and Valacich, 1998); others find inhibited information sharing (Hightower and Sayeed 1996).

Given these issues, several research opportunities related to HCI and collaboration emerge. Additional research is needed to study media conditions and social factors that influence how groups perceive and use technology, and the social structures created by collaborative technologies (Yoo and Alavi, 2001).

LITERATURE

Voice effect is the notion that having the opportunity to provide input on a decision will enhance judgments of process fairness (Folger, 1977). Alternative explanations for the *voice effect* are grouped as *instrumental and non-instrumental explanations*.

Instrumental explanations claim that voice enhances procedural justice because individuals assume that expressing their views will increase the chances for favorable outcomes. The instrumental perspective explains that voice enhances perceptions of procedural justice because participants hope to influence decision makers to enhance the likelihood of favorable outcomes (Brett, 1986).

Non-instrumental explanations focus on informational and symbolic results of procedures (Lind, Kanfer, and Early, 1990) rather than on the ability of procedures to enhance instrumental benefits for voicing individuals. This perspective attributes the voice effect to desires by participants to express their opinions and be listened to, regardless of outcomes (Tyler, 1987). Non-instrumental motives to voice can be divided into two related constructs (Barry and Shapiro, 2000): *Non-instrumental motives to express opinions*, which is the desire to express feelings to a group to feel better, regardless of the outcome; and (2) *non-instrumental motives to vent*, which is the desire to vent opinions, regardless of the outcome.

The *desire to voice* reflects on one's motivation to participate in group processes. A large part of this desire is whether group members believe that they can potentially influence group outcomes (Barry and Shapiro, 2000). Although the impact of voicing opinions likely differs depending on a group's context, one's desire to voice opinions should not vary significantly within a given context.

Opportunity to voice is defined as to the degree to which a group allows group members to express their opinions before decisions are made (Barry and Shapiro, 2000).

Social presence is defined as "the degree to which [a] medium facilitates awareness of the other person and interpersonal relationships during the interaction" (Fulk et al. 1990, p. 118). Most studies have operationalized social presence from low to high (Miranda and Saunders, 2003). Electronic and paper-based communication media are generally viewed as low in social presence, while FtF communication is viewed as high in social presence (Miranda and Saunders, 2003).

Variations in social presence occur through both proximity choices and media choices—as distributed groups naturally have less presence than FtF groups.

HYPOTHESES

FtF groups tend to have high social presence while distributed groups have low social presence (Miranda and Saunders, 2003). The results of distributed groups tend to be diminished by having less media richness and socialization than FtF groups (Burke, and Chidambaram 1999).

Although FtF work is superior to distributed work in terms of social presence, it is not necessary superior in all aspects. Several dozen potential process losses are typical in FtF groups (Nunamaker et al., 1991); however, most research has focused on evaluation apprehension, domination, and production blocking—all of which affect voice.

Evaluation apprehension occurs when group members withhold ideas because they fear the ideas they suggest may be criticized or ridiculed by other group members (Diehl and Strobe, 1987), and is stronger in FtF groups. *Domination* occurs when a group member forces his/her will upon other group members (Nunamaker et al., 1991), which is also stronger in FtF groups. *Production blocking* occurs when only one member can communicate at once, which causes the suppression or forgetting of group members' ideas; all of which can require one to focus on remembering a particular idea, while waiting to express it to the group, rather than creating new ideas; and may cause one to listen closely to the ideas of others, rather than creating new ideas (Diehl and Strobe, 1987). This can occur more in FtF because domination is more likely, and dominant people cause group production blocking. Although distributed groups have less social presence than FtF groups, this limitation will likely be counterbalanced by having fewer negative effects from evaluation apprehension, domination, and production blocking than FtF groups.

H1A: The *desire to voice* will be similar for group members , regardless of proximity.

H2A: *Instrumental motives* for voicing will be similar for group members, regardless of proximity

H3A: *Non-instrumental motive*s of *expressing opinions* will be similar for group members, regardless of proximity.

H4A: *Non-instrumental motives* of *venting* will be similar for group members, regardless of proximity.

More social presence in FtF groups should directly translate into more opportunity to voice, because there are more opportunities for interactivity and greater communication bandwidth.

H5A: The *Opportunity* to voice will be greater for F-t-F group members than for dispersed group members.

Collaborative software can have more social presence due to media richness improvements that include support for parallelism, anonymity, group memory (Zigurs and Buckland, 1998), and group awareness (Lowry and Nunamaker Jr. 2003).

Parallelism is the ability of group members to contribute information simultaneously without waiting for other group members (Dennis, Wixom, Vandenberg, 2001). *Parallelism* mitigates *production blocking* (Gallupe et al., 1994) by creating more equal participation (Dennis and Garfield, 2003).

Anonymity enables group members to contribute to group discussions and collaborations without being identified, and often increases motivation of individual group members to participate (Dennis, Wixom, Vandenberg, 2001). Without anonymity, participants may withhold ideas or comments due to *evaluation apprehension* (Diehl and Strobe, 1987) or may conform to the group majority or leaders' views (Hackman and Kaplan, 1974). *Anonymity* may alleviate conformance by shielding a contributor from a group's reactions (Hayne and Rice, 1997). *Anonymity* can reduce the reluctance of group members to challenge the views of others (Nunamaker et al. 1991).

Collaborative software has been shown to increase group participation. Teams are more participative when those in power choose to listen to and act on a team's interactions, and collaborative software generally increases participation (Dennis and Garfield, 2003). This occurs because of equality provided by anonymity (Dennis and Garfield, 2003) and being able to work in parallel (Dennis et al., 1999).

H1B: The *desire* to voice will be greater in groups using collaborative tools than in non-collaborative tool groups.

H2B: *Instrumental motives* for voicing will be greater in groups using collaborative tools than in non-collaborative tool groups.

H3B: *Non-instrumental motives* of *expressing opinions* will be greater in groups using collaborative tools than in non-collaborative tool groups

H4B: *Non-instrumental motives* of *venting* will be greater in groups using collaborative tools than in non-collaborative tool groups.

H5B: The *opportunity* to voice will be greater in groups using collaborative tools than in non-collaborative tool groups.

Small groups tend to have more social presence than large groups. Increased group size has been shown to increase process losses in verbally interacting groups, either exponentially (Steiner, 1972) or linearly (Bouchard and Hare, 1970). The number of ideas contributed per person decreases sharply as group size increases (Steiner, 1972). Group research involving heuristic evaluation performed with non-collaborative software concludes the optimal

team size for HE is three to five people (Nielsen and Landauer 1993). In this scenario, teams larger than three to five members often report too many duplicate usability issues, have difficulties coordinating, and fail to find enough additional usability issues to justify size increases (Nielsen and Landauer 1993).

Much of the losses that occur as groups increase in size can be attributed to process losses such as *evaluation apprehension* and *production blocking* (Nunamaker et al., 1991). These phenomena should decrease instrumental motives to voice and opportunity to voice. Likewise, a similar decrease should be seen in non-instrumental motives to voice opinions and to vent.

H1C: The *desire* to voice will be greater in groups of three than similar groups of six.

H2C: *Instrumental motives* for voicing will be greater in groups of three than similar groups of six.

H3C: The non-instrumental motive of *expressing opinions* will be greater in groups of three than similar groups of six.

H4C: The non-instrumental motive of *venting* will be greater in groups of three than similar groups of six.

H5C: *Opportunity to voice* will be greater in groups of three than similar groups of six.

METHOD

Task / Tools

Participants were asked to perform a heuristic evaluation (HE) task. HE is a group-oriented usability evaluation technique and was chosen because it is efficient, economical, easy for non-experts to understand and perform, and is most effective when performed in group settings (Nielsen and Molich, 1990). The purpose of HE is to evaluate quickly the usability of a system's interfaces during software development, using heuristics for software usability. The evaluation task included evaluating a website and categorizing software bugs. Word™ was chosen as the non-collaborative tool for the control groups. Collaboratus was chosen for conditions B and C because it supports both FtF and distributed group work.

Treatments

The design of the experiment involved a three-way ANOVA with a 2x2x2 design. The three manipulated conditions include *proximity* (FtF vs. distributed), *tool use* (non-collaborative software, Word™, vs. collaborative software, Collaboratus), and *group size* (three people versus six people).

The control groups performed HE FtF using traditional processes; conducting step one of HE in parallel without awareness of other group member's work. Instead, they recorded individually their bugs using Word™ without knowledge of what bugs other group members were submitting. In step two, control groups discussed FtF the bugs they found and combined them into one document in Word™.

The first treatment performed HE FtF in step one using Collaboratus. This tool allowed participants to see the contributions of others, but did not allow for any direct communication. In step two, the first treatment groups discussed their bugs FtF and combined them into one document in Collaboratus.

The second experimental treatment performed HE in step one in a distributed-synchronous work mode using Collaboratus. Just like the FtF Collaboratus treatment, these groups had no explicit communication capabilities in step one. In step two, these distributed treatment groups had to discuss their bugs and consolidation using the chat features of NetMeeting™.

Participants

The participants were all members of a 200-level IS class at a large Midwestern university. 300 students were enrolled in the course over two semesters. 550 students volunteered for the two experiment sessions. The first session was conducted with three-member teams. The second session was conducted using six-member teams. In total, 512 students participated, however, 97 of these participants' data was subsequently dropped. 415 students provided demographic data: age (M=20.2, SD=1.9); GPA (M=3.3, SD=.46), years of education (M=13.7, SD=1.2.); gender (57.5 % male, 42.5% female).

Procedures / Measures

All students were given training on HE in class. Next, students attended their assigned laboratory sessions, where their assigned conditions were executed. A given lab session was dedicated to only one condition. None of the participants were allowed to talk during Step One, and only the control groups and FtF Collaboratus groups were allowed to communicate orally during Step Two. The same facilitator and assistants oversaw each session. All aspects of the session were scripted, timed, and read carefully by the facilitator. Table 1 shows the measures used to evaluate voice.

Study Measurements	Alpha
Desire to Voice	.6341
Instrumental Motives Voice	.7996
Expressing Opinions	n/a.
Venting	n/a.
Opportunity	n/a.

Table 1. Measures and Alphas

All are from (Barry and Shapiro, 2000), except opportunity from (Tyler, 1994)

ANALYSIS

The method of analysis was three-way ANOVA on each DV, with proximity, tool, and size as the IV's with alpha=.05. Multiple comparisons were conducted using a Tukey's procedure.

DISCUSSION OF RESULTS

No significant differences were measured between FtF groups and distributed groups in terms of desire to voice, instrumental motives, expressing, and venting. These results suggest that proximity has no real bearing on desire and motives related to voice. The results also show that FtF groups provide greater opportunity to voice than distributed groups. Yet, large distributed groups have greater voice opportunity than traditional large FtF groups; and even greater opportunity is given to FtF groups using collaborative tools than traditional FtF groups at both sizes. This supports the claim that negative processes losses that often occur in FtF groups may be alleviated by collaborative tools.

No significant differences were shown between non-collaborative-tool groups collaborative-tool groups, in terms of desire and instrumental motives. However, predictions were confirmed that collaborative software would increase expressing, venting, and opportunity. This suggests that participants' desire to voice is too ingrained in one's self-concept to be affected by tool choices, and that participants did not believe that collaborative software would give them more power to influence their groups. However, participants did feel that collaborative software empowered them to express themselves; even though they did not believe they would greatly influence outcomes. Collaborative software also allowed participants more voice opportunity; likely because of parallel work, group awareness, and anonymity.

Participants in large groups had less desire to voice, less expressing, less venting, and fewer opportunities to voice. These results indicate that increases in group size are detrimental to these voice constructs. Finally, there were no significant differences between large and small groups in terms of instrumental motives.

The contribution of this research is to show how variations in social presence in groups (manipulated through proximity, tools, and group size) affect desire to voice, instrumental motives, non-instrumental motives, and opportunity. We showed that distributed work does not negatively affect desire to voice, instrumental- and non-instrumental motives, and that large distributed groups using Collaborative tools had more opportunity to voice than large FtF groups not using Collaborative tools. This provides evidence that distributed work may be more viable than previously believed, when conducted with collaborative software.

Our results also clarify the relationship between desire to voice, instrumental- and non-instrumental-motives. The results of comparing collaborative software teams with traditional software teams suggest that, since desire and instrumental motives remained constant while non-instrumental motives increased, there are additional factors that affect an individual's desire to voice. It appears these additional factors decreased the effect of the increase in non-instrumental motives to voice, so that overall desire remained the same.

Our results show that collaborative tool use is directly related to increased non-instrumental motives, venting, and expressing. This provides a unique understanding and new set of benefits to collaborative software use. Collaborative software may therefore provide distributed groups with the tools and structures needed to provide practical alternatives to FtF interaction, especially in activities which require high levels of participation among group members. The key to collaborative software effectiveness is a well-designed interface. The interface provides the means under which group awareness, parallelism, anonymity, and group memory are provided so that social presence can be increased.

FUTURE RESEARCH

Given the limited generalizability of these findings, several streams of research should be conducted. Research could explore the applicability of the results in real business settings through field research or through controlled laboratory studies with usability experts working on systems that have specific business purposes. Replication of this experiment with varying levels of expertise and different screens and tasks would also be helpful. It could also be useful to explore the social presence, and subsequent effects on voice, of asynchronous-distributed (AD) settings.

CONCLUSION

As work with collaborative technology becomes more prevalent, there is an increased need for understanding how such technology can affect team interactions. This study has demonstrated that appropriate choices on technology, proximity, and group size significantly increase the social presence among group members which positively affects the motivations and opportunity of members to voice their opinions. Increased voice helps members feel more satisfied with the group outcome and is associated with increased productivity. Future research should continue to explore ways to improve social presence and voice effects in HCI environments.

REFERENCES

1. Barry B. and Shapiro, D. (2000) When will grievants desire voice?: A test of situational, motivational and attributional explanations," *International Journal of Conflict Management*, 11,2, 106-134.
2. Bouchard, T. and Hare, M. (1970) Size, performance, and potential in brainstorming groups, *Journal of Applied Psychology*, 54, 51-55.

3. Brett, J. (1986) Commentary on procedural justice papers." In R. Lewicky, M. Bazerman, and B. Sheppard (Eds.), *Research on Negotiation in Organizations* (Vol. 1, pp. 81-90). Greenwich, CT: JAI Press.

4. Burke, K. and L. Chidambaram. (1999). How much is enough? A longitudinal examination of media characteristics and group outcomes, *MIS Quarterly*, 23,4, 557-580.

5. Dennis, A. (1996) Information processing in group decision making: You can lead a group to information, but you can't make it think, *MIS Quarterly*, 20,4, 433-458.

6. Dennis, A. and Garfield, M. (2003). The adoption and use of GSS in project teams: Toward more participative processes and outcomes, *MIS Quarterly*, 27, 2, 289-323.

7. Dennis, A., Hilmer, K. and Taylor, J. (1997-1998). Information exchange and use in GSS and verbal group decision making: effects of minority influence, *Journal of Management Information Systems*, 14, 3, 61-88.

8. Dennis, A., Wixom, B., and Vandenberg, R. (2001). Understanding fit and appropriation effects in group support systems via meta-analysis, *MIS Quarterly*, 25, 2), 167-193.

9. Folger, R. (1977). Distributive and procedural justice: Combined impact of "voice" and improvement on experienced inequality, *Journal of Personality and Social Psychology*, 35, 108-119.

10. Fulk, J., Schmitz, J., and Steinfield, C. (1990). A social influence model of technology use, in J. Fulk and C. Steinfield (Eds), *Organizations and Communication Technology*, Sage Newbury Park, CT, 117-142.

11. Gallupe, R., Bastianutti, L., and Cooper, W. (1994). Unlocking brainstorms, *Journal of Applied Psychology*, 76, 1, 137-142.

12. Hackman, J. and Kaplan, R. (1974). Interventions into group processes: An approach to improve the effectiveness of groups, *Decision Sciences*, 5, 3, 459-480.

13. Hayne, S. and Rice, R. (1997). Attribution accuracy when using anonymity in group support systems, *International Journal of Human Computer Studies*, 47, 3, 429-450

14. Hightower, R. and Sayeed, L. (1996). Effects of communication mode and prediscussion information distribution characteristics on information exchange in groups, *Information Systems Research*, 7, 4, 451-465.

15. Gigone, D. and Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment, *Journal of Personality and Social Psychology*, 65, 5, 959-974.

16. Lind, E., Kanfer, R. and Early, C. (1990). Voice, control, and procedural justice: Instrumental and non-instrumental concerns in fairness judgments, *Journal of Personality and Social Psychology*, 59, 952-959.

17. Lowry, P. and Nunamaker, J. (2003). Using Internet-based, distributed collaborative writing tools to improve coordination and group awareness in writing teams, *IEEE Transactions on Professional Communication*, In Press.

18. Mennecke, B. and Valacich, J. (1988). Information is what you make of it. The influence of group history on information sharing, decision quality, and member perceptions, *Journal of Management Information Ssystems*, 15, 2, 173-198.

19. Miranda, S. and Saunders, C. (2003). The social construction of meaning: An alternative perspective on information sharing, *Information Systems Research*, 14, 1, 87-106.

20. Nielsen, J. and T. Landauer. (1993). A mathematical model of the finding of usability problems, *Proceedings of the Interchi*, 206-213.

21. Nielsen, J. and Molich, R. (1990). *Heuristic evaluation of user interfaces.* Paper presented at the Computer Human Interaction (CHI).

22. Nunamaker , J., Dennis, A., Valacich, J., Vogel, D., and George, J. (1991). Electronic meeting systems to support group work, *Communications of the ACM*, 34, 7, 40-61.

23. Steiner, I. (1972). *Group process and productivity*, New York: Academic Press.

24. Tyler, T. (1994). Psychological models of the justice motive: Antecedents of distributive and procedural justice, *Journal of Personality and Social Psychology*, 67, 850-863.

25. Tyler, T. (1987). Conditions leading to value expressive effects in judgments of procedural justice: A test of four models, *Journal of Personality and Social Psychology*, 62, 333-344.

26. Yoo, Y. and Alavi, M. (2001). Media and group cohesion: Relative influences on social presence, task participation, and group consensus, *MIS Quarterly,* 25, 3, 371-390.

27. Zigurs, I. and Buckland, B. (1998). A theory of task-technology fit and group support systems effectiveness, *MIS Quarterly*, 22, 3, 313-334.

An Empirical Investigation of Antecedents of Internet Abuse in the Workplace

Dennis F. Galletta
University of Pittsburgh
galletta@katz.pitt.edu

Peter Polak
University of Miami
ppolak@miami.edu

ABSTRACT

This study examined the extent to which employees engage in Internet abuse, and whether any of 15 antecedents predict the amount of that abuse. Data were collected from 571 Usenet users in an on-line survey. Aggregating the time for each of the eleven listed methods of Internet abuse revealed a total of 5.8 hours per week, on average. Most of the antecedents in two of the three Theory of Planned Behavior (TPB) categories (Attitudes and Subjective Norms), were significant, and none of the antecedents in the third TPB category (Perceived Behavioral Control) showed significance. addiction, self-justification, job satisfaction, peer culture, and supervisor culture were significant predictors of Internet abuse. Exploratory demographic factors computer experience, gender, and firm revenue also showed predictive power.

Keywords

Cyberloafing, Internet Abuse, Theory of Planned Behavior, Attitudes, Subjective Norms

INTRODUCTION

According to the latest government study, the personal computer has become the "hallmark of the workplace in postindustrial America" (Hipple & Kosanovich, 2003). Almost 80% of managerial and professional workers have access to a personal computer at work, and nearly 66% use the Internet at work.

The widespread use reflects many business benefits (Vogt, 1997), but the "double-edged sword" (Lim, 2002, p. 676) that is the Internet unfortunately carries with it extra "baggage" that accompanies those benefits. Several recent studies reveal much abuse of the Internet in the workplace by employees; users exchange personal e-mails, shop on line, check scores on sporting events, gamble on line, view pornographic material, and chat on instant messaging services.

In the most recent U.S. study (Colby and Parasuraman, 2002), it is estimated that employees spend between 3.7 and 6.5 hours per work week on personal Net use. Earlier studies (Lim, 2002) revealed that between 64% and 90% of U.S. workers engaged in personal activities while at work. Financial losses from this abuse have been estimated to reach 64% of organizations, costing $378 million in 2001 (Computer Security Institute, 2001).

In this study we place Internet abuse into a framework that examines the antecedents of that abuse, and report on a study of 571 Usenet users.

BACKGROUND

The Computer Security Institute reported that 31% of businesses said they had experienced financial losses from reduced productivity as a result of employee misuse of Internet privileges (McCollum, 1998).

Internet Access and Productivity

There are two views about the effect of Internet abuse on productivity. One asserts that productivity suffers due to wasted time, and the other takes the opposite view, stating that employees need breaks to recharge their creative potential and relax while performing their duties, leading especially to improved team-building and communication (Guthrie and Gray, 1996).

While it is difficult to pinpoint the net result of gains in creativity and restfulness pitted against potential problems such as the waste of time (McCollum, 1998), reduced bandwidth, legal exposure (Manhasset, 1997), or ethical issues (Lee et al., 2002), most researchers make the assumption that there is a net negative effect. This study attempts to examine possible antecedents of Internet abuse for greater understanding of the problem.

RESEARCH MODEL AND EXPECTATIONS

Several factors can lead to Internet abuse in the workplace. Informal discussions with workers have revealed that some are unable to spend such time because they are not interested, they are too conscientious, they are in full view of others, they are too busy, etc.

This study builds on previous research by looking at potential factors that influence Internet abuse in the workplace. According to previous studies, several candidate factors lead to Internet browsing by employees. Using the Theory of Planned Behavior (Ajzen, 2001), Lee and Lee (2002) arranged some potential factors into those relating to attitudes, subjective norms, and perceived behavioral control (see Table 1). We expanded the list of possible factors in each of the areas as described below,

and augmented this list with other exploratory factors. While this study does not test the statistical appropriateness of where each of our measures fit, we present them in this structure for greater understandability. Each will be discussed in turn.

Attitudes	Perceived Beh. Control
• Job satisfaction • Playfulness • Engagement • Internet Addiction • Self-Justification	• Abuse Policy • Workplace Privacy • Productivity Measurement • Work Monitoring
Subjective Norms	**Demographic Variables**
• Peer Culture • Supervisor Culture	• Gender • Age • Exp. with Computers • Exp. with the Internet • Size of the Firm

Table 1. Antecedents of Internet Abuse

Attitudes

In a conceptual paper, Lee and Lee (2002) identified attachment, involvement, commitment, and beliefs as several attitude factors that can be powerful determinants of employees' willingness to commit computer abuse. An empirical study by Stanton (2002) examined the relationships between several dimensions of job attitudes and the frequency of Internet use.

Several possible measures can be used to address attachment, involvement, commitment, and beliefs. We examined playfulness and engagement in this study. The first of the two variables, playfulness, was measured by adapting an instrument from Webster and Martocchio (1992). The second, related measure of engagement (Webster & Ho, 1997) is concerned with users' subjective experiences of pleasure and involvement due to their intrinsic interest. It is expected that people scoring high on the playfulness and engagement measures will generally tend to abuse the Internet to a greater degree.

In addition to those measures, we asked users if they believed they were addicted to the use of the Internet. This exploratory item was used so that we could focus on one important aspect that might go beyond playfulness and engagement (Stanton, 2002).

Job satisfaction, Stanton's (2002) main variable of interest with respect to the frequency of Internet use, presents many possibilities as an antecedent of Internet abuse. One heavily used and validated scale is the Job Satisfaction Survey (JSS) (Spector, 1997), including fringe benefits, communications, operating procedures, co-workers, pay, promotion, contingent rewards, supervision, and the nature of the work itself. Stanton's study showed that most of the dimensions seemed to follow the pattern that lower job satisfaction led to heavier Internet use, perhaps due to users' detachment with aspects of their jobs and desire to disengage by substituting other activities.

Another exploratory item was added to address self-justification of the activity (Lim, 2002). We asked users if they believed their rewards matched their efforts at work. It was expected that both addiction and feelings of inadequate rewards would lead to more Internet abuse.

Our expectations are that:

H1: Attitudinal factors affect the extent of Internet abuse:

H1a: Lower job satisfaction will promote Internet abuse.
H1b: Computer playfulness will promote Internet abuse.
H1c: Engagement will promote Internet abuse.
H1d: Internet addiction will promote Internet abuse.
H1e: Perceived inequity will promote Internet abuse.

Subjective Norms

Lee and Lee (2002) provide two subjective norms, "co-workers influence" and "seniors influence." In this study, we examine both dimensions but focus the latter on the user's supervisor. We developed items that asked if personal Internet activity was seen as appropriate by peers and by supervisors, respectively.

Subjective norms have been powerful determinants of behavior in previous studies. The organizational behavior literature has for many years shown powerful effects of norms on worker behavior (for example, Milgram, 1965). In the Marketing literature, it is well known that a consumer's expectations are influenced more by peers than any other factor (Webster, 1991). A previous study (Galletta et al., 1995) brought this to the realm of information systems by examining training in a new software package: subjects were reliably influenced to reject a new package by their peers.

H2: Subjective norms affect the extent of Internet abuse:

H2a: Supportive peer culture promotes Internet abuse.
H2b: Supportive supervisor culture promotes abuse.

Perceived Behavioral Control

The final category of TPB is perceived behavioral control, described by Lee and Lee (2002) as the perceived ease or difficulty of performing a particular task. Organizations can limit undesirable activities by imposing policies, monitoring work, placing workers in publicly-visible settings, and deploying strict productivity measurement.

Work monitoring and lack of workplace privacy are related and strong limiting mechanisms on abusive behavior. Previous studies have determined that workers will be motivated to engage in social loafing when they think that their behavior is not being monitored (Jones, 1984). Workplace privacy varies by employee, and it is expected that employees with full privacy and without monitoring of their behavior will tend to be more abusive of the Internet than employees working in full view of others and with their use monitored by a supervisor.

To our knowledge, no previous studies have addressed the potential hindering effects of productivity measurement

on Internet abuse. While it is perhaps somewhat obvious that having privacy will provide a great deal of freedom to the Internet abuser, it is more subtle to consider the measurement of worker output. Some workers have their output measured very precisely in terms of lines of code, keystrokes, or customer service call quotas. The more objective, short-term in focus, and clear the measurement of a person's productivity, the less he or she will be able to abuse the Internet, even if they are tucked away with near invisibility. Theoretical grounding for the output measurement factor can be found in the area of social psychology (Williams, et al., 1981). When people's outputs are unidentifiable, they are less motivated to perform well because they can "get away" with less work without being criticized or punished.

We created original three-item scales to address the existence of limiting factors such as policies against Internet abuse, measurement of work output, monitoring of Internet traffic, and lack of privacy. We expect that:

H3: Perceived behavioral controls hinder Internet abuse:

H3a: More restrictive policies will hinder Internet abuse.
H3b: Workplace privacy will promote Internet abuse.
H3c: Productivity measurement will hinder abuse.
H3d: Monitoring of traffic will hinder Internet abuse.

Demographic Variables

Several demographic measures were also developed, without well-formed expectations about their effects on Internet abuse. Such factors have not been examined closely, so we treated them as exploratory in this study; hypotheses are stated in null form:

H4: Demographic factors will not affect Internet abuse:

H4a: Gender will not affect Internet abuse.
H4b: Age will not affect Internet abuse.
H4c: Computer experience will not affect Internet abuse.
H4d: Internet experience will not affect Internet abuse.
H4e: Firm size will not affect Internet abuse.

PROCEDURE

All of the items described above were assembled into an on-line instrument that contained 106 items. Completion of the entire instrument required about 15-20 minutes.

A short invitation to participate in the study was sent to a large number of Usenet newsgroups. It was difficult to determine the exact number of newsgroups; however, it is estimated that about 3,000 messages were placed.

Over a period of a week, 835 completed surveys were received. In our data set, however, we found many incomplete or otherwise unusable entries. Two judges reviewed the entries separately and flagged all that should be removed. Pooling the results, we deleted 264 entries, the union of the two lists, for a final sample size of 571.

Our sample provides for a useful test of our list of antecedents for several reasons. First, tapping only Usenet provides a set of users that is quite homogeneous, limiting the error variance in our analysis. Also, Usenet users have specialized interest, knowledge, and experience that is not shared by most Internet users, and we expected a substantial number of serious respondents.

RESULTS

Analysis of the demographic information revealed a highly experienced group (age=40.2; experience=17.8 yrs.). Company revenues of the respondents varied widely, with 24% of subjects indicating revenues of less than $50 million, and 23% of subjects over $601 million.

The majority of subjects, 71.8% reported using their computer at a private desk or cubicle, while 19.3% reported usage in a public location or with access to the computer shared with several employees. Others reported usage in a location highly visible to others (4.9%) such as in a large office with several desks, while 3.9% reported usage at home. Computers were connected via a high speed Local Area Network in most of the cases (87.7 %). Respondents used their computers, on average, 27.7 hours per week, with Internet usage being less than one hour.

When respondents were asked to estimate the amount of such abuse directly as one number, the average was 4.8 hours per week (see Table 2). In contrast, when asked to estimate each of several categories of Internet abuse, the total is instead 5.8 hours per week. To provide a more complete picture, we chose to also ask how much non-Internet personal time was spent at work, and the estimate was 4.6 hours. Although it is possible that this is also an underestimate, the total amount of personal time reported is 10.4 hours per week, or fully 25% of all work time.

Abuse Activity	Mean hours Per Week (standard deviation)
Non-Internet personal matters at work *	4.6 (7.5)
Estimated Total Internet Abuse	4.8 (5.8)
Summed Total of all categories below	5.8 (6.6)
- Personal communications (including e-mail)	1.7 (2.3)
- Shopping	.3 (.8)
- Selling	.1 (.6)
- Finance and Investing	.3 (.8)
- News	1.0 (1.8)
- Travel	.1 (.4)
- Adult	.1 (.6)
- Nonessential computer maintenance	.5 (.8)
- Hobbies	1.1 (2.3)
- Entertainment	.6 (1.5)
- Self-Education	.01 (.3)

**Examples given to subjects included personal phone calls, chatting about personal matters, napping, and playing golf.*

Table 2: Self-Reports of Internet (and other) Abuse

Reliability Analysis

To measure the independent variables, scales were constructed by averaging the scores on individual items. The scale with lowest reliability, privacy, only reached an alpha score of .65 and was dropped form further analysis. Future research should examine the privacy issue in more detail, and perhaps develop a more reliable instrument.

All other multiple-item scales displayed adequate levels of reliability immediately or reached it after dropping an item (if the initial reliability was below .8).

Multicollinearity Analysis

All independent variables were intercorrelated in preparation for our regression analysis. There was surprisingly little evidence of multicollinearity, as all correlations were well below the generally-accepted .6 level. The two highest correlations were -.546 between supervisor and policy, and .544 between peer culture and supervisor culture. Only two fell between .4 and .5, and all others were well below .4. Even more assuring were the VIF scores in the regression analysis described below; the highest VIF score was 1.8, well below the common threshold of 10 for regular regression and 2.5 for weaker models such as logistic regression.

Regression Results

Regression was used to determine which factors are the most important. Out of the fifteen items shown earlier (with Workplace Privacy excluded), eight were significant in the regression equation as follows. Table 3 presents the regression results.

Item	Standardized Beta	Significance
Peer culture (low=restrictive)	.178	.003
Addiction	.154	.003
Computer Experience	-.211	.000
Supervisor culture (low=restrictive)	.152	.013
Self-justification (low=surplus)	-.212	.000
Job Satisfaction	-.191	.002
Gender (0=F; 1=M)	.126	.017
Revenue	-.107	.038

Table 3. Results of Regression on Total Amount of Internet Abuse – Variables That Entered

The regression equation's adjusted R^2 is .192; nearly 20% of the variance in Internet abuse can be explained by the set of antecedents. Variables that did not enter the equation are playfulness and engagement in the attitudes group, age and Internet experience in the demographic group, and the entire perceived behavioral control group, including abuse policy, productivity measurement, and work monitoring.

There were no surprises with the directionality of any of the coefficients with the exception of self-justification. It was originally expected that employees would behave according to the "ledger" described by Lim (2002). In retrospect, it is possible that employees in our sample with more feelings of surplus have higher levels of freedom and privileges, and are more able to get their own way. Our speculation requires additional research before this explanation can be taken seriously.

Items without previously-held expectations include gender, revenue, and computer experience, which show explanatory power in the model. Males, computer novices, and employees in small firms are more likely to abuse the Internet than females, more experienced employees, and those in large firms.

Although it is difficult (and perhaps dangerous) to speculate why males commit more Internet abuse than females, the other two demographic factors invite speculation. Employees with less computer experience might be undergoing a temporary "infatuation" while those with more experience have already gotten satiated in the past. In smaller firms, there might be less formality and wider latitude of behavior. It is possible that there are fewer peers or supervisors present, and therefore no consistent source of subjective norms.

Finally, it was puzzling to see the failure of all of the items of TPB Perceived Behavioral Control to provide any significant explanation of the extent to which respondents reported Internet abuse. Regarding policy, it is possible that policies not only lack legal grounding (Siau et al., 2002), they also lack behavioral grounding. The study by Lim et al. (2002) indicates that only 60% of employees accept usage policies.

The failure of productivity measurement and work monitoring to provide explanatory power are more difficult to explain. Both variables were normal and exhibited a wide range, so a lopsided distribution or a restricted range cannot account for the failure. Perhaps the items need to be adjusted in further studies. Table 4 summarizes the results.

CONCLUSION

This study attempted to examine the extent to which employees engage in Internet abuse, and whether any of fifteen antecedents show a significant relationship with the amount of Internet abuse. Data were collected from a sample of 571 Usenet users, using an online survey.

An aggregated total of 5.8 hours of self-reported Internet abuse was reported plus a 4.6 hour estimate of non-Internet personal time at work. The total is 10.4 hours per week on personal tasks, or 25% of a 40-hour work week.

Examining the antecedents in a regression analysis revealed that most of the antecedents in two of the three Theory of Planned Behavior (TPB) categories (Attitudes and Subjective Norms), were significant, and none of the antecedents in the third TPB category (Perceived Behavioral Control) showed significance.

Attitudes	
H1a: Job Satisfaction limits abuse	Supported
H1b: Playfulness promotes abuse	Not supported
H1c: Engagement promotes abuse	Not supported
H1d: Internet Addiction promotes abuse	Supported
H1e: Self-Justification – inequity promotes abuse	Not supported (Reverse)
Subjective Norms	
H2a: Supportive Peer Culture promotes abuse	Supported
H2b: Supportive Supervisor Culture promotes abuse	Supported
Perceived Behavioral Control	
H3a: Abuse Policy limits abuse	Not supported
H3b: Workplace Privacy promotes abuse	Not supported
H3c: Productivity Measurement limits abuse	Not supported
H3d: Work Monitoring limits abuse	Not supported
Demographic Variables	
H4a: Gender	Entered the model
H4b: Age	Did not enter
H4c: Experience with Computers	Entered the model
H4d: Experience with the Internet	Did not enter
H4e: Size of the Firm	Entered the model

Table 4 – Results of Hypothesis Testing

The issue of Internet abuse in the workplace is only beginning to be investigated. By examining the factors that lead to increased abuse, researchers and managers might better understand the phenomenon and its antecedents. It appears that policies and other restrictive practices fail to restrict Internet abuse, and perhaps it would be more effective to try and foster a culture that does not support the practice. Better understanding of the Internet abuse phenomenon that this and future studies will help provide, might eventually help the workplace become a more productive and creative work environment.

REFERENCES

1. Ajzen, I. "Nature and Operation of Attitudes," *Annual Review of Psychology*, 52, 2001. 27–58
2. Colby, C. and Parasuraman, A. (2002) "The 2002 National Technology Readiness Survey," University of Maryland Center for E-Service, http://old-www.rhsmith.umd.edu/ntrs2002/
3. Computer Security Institute (2001). Issues and Trends: 2001 CSI/FBI Computer Abuse and Security Survey, CSI, San Fransisco, CA, March.
4. Hipple, S. and Kosanovich, K. "Computer and Internet Use at Work in 2001," *Monthly Labor Review*, February 2003 (also available at http://www.bls.gov/opub/mlr/2003/02/art2full.pdf)
5. Jones, G. R. (1984). "Task visibility, free riding, and shirking: Explaining the effect of structure and technology on employee behavior," *Academy of Management Review,* Vol. 9, No. 4: 684695.
6. Lee, J and Lee, Y. (2002). "A Holistic Model of Computer Abuse Within Organizations," *Information Management & Network Security*, 10(2), pp. 57-63.
7. Lee, Y; Lee, Z; and Kim, Y. (2002) "Factors Affecting Internet Abuse in the Workplace" Academy of Management Meeting, Denver.
8. Lim, V. (2002) "The IT Way of Loafing on the Job: Cyberloafing, Neutralizing, and Organizational Justice," *Journal of Organizational Behavior*, 23, 675-694.
9. Lim, V. and Teo, T. (2002). "How Do I Loaf Here? Let Me Count The Ways," *Communications of the ACM*, January, 45(1), pp. 66-70.
10. McCollum, T. (1998). "Preventing a productivity drain" Nation's Business.
11. Milgram, S. "Some Conditions of Obedience and Disobedience to Authority," *Human Relations*, V. 18, 1965, pp. 57-76.
12. Siau, K., Nah, F.F., and Teng, L. (2002). "Acceptable Internet Use Policy," *Communications of the ACM*, 45(1), pp. 75-79.
13. Spector, P.E. (1997), *Job Satisfaction: Application, Assessment, Causes, and Consequences*, Thousand Oaks, CA: Sage
14. Stanton, J. (2002). "Company Profile of the Frequent Internet User," *Communications of the ACM*, 45(1), pp. 55-59.
15. Vogt, P. (1997). "Just one more site…," *Credit Union Management*, Vol. 20, No. 5: 24 25.
16. Webster, Cynthia. (1991), "Influences upon Consumer Expectations of Services," *Journal of Services Marketing*, V.5, N. 1, Winter, pp. 5-17.
17. Webster, J. and Ho, H. (1997). "Audience engagement in multimedia presentations," *The DATA BASE for Advances in Information Systems,* Vol. 28, No. 2: 63-77.
18. Webster, J. and Martocchio, J. J. (1992). "Microcomputer playfulness: Development of a measure with workplace implications," *MIS Quarterly*, Vol. 16: 201-226.
19. Williams, K., Harkins, S. and Latane, B. (1981). "Identifiability as a deterrent to social loafing: Two cheering experiments," *Journal of Personality and Social Psychology,* Vol. 40: 303 311.

HCI Research Transfer to Practice: Better Together

Mary Czerwinski
Microsoft Research
marycz@microsoft.com

Izak Benbasat
University of British Columbia
izak.benbasat@sauder.ubc.ca

Julie Ratner
Infospace Mobile
julie.ratner@infospace.com

Radhika Santhanam
University of Kentucky
santhan@uky.edu

Peter Todd
University of Virginia
ptodd@virginia.edu

ABSTRACT

Currently, HCI researchers and HCI practitioners work in relatively separate spheres of influence. Practitioners often question the value of academic HCI research and desire more practical directions. HCI researchers often wonder if their research findings are communicated via the optimal channels for influencing practitioners' process and direction, or whether their results generalize to the real workaday world of HCI. This panel attempts to outline what practitioners need from their academic partners, and how they think these needs can be addressed by academic research. Academics on the panel will state what they see as interesting future research challenges, and whether or how they think they can address the practitioner community's interests. The practitioners on the panel will then state their opinions about the opportunities for technology transfer from academia to practice.

Keywords

Academic HCI research, HCI practitioners, technology transfer, usability, generalizability, work practice

INTRODUCTION

For some time now the practitioner community has taken issue with the research activities of their academic counterparts. Likewise, the academic community in HCI has often felt a level of frustration associated with knowing whether or not their research findings generalize to the real world, are timely enough, or "cutting edge" enough, to benefit the practitioner community. Additionally, transferring research technology or techniques to the practitioner can be problematic, since the two communities might not attend the same conferences or read the same materials. In fact, the practitioner community may not have the time to attend conferences or publish how they perform their daily tasks, further exacerbating the flow of knowledge between the two areas of discipline.

What follows are a series of position statements from the panelists. We will have three academic panelists and two industrial participants. Of our industrial participants, one is an active user researcher in the midst of real product design, while the other is an industrial researcher who works closely with product teams and has worked in product development in the past. Our format will be as follows:

The practitioners will outline what they feel their community would like to see more of or need from their academic counterparts. The academic researchers will describe what they think is of interest to academic researchers in HCI in the short and long run. Practitioners will then respond to this research agenda in terms of whether or not they think these topics are of interest to practice, whether or not they are too esoteric, timely, and whether or not the practitioner community could provide resources to help with the technology transfer. After 45 minutes or so of debate on these topics, we will open the panel to audience participation.

MARY CZERWINSKI

I have worked as a usability engineer on product teams, as an adjunct professor at universities and as an industrial research scientist doing applied research. Because of these experiences, I have come to have a keen understanding of how difficult it is to transfer HCI research knowledge, especially from within an academic setting, to the practitioner community. As academics, we tend to think very long-term, and often more theoretically and systematically, than our practitioner counterparts have the luxury to afford. As practitioners, we tend to need to have a myriad of tools and techniques at our fingertips, ready for application quickly as our product development cycle dictates. Practitioners rarely have the time to perform research necessary for refining or iterating on a problem or an aspect of their craft, much less publish methods or techniques that they have developed to solve a practical problem. In my opinion, academic HCI researchers need to partner with their practitioner counterparts. The academic researcher needs access to real user scenarios, and real data or artifacts, and real design challenges for their research to have the proper scope of influence. By product teams partnering with academics or their students, everyone wins. The

product team benefits from the perspective and technology that the researcher can bring to bear, in addition to the benefit of theoretical or systematic research findings. They also get access to technological resources that they might not have had otherwise. The academic quickly comes to realize which parts of their research program do or don't make sense in a real world context, given real world time pressures and resources.

IZAK BENBASAT

Academics and practitioners are sometimes characterized as the two solitudes though neither party desires such a state of affairs. As my fellow panelists describe in more detail below there is a strong and genuine desire to cooperate, but very often the realities, rewards structures, and constraints imposed by our separate environments make it difficult to do so. In the information systems academic literature there has been interest in recent years in exploring the means for cooperation and knowledge transfer, including a commentary that I wrote with Bob Zmud (Benbasat & Zmud, 1999). However, the papers written have put forth the views of academics only. This panel will give academics an opportunity to hear the opinions of the practitioners, and allow us to adjust our thinking and tactics to fit their needs and constraints, and will do the same for practitioners. It is my view that though cooperation is a desired goal the means to achieve it is not easy. Hence, I hope that we will be able to come up with a few but concrete means of achieving cooperation, and measure our success in doing so in follow-up panel discussions that will take place at ICIS in future years.

JULIE RATNER

After working in industry for nearly a decade, I perceive my years doing academic research and longitudinal government research through a different lens with a more strategic business perspective. Today, I interpret academic research results I read with keen interest and notice that I usually yearn for timely reporting and a focus on practical details and less theory.

Since I work with engineers and designers on wireless applications with 1-10 week definition to delivery timeframe; the key to successful collaboration between InfoSpace Mobile and academia is efficacy and flexibility. To use a common metaphor, HCI results are relevant to my product teams "when the rubber hits the road;" when they impact the bottom line, *before* applications launch. Success is measured by initial user experience; if a mobile user's *first* experience is intuitive, user adoption of wireless applications is likely to increase.

I have had a few successful collaborative research projects with academia since I have worked in the software industry. When I worked at on the east coast, we sponsored semester long research projects each year. The reason they were successful is that the graduate and undergraduate students stayed in budget, delivered what they promised, and listened to and answered the product teams' questions about users. The value of working with these students was multi-faceted; we were able to delegate 6-month field studies that the company did not have the resources to conduct and we frequently hired exemplary students as interns once the semester ended. The students in turn gained practical experience about the value of research in industry and a few even received offers of employment with the company.

Depending on how collaborative projects are structured in the academic settings, my reaction is initially mixed, not because I don't value and appreciate academic research (which I do) and not because I don't see the benefit of partnering with academia (which I totally support), but because I know for a fact that our time-frames are out of sync. In one academic semester for example, our business goals typically shift repeatedly and oftentimes the HCI research that would have been priceless in January is not relevant by May.

RADHIKA SANTHANAM

Though I have worked in the industry, it was not related to HCI work and I consider myself to be primarily an academic researcher. Therefore, my views may seem a little radical to the practioner panelists, and I do welcome them to convince me otherwise. While I think it is important for academic research to be relevant to practice, I also feel that we will and must continue to have a certain areas of research space that is distinctly different, and which will seem somewhat irrelevant to the other group. In fact, I feel that if we did similar kinds of research and chased the same specific problems, we will not have much to offer to the other group. I clarify this premise up-front so that we can better discuss what knowledge we academics need to transfer to practice and vice-versa. It will also help us identify those intersecting areas of interest.

First, the research goals of our two communities are fundamentally different. Our goal as academic researchers in HCI is to understand underlying, and (hopefully) enduring, principles of human behavior that come into play when interacting with computers. We focus on building a cumulative body of knowledge. As I see it, HCI practioner researchers are also interested in understanding these underlying principles, but want it in a form that they can readily apply to system design and product development. They usually do not have the time to investigate and develop underlying principles. Therefore, one way I see for academics to communicate this knowledge is to get together every couple of years in a workshop with the sole goal of "Knowledge Transfer between Academics and Practioners". In such a workshop, an expert in specific areas of HCI, (e.g. on the topic of visualization, decision making or training) will present all the key findings/ideas that have been generated in academic research in the last few years. The expert will also indicate how these ideas could be applied to

practice. This will enable the transfer of knowledge discovered in academia by eschewing the theory, the complicated statistics, the obfuscating language and all other things that practioners do not want to sift through. Practioners could ask for clarifications, quiz and maybe even disparage these findings! But this interactive process will provide good feedback to academic researchers on what aspects of research was useful to practice, and practioners will hopefully obtain nuggets of knowledge in a speedier fashion.

As an academic researcher, what I want to know from practioners is about repetitive problems that they have faced, and about issues they have not been able to solve. I emphasize repetitive because a user problem in one specific application is something that academic researchers should not worry about. Using the wireless applications as an example, if there are persistent problems about displays, or issues relating to user learning/adoption that are perplexing, then it should be brought to our attention. We as academics can search for some underlying issues that could perhaps explain these persistent problems. Once again, I think the workshop setting is an avenue where this can occur. I think an important way by which practioners can help transfer and also help develop knowledge that is useful to practice, is to share data on these problems.

While I like the idea of collaborative projects, I think knowledge transfer has to occur at a higher level of abstraction than single projects. These projects do have value to the extent that each group can get to know members of the other group and understand their perspectives. But I am afraid that too many such projects will lead to a situation where academic researchers are also huffing and puffing about product development cycles and delivery schedules. Furthermore, if we academics also start to focus on immediate problems and specific products, I fear that in the long run, we will become even more irrelevant to practice.

PETER TODD

Should we build bridges between academic research and the practice of HCI? Most of us would agree this is a laudable goal. A motherhood and apple pie agenda. But as Professor Santhanam notes above, such a goal may have unintended consequences. As academics are neither trained nor motivated to examine issues in the short term or to provide rapid results. As a consequence by following the needs of practice we risk making academic research, which is narrow, focused, long term and cumulative in nature less relevant as we try to meet the needs of practice, to provide rapid results to immediate issues with bottom-line impact. And to do it with fewer resources, with less sense of the market and ultimately less well than do our colleagues addressing the same issues in practice.

Those absorbed in practical issues of systems design and implementation are likewise not well-attuned or motivated to the possibility of taking our narrow theoretical notions and applying them to their practical efforts. In this context the chasm between our research abstractions and the immediate needs of practice appear to be nearly insurmountable. What then can we do?

My colleague Izak Benbasat suggests the way, we academics, can get practical. Not practical in our substance, but practical in our approach. We need to look for the few things we can practically do that will help to build bridges. Our colleagues in practice can also become more open to the importance ideas that evolve over the long term. In addition we can all be a little more patient.

Lets start with patience. Recently I was preparing a graduate class on decision-making in our executive-format Master of IT Management Program. As is often the case for these classes I turn to sources such as the Harvard Business Review to find coverage that will be accessible and acceptable to them. In this particular instance one of the articles I chose was:

Delusions of Success: How Optimism Undermines Executives' Decisions by Dan Lovallo and Daniel Kahneman (HBR July 2003). Kahneman, of course, was the recipient of the Nobel Prize in economics in in 2002 for his landmark work with the late Amos Tversky (who also received the award). Their initial work dates to the early 1970's and formed the basis for the HBR article. Thirty years from theory to practice. Lets learn to be patient.

While we are waiting there are few other things we can think about.

- **We should learn to talk**. I have often found it is possible to have interesting and productive conversations with practitioners about theory and research results.
- **We should learn to listen.** Practice is a great source of interesting questions. Often not the question that is being posed but higher level questions that really are enduring. The issues and questions do not change as quickly as they are made out to.
- **We should learn to cooperate.** Unlike almost any other area of research the HCI field has a remarkable opportunity to collaborate to collect information that can lead to important theoretical insights and inform practice.

Our panel discussion should provide us with an important opportunity to examine these and other issues. One thing we can be sure of we will all be optimistic about the possibilities, pessimistic about the ability to act on those opportunities and impatient for results.

REFERENCES

Benbasat, I. And R. W. Zmud, "Empirical Research in Information Systems: The Practice of Relevance", *Management Information Systems Quarterly*, March 1999, pp. 3-16.

A Model Made of Paper: Clinicians Navigate the Electronic Health Record

Catherine Arnott Smith
Syracuse University
casmit07@syr.edu

ABSTRACT

The electronic health record (EHR) is actually an aggregation of individual clinical documents. Medical records document not only the knowledge domains of clinical practice, but the work processes and practices that support these domains. Human-computer interaction is an important factor in EHR system success: researchers have argued that clinician readers consciously perceive the context of production, and integrate an understanding of the producer into their understanding of the data. In support, this paper reports findings of an information retrieval study using a simulated EHR containing deidentified clinical documents. Physician subjects verbally demonstrated use of a mental model of the paper medical record during their navigation of the system. Clinicians may actively apply a mental representation of their domain of practice—and actively refer to this paper-based knowledge base—when they access medical data. An understanding of the mental models that clinicians use would greatly inform our understanding of EHR systems.

Keywords

Clinical information systems; medical records; passage retrieval; mental models

INTRODUCTION

The electronic health record, or EHR, has constituted one of the great unattained goals of medical informatics since the early 1970s: the 'quest for the Holy Grail' (Gregory, Mattison, and Linde, 1995 p. 59; Nygren and Henriksson, 1992). Facilitating access to the medical record has historically been viewed as such a resource-consuming task that electronic medical record construction was the driving force behind early hospital information systems (Collen, 1987). And as early as 1975, researchers in medical informatics argued that "the psychological characteristics of the user should be taken into account in the design and implementation of medical information systems" (Herbst, 1989, p. 389). To understand the problem of clinical information retrieval in the context of an EHR, it is necessary to understand not only its users, but the nature of the data that comprise it.

The medical record is a feature of patient care as much as 3000 years old (Spiegel and Springer, 2001). Frisse (1992) has identified four principal ancestors of the modern medical record: (1) the case record collection of the 19th century, resembling "diaries or research notebooks"; (2) the bedside chart containing individual patients' vital signs and observations; (3) the physician order, used for workplace communication; and (4) the financial ledger, or record of physician charges and transactions. Today, still patient-centric, still resembling its ancestors, the typical record is still kept on paper. EHRs have penetrated only 5-10% of the U.S. market (Carpenter, 2002). Small wonder that one author has asked: "Is a user-friendly, secure and interactive electronic medical record a figment of the collective imaginations of overzealous techies?" (Thompson, 1996, p. 29). What has prevented attainment of this particular future? Morrissey (2001) blames the "best of breed" mentality prevalent in healthcare IT: "Healthcare applications were selected to satisfy a particular department rather than their ability to share and consolidate information with other applications and the healthcare system as a whole". Thus, the integration of data from applications that were best for different things has only reinforced and perpetuated a pre-existing lack of communication and disdain for standards.

The most fundamental function of the medical record is that throughout its development, in whatever medium, it always documents not only the knowledge domains of clinical practice, but the work processes and practices that support and maintain the operation of these domains. Sociologist Marc Berg (1996) wrote that the record "is part and parcel of the production of hierarchical relations, of the shaping of the doctor-patient encounter, of the processes that constitute the socialization of interns, and so forth" (p. 501). Rees pointed out that "the very possibility of understanding the record's entries is based on a shared, practical, and entitled understanding of common tasks, experiences, and expectations" (Atkinson and Heath, 1981, pp. 200-201). Whether digital or paper, the medical record encodes work processes and subprocesses.

The traditionally oriented medical record is organized around sources of medical data, such as patient encounters with the physician. For example, in the MARS (Medical Archival and Retrieval) System in use at the University of Pittsburgh Medical Center, the

organizational scheme is traditional: unstructured free-text narratives are classed by one of 19 clinical report types, with the text providing a further account according to that central event. For example: a "Radiology Report" breaks down into components describing the procedure performed, clinical data associated with/generated by the procedure, and the final conclusions of the radiologist regarding the data. The EHR as "medical record" is actually an aggregation of individual clinical documents like this.

An EHR system thus needs to be understood as a document base rather than a database, "based primarily upon a store or collection of documents, rather than a store or collection of structured data" (Chen and Dhar, 1991, p. 406). Typical clinical tasks performed using EHR systems include the following (Laerum et al., 2001): Reviewing the patient's problems; seeking specific information from patient records; Following results over time; obtaining new results; and reviewing cohort data. Clinical information retrieval occurs in situations in which every second, clinically, counts, which makes human-computer interaction an important factor in EHR system success. Retrieval in this domain simply "can't be more time-consuming than reading from a conventional paper record" (Nygren and Henriksson, 1992, p. 1) or clinicians have no incentive to use, let alone rely upon, the system. An understanding of the mental models that clinicians use when they access information retrieval systems in medicine would greatly inform our understanding of the systems.

MENTAL MODELS

I use here Borgman's definition (1999): "a cognitive mechanism for representing and making inferences about a system or a problem which the user builds as he or she interacts with and learns about the system" (p. 436). Donald Norman stated early (1983) that a person's mental model "reflected his or her beliefs about the physical system, acquired either through observation, instruction, or inference" and furthermore that an individual's "beliefs about a system lead to expectations of the system's capabilities" (Norman, 1983). Kieras and Bovair (1984) early on found that imparting device model information to the user had strong effects on that user's ability to use the system; Fein, Olson and Olson (1993) followed on these findings to investigate a continuum of mental models at work in the use of a control panel. To support the model-understanding connection, some have blamed the "inadequacy" of mental models for users' inability to cope with system failures (Cardinale, 1991) and subjects' mental models have been found to interfere with learning when mapping from "print" (typewriter mental model) to "digital" (computer mental model) (Borgman, 1999, citing Douglas, 1983).

According to Borgman (1999), the bulk of research in the area of mental models and computers has related to text editors and calculators; see, for example, Halasz and Moran (1983). The body of IR research that has considered mental models has done so primarily in the service of training. Dimitroff (1992) focused on the relationship between mental models and users' outcomes in searching a bibliographic retrieval system and found that subjects whose models were most "complete" found significantly more items when searching. Borgman (1999), like Dimitroff, investigated the contribution of users' mental models to success in task performance using an IR system. She found that even subjects who were not trained in the use of mental models were able to develop such models without assistance, echoing the conclusions of Fein et al..

PASSAGE RETRIEVAL

The research reported in this paper was based in the theory of passage retrieval, a subset of corpus-based information retrieval: "the task of identifying and extracting fragments from large, or short but heterogeneous, full text documents" (Melucci, 1998). Passage retrieval is thought to enable more precise retrieval because it concentrates the reader's attention on those parts of the text that have a "high density" of relevant information, thus providing an "intuitive overview" of the knowledge base (Salton and Allan, 1993).

Can this be ascribed to a mental model? Eveland and Dunwoody (2000) proposed that hypermedia learners have their own model-building facilitated by the visual and ontological scaffolding provided by hypertext links. When the learner uses the scaffolding, "The structured representation acts as an intensional definition, in the particular vision of a world embedded in a structure." (Rossi Mori, Galeazzi, Consorti, and Bidgood, 1997).

The clinical documents that make up the EHR can be considered to be composed of passages, since they contain units of textual discourse such as sentences, paragraphs, and sections. Clinical documents are typically extremely short, and have unique and nonredundant text. However, clinical document section headings are more like database fields than the content summaries seen in passage retrieval research (for example, Hearst and Plaunt, 1993). These "section headings", "labels", or "segment labels", as they are variously called in the literature, serve as the means by which readers navigate the documents. Nygren, Johnson, and Henriksson (1992) identified three reading techniques of medical records: first, skipping over irrelevant sections; second, skimming sections identified as possibly relevant; and third, reading needed information carefully. Labels thus signal content to the reader, both denoting the structure and defining the domain of knowledge.

This paper reports findings of a clinical information retrieval study using a simulated EHR system with a document base of deidentified but authentic clinical documents. The purpose of the study was to assess the contribution of XML markup to improved retrieval of

clinical documents. Subjects were assigned information retrieval tasks to which the document base provided the gold standard answer. The control group searched a flat file of ASCII full-text clinical documents; the treatment group could pose field-based queries enabled by XML markup of passages denoted by section headings. A side effect noted during the experiment was the demonstration by physician subjects that they incorporated a mental model - a model of the paper document base - into their own navigation of the simulated system.

METHODS

This experiment was conducted during April and May, 2002. One thousand clinical documents from the MARS system in place at the University of Pittsburgh Medical Center (UPMC) were randomly selected and automatically deidentified; that is, all individual identifying information was removed and replaced with pseudonymizing text. These 1000 documents were evenly distributed among the 8 most frequently occurring types found in a pilot study: radiology reports; progress notes; physician letters; operating room notes; history and physical notes; surgical pathology reports; discharge summaries; and emergency room visits. Subjects were 10 physicians (9 M, 1 F, ages 28-45) drawn from a convenience sample, experientially varying from a medical school graduate to attending faculty members. Results are also reported here from a pilot study involving 5 additional male physicians meeting the same criteria.

The simulated EHR was built with an open source XML database called Xindice. The simulation offered the same search capabilities as did MARS, but used a purposely simple Web-based browser interface. Subjects could use Boolean operators and partial string matching to search full-text XML documents, return a list of results, and display full-text documents for further browsing. A session log running behind the scenes captured user activity. A written log of subject comments was kept by the investigator during each search session. Subjects are referred to by number, for example, Subject Five = S5; Pilot Study Subject Six=PS6.

THE SEARCH EXPERIENCE

Comments about the Experiment

Several subjects were careful to make a distinction between the artificiality of their environment and the clinical setting the environment attempted to replicate. S5 used the phrase "in real life" while S6 told me what he would do if it wasn't an experiment:

[There's a] difference between the two searches—why? Have to figure out. In real life would go on and compare. (S5).

I would look at 49 – they have to be in here! (S6).

And S1 noted:

In real life, I probably wouldn't be going through all these! (S1)

However, one subject did twice comment that the experiment was realistic:

That's exactly how you would look for this sort of thing. (S8). … They were Discharge Summaries and Consult Notes—just what I would have wanted! (S8).

About the Search Process

Comments spoke to the role of the searcher and to the searcher's understanding of his or her role in interpreting the document. As S8 remarked rhetorically:

Am I a clinician or a research assistant?

And PS S3 achieved the same effect by stating firmly:

I'm not a neurologist!

Most subjects indicated by their comments that they knew likely locations for information:

If I could search 'Past Medical History', something would be structured … (S6)

It's probably going to be under 'Social History'! (S8)

'Adenocarcinoma' would have to be in the 'History'. (S4)

In my mind there's an idea that this would go with 'Procedure' or 'Techniques'. I was happy to click on 'Procedure' and try to search! (S5)

I expected it to be in 'Hospital Course' because it's medicolegal. Anything that happens to the patient while they're in the hospital is going to be in there! (S6)

I read 'Techniques' for 'premedication' – that's where I would expect to see it. (S5).

Conversely, some information was *not* expected to be in particular locations; subjects were capable of expressing surprise:

'Physical Exam' – shouldn't be there, but we don't know! (S9)

The only other place to look would be the 'History' part, but there are only a couple of those (S3)

'Substance abuse' was not in 'Social History'—that's where it should be, but it's not! (S6)

'Pleural effusion' should be in the 'Reason' part, not 'Hospital Course!' (S5)

Strange to find it there! I thought it would be under 'Cytogenetics', but it's molecular genetics. (S5).

["Why did you go straight to the 'Description of Operation' field?"] Because that's where they'd deal with 'resection'. It wasn't where I expected it to be! (S10)

About the Documents

One common subject response was to verbally place themselves in the role of the document creator, sometimes recreating what they considered to be the thought processes of the document creator. S8 explained that the length of a clinical document related directly to medical billing, since "extensive documentation of history" (i.e., "No history of diabetes") enabled "upcoding" for a larger bill matching the longer dictation; or "some people just keep on going." Similarly:

Some people say 'Past Medical History', some people say 'Past Surgical History'. (S10).

But some subjects made guesses as to what kind of clinician dictated the note, apparently as part of their sense-making about the content:

I think a medical student wrote this! (PS S3)

A dermatologist probably didn't write this. They would have more accurate terminology! … An internist wrote this. Would have been easier if a dermatologist wrote it! (S4).

This is just my observation: Consultants say 'Impression' and primary care [physicians] say 'Plan!' (S8)

Two subjects verbally assumed the creator's role:

I'll just search for 'colon' because I'm not mentioning a normal guy's colon! (S8).

I just thought that if I were dictating a note…where would I put that information? (S7)

In two cases, the subjects attempted to second-guess the document author's diagnosis as part of their search strategy:

If I could remember the things that cause phlebitis, I could look those up! (S1)

Even though it says 'rectal', I consider this colon cancer! Rectal cancer is similar to colon cancer. (S12)

And two subjects verbally corrected the absent document creator:

I have used 'liver mass' in this situation – [the author of the document] should have used other words. (S6)

The word wouldn't be 'ileoscopy.' It would be 'ileostomy.' … The syntax is weird in the diagnosis. (S7).

This expressed itself in one case in a dialogue with the phantom author:

How dare you say 'degenerative joint disease' instead of 'arthritis?' Don't you know I'm going to be searching for this, you jerk? (S8)

DISCUSSION

Context is vitally important to communication of medical data. Whether an EHR is document-centric or data-centric, when it reflects clinical work processes, it becomes a working model of the clinical knowledge domain. As a result, Panko et al. (1999) have noted a drawback in constructing such systems via relational databases: "[T]he loss of both context and integrity when such elements are extracted and isolated from the original report" (p. 5). This context is "a knowledge base… composed of expert knowledge about the domain (medical application) and knowledge about documentation in the domain" (Poullet, Pinon, and Calabretto (1997), p. 120; italics mine).

Once the EHR is fully understood as a knowledge base, requiring context for accurate interpretation, the question then for systems developers becomes how best to represent that knowledge. Clinicians may actively apply their own representation of their domain of practice—actively refer to this paper-based knowledge base-- when they access medical data: "The data are transferred embedded in the significance-functions contributed by the conceptual frameworks of the relevant parties" (Kluge, 1996, p. 88). This framework is a filter that clinicians use to process the data they read. "In and of themselves, these data are not related", but the connections between the data points are the "information-space for the set of data" (p. 90). Or, as Essin puts it more succinctly, in his own Information Model: "Facts originate in events. Facts require context to be informative." (quoted in Royal College of General Practitioners, 1999).

Clinical information and work context are intimate and inseparable, and "The further information has to be able to circulate, the more work is required to disentangle the information from the context of its production" (Berg and Goorman, 1999, p. 52). Human readers of medical information interpret and reinterpret to assess the information "in the light of who generated it" (Berg and Goorman, 1999, p. 55), whether that generator be a human being or a machine. Readers consciously perceive the context of production, and integrate an understanding of the producer into their understanding of the data. I hope in future research to further explore the relationship between clinician readers and the mental models that document their clinical worlds.

REFERENCES

1. Atkinson, P., and Heath, C. (Eds.) (1981) *Medical work: Realities and routines*. Farnborough, England: Gower.
2. Berg, M. (1996) Practices of reading and writing: The constitutive role of the patient record in medical work. *Sociology of Health and Illness*, 18(4), 499-524.

3. Berg, M., and Bowker, G. (1997) The multiple bodies of the medical record: Toward a sociology of an artifact. *Sociological Quarterly, 38*(3), 513-537.
4. Berg, M., and Goorman, E. (1999) The contextual nature of medical information. *International Journal of Medical Informatics*, 56, 51-60.
5. Borgman, C.L. (1999) The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Human-Computer Studies,* 51, 435-452.
6. Callan, J. P. (1994) Passage-level evidence in document retrieval. *Proceedings of ACM SIGIR Conference* (pp. 302-310). New York: ACM.
7. Cardinale, L.A. (1991) Conceptual models and user interaction with computers. *Computers in Human Behavior,* 7, 163-169.
8. Carpenter, D. (2002) The paperless chase. *Hospitals and Health Networks, 76*(1), 47.
9. Chen, H.C., and Dhar, V. (1991) Cognitive process as a basis for intelligent retrieval-systems design. *Information Processing and Management*, 27, 405-432.
10. Collen, M. F. (1987) Health care information systems: A personal historic review. *Proceedings of ACM Conference on History of Medical Informatics* (pp. 123-136). New York: ACM.
11. Dimitroff, A. (1992) Mental models theory and search outcome in a bibliographic retrieval system. *Library and Information Science Research,* 14, 141-156.
12. Eveland, W.P., Jr., and Dunwoody, S. (2000) Examining information processing on the World Wide Web using thinkaloud protocols. *Media Psychology*, 2, 219-244.
13. Fein, R.M., Olson, G.M., and Olson, J.S. (1993) A mental model can help with learning to operate a complex device. *Proceedings of the Conference on Human Factors in Computing Systems [Interact '93]:* 157-158. New York: ACM.
14. Frisse, M. (1992) The health of the computer-based patient record. *Academic Medicine, 67*(7), 441-443.
15. Gregory, J., Mattison, J. E., and Linde, C. (1995) Naming notes: Transitions from free text to structured entry. *Methods of Information in Medicine*, *34*, 57-67.
16. Halasz, F.G. and Moran, T.P. (1983) Mental models and problem solving in using a calculator. *CHI Proceedings*, 212-216.
17. Hearst, M. A., and Plaunt, C. (1993) Subtopic structuring for full-length document access. In *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 59-68). New York: ACM.
18. Herbst, A.J. (1989) The evaluation of user interface design in medical software. *Medinfo,* 368-371.
19. Kieras, D.E., and Bovair, S. (1984) The role of a mental model in learning to operate a device. *Cognitive Science*, 8, 255-273.
20. Kluge, E.-H. W. (1996) The medical record: Narration and story as a path through patient data. *Methods of Information in Medicine, 35*, 88-92.
21. Laerum, H., Ellingesen, G., and Faxvaag, A. (2001) Doctors' use of electronic medical records systems in hospitals: Cross sectional survey. *BMJ, 323*(7325), 1344-1348.
22. Melucci, M. (1998). Passage retrieval: A probabilistic technique. *Information Processing and & Management,* 34(1), 43-67.
23. Nygren, E., and Henriksson, P. (1992) Reading the medical record: I. Analysis of physicians' ways of reading the medical record. *Computer Methods and Programs in Biomedicine*, 39, 1-12.
24. Panko, W., Silverstein, J., and Lincoln, T. (1999) Technologies for extracting full value from the electronic patient record. *Proceedings of the 32nd Hawaii International Conference on System Sciences.* New York: IEEE.
25. Poullet, L., Pinon, J.-M., and Calabretto, S. (1997) Semantic structuring of documents. *Proceedings of the 3rd Basque International Workshop on Information Technology (BIWIT)* (pp. 118-124). New York: IEEE.
26. Rossi Mori, A., Galeazzi, E., Consorti, F., and Bidgood, W. D. (1997) Conceptual schemata for terminology: A continuum from headings to values in patient records and messages. *Proceedings/AMIA Annual Symposium*, 650-654.
27. Royal College of General Practitioners Health Informatics Task Force. (1999) *Scoping the primary care view of the option for the development of the EPR (ScopeEPR).* London: Author.
28. Salton, G., and Allan, J. (1993) Selective text utilization and text traversal. *Hypertext '93* (pp. 131-144). ACM.
29. Spiegel, A.D, and Springer, C. R. (1997) Babylonian medicine, managed care and Codex Hammurabi, circa 1700 B.C. *Journal of Community Health, 22*(1), 69-89.
30. Thompson, T. R. (1996) Easier said than done. *Hospitals and Health Networks*, 70(13), 29-31.

Effect of Presentation Flaws on Users' Perception of Quality of On-Line Stores' Web Sites: Is it Perception that Really Counts?

Andrea Everard
University of Delaware
everarda@lerner.udel.edu

Dennis F. Galletta
University of Pittsburgh
galletta@katz.pitt.edu

ABSTRACT

Presentation flaws are abundant in web sites, but there has been no study to determine how presentation flaws affect consumers' perceptions of quality of an on-line store, trust in the store, and ultimately the intention to purchase. The theoretical foundation stems from various relevant streams of literature: trust and credibility, impression formation, and impression management. A laboratory experiment examined three main factors, incompleteness, error, and poor style, and used 160 student subjects in a completely balanced, fully factorial design (2x2x2). It was found that error, incompleteness, and poor style affected consumers' perceived quality of the web site. Furthermore, it was found that the relationship between the factors and perceived quality was mediated by the perception of the flaws. The perception of flaws rather than the actual flaws influenced users' perception of quality.

Keywords

Presentation flaws, perception, web site quality, trust, intention to purchase.

INTRODUCTION

Many of the activities performed over the Internet involve financial and confidential transactions; it is of crucial importance that users perceive such systems to be credible. Credibility, often equated with believability, is composed of trustworthiness (perceived accuracy and goodness) and expertise (perceived knowledge, skills, and competence of the developer) (Fogg & Tseng, 1999).

Before divulging personal or confidential information, users need to judge a web site worthy of trust. Hoffman, et al. (1999) suggest that the main reason consumers are resistant to providing personal information and to buying on-line is a fundamental lack of trust (Garbarino & Johnson, 1999; Doney & Cannon, 1997).

PRESENTATION FLAWS

While credibility and trust can be enhanced by users' perceptions of reliable and accurate information being supplied by the computer, flaws in the information provided may serve to destroy that trust. In some extreme cases, flaws could prevent users from using the system in a meaningful manner (Molich & Nielsen, 1990).

For this research, presentation flaws are grouped into three categories: (1) Poor Style, (2) Incompleteness, and (3) Error. Poor style includes graphical and visual elements such as backgrounds that interfere with page text, inconsistent word and line size and spacing, and improperly formatted tables. Incompleteness addresses missing structural elements of the web site, including images that fail to load, "under construction" pages, and tables with empty cells. The third type of flaw, error, includes typographical, grammatical, and factual errors.

LITERATURE REVIEW

Relevant to the study of the effect of presentation flaws is research on trust and credibility, impression formation and impression management.

Trust

Trust has been defined in various ways, often depending on the context in which it appears (Rousseau, et al., 1998). Sultan and Mooraj (2001) found that managers distinguish between two types of trust environments: trust in the relationship among businesses, consumers, and other stakeholders; and trust in the web site and its functionality. The view adopted in this research is of the latter kind, trust in the on-line store via its web site.

Impression Formation

Research on impression formation dates back to 1946 (Asch) and considers the way people perceive others as a process by which an integrated impression is formed from stimulus information that is provided. Early models of impression formation (Asch, 1946; Anderson, 1965) assume that when an individual is presented with information about a previously unknown or unfamiliar person, the individual creates a sort of mental slot in which information is received and processed.

Research has shown that attributes that are negative and that have extreme evaluative meaning weigh more heavily on an individual's impression than neutral items, because of their novelty and unusual nature (Fiske, 1980).

One reason that impression formation is so important is that, according to Cotlier (2001), the first seven seconds that a visitor views a firm's web site are the most crucial as it is within that time period that a prospective customer can be turned off for good.

Impression Management

Impression management, also referred to as self-presentation, is the process whereby individuals seek to control the impressions that other persons form of them (Goffman, 1959; Rosenfeld et al., 2002). The information provided on the web site tends to be imperfect and incomplete and thus requires the consumer to make inferences based on the information presented (Jarvenpaa & Tractinsky, 1999).

RESEARCH MODEL

This study explores the effects of flaws on several outcomes. The research model for this study is shown in Figure 1. From left to right, the user's perception of the different presentation flaws (poor style, incompleteness and error) affects the users' perception of quality of the on-line store. This perceived quality in turn affects the user's level of trust in the on-line store which in turn influences the users' intention to purchase from the on-line store.

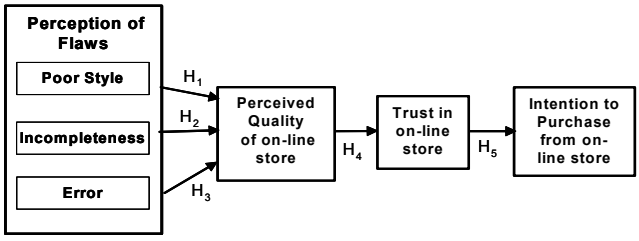


Figure 1. Research Model

Because it is strategically crucial to manage consumers' impressions of the web site, firms need to ensure that any factors that may convey a lack of integrity are reduced. Lynch and Horton (2002) recommend that to convey to users that what the firm is offering is accurate and reliable, high editorial and design standards need to be achieved; "a site that looks sloppily built, with poor visual design and low editorial standards, will not inspire confidence" (p.25).

They further state that the overall organization of the site will have the greatest impact on the user's experience visiting the web site. Furthermore, because of the higher importance of early information and negative information (Anderson, 1965; Fiske & Taylor, 1991), it is important that on-line stores present web sites that are properly formatted and that have an overall organized look.

> H1: A web site that is perceived to be in poor style will result in lower perceived quality of the on-line store than a web site that is not perceived to be in poor style.

On-line stores can manage the impressions that consumers form of the store's web site by establishing legitimacy. Cotlier (2001) asserts that a firm can establish legitimacy by providing users with a finished product in terms of its web site; this can be achieved by avoiding broken links, "coming soon" pages, and images that do not load. Broken links shake the user's confidence with respect to the user's validity and timeliness of the web site's content (Lynch & Horton, 2002). The web site serves as a signal to the consumers; for this reason, Lynch and Horton (2002) warn against letting a site go stale, that is, not checking periodically whether the links to pages outside of the firm's web site are still working. It is important for on-line stores to maintain high standards for their web sites; otherwise the impressions that users have will fall (Rosenfeld, et al., 2002). Users are less likely to come back to the site if they are disappointed with their initial visit; it is always more difficult to attract users back to the site once they have been disappointed (Fiske & Taylor, 1991).

> H2: A web site that is perceived to be incomplete will result in lower perceived quality of the on-line store than a web site that is not perceived to be incomplete.

As Molich and Nielsen state, "spelling errors distract users and make them suspect a generally poor quality" of a system (1990, p.344)." Moreover, spelling errors can be used to form impressions about competency and attention to detail (Liu & Ginther, 2001). In computer-mediated communication, communication style (for example, word choice, paralinguistic cues, typographic information) can beget impression-relevant information; for example, if messages contain several errors, it may be interpreted that the sender is careless or incompetent (Lynch & Horton, 2002). Moreover, Goffman (1959) warns "… we must be ready to examine the dissonance created by a misspelled word…" (p.55) and that "…the impression of reality fostered by a performance is a delicate, fragile thing that can be shattered by very minor mishaps." (p.56).

> H3: A web site that is perceived to contain errors will result in a lower perceived quality of the on-line store than a web site that that is not perceived to contain errors.

Trust is increasingly becoming a significant strategic issue in organizational web site development. Not only is it fragile, as Goffman (1959) stated, but it is also hard to generate, easily lost, and once lost, difficult to regain (Hanowski et al., 1994; Muir & Moray, 1996; Shneiderman, 2000). Fogg and Tseng (1999) concentrate on the trust that forms between individuals and that is mediated by technology: "trust indicates a positive belief about the perceived reliability of, dependability of, and confidence in a person, object, or process" (p.81). It follows that reliability, dependability, and confidence will increase its perceived quality. Furthermore, McKnight et al., (2002) assert that perceived web site quality should positively influence the users' trusting beliefs and trusting intentions as using the web site provides the first

experiential feel of the on-line store's presence and confirms first or initial impressions: "if consumers perceive the Web site is of high quality, they will assume that the Web vendor has positive attributes and will form trusting intentions" (p.341). Thus,

> H4: Perceived quality of the on-line store will influence the user's trust in the on-line store.

Trust facilitates cooperative behavior (Shneiderman, 2000). By trusting someone or something, individuals make themselves vulnerable in a variety of ways. However, individuals trust when, although they are aware that they are vulnerable to harm from others, they do not believe that these others would harm them even if they could (Friedman et al., 2000). Technology designers aim to inspire a cognitive state of trust in users so that users will engage in trusting behaviors, which will enable the transaction to progress without problems (Cassel & Bickmore, 2000).

Low trust leads to hesitation or failure to complete a purchase or disclose personal information (Cassel & Bickmore, 2000; Jarvenpaa & Tractinsky, 1999; Doney & Cannon, 1997). Gefen (2000) examined the relationship between familiarity and trust on electronic commerce and found that trust was a good predictor of intention to purchase. Others demonstrate that trust influences intentions to purchase (Dwyer, et al., 1987; Ku, et al., 2002).

> H5: Trust in the on-line store will influence the user's intention to purchase from the on-line store.

RESEARCH METHODOLOGY

Hypothesis testing was carried out using a between-subjects 3-way fully factorial laboratory experiment, with 20 subjects per cell. Participants were used only once and were randomly assigned to one of eight experimental conditions. This between-subject design avoids any order or learning effects and prevents contamination of subjects' responses on the main task due to manipulation check questions. Eight different versions of the web site were designed, with all possible combinations of presentation flaws (each of 3 flaws absent or present).

The experimental materials consisted of a fictitious web site, with which participants were asked to find specific information on the web site and record the answers. To answer the questions participants had to browse the web site. The task was followed by an on-line questionnaire with questions pertaining to dependent variables measuring perceived quality, trust, and intention to purchase, as well as control variables (computer experience, web experience, and computer efficacy) and three manipulation checks (one for each condition).

Data Analysis

Reliability analyses were calculated for the scales used. All alphas were well over .8, showing adequate reliability for further analysis.

Perception of Flaws

The participants' perceptions of the three types of presentation flaws were recorded by their answers to the manipulation check questions. We were reassured that when a flaw was present, participants perceived the flaw. However, when the flaw was not present participants seemed wary of declaring the site to be flawless. We speculate that participants were reluctant to commit to either the presence or the non-presence of a flaw, for example, in the Incomplete, No Errors, Good Style cell, the score with the highest frequency was 4 (12 out 20 participants), which suggests that participants did not feel comfortable declaring an absence of flaws. Instead they preferred to "straddle the fence." The same phenomenon occurred with the Complete, No Errors, Poor Style and the Complete, Errors, Poor Style treatments.

RESULTS

Perception of poor style and perception of errors were found to be significant predictors of perceived quality of the site (β=-.274, p=.001 and β=-.556, p=.000, respectively). Contrary to our predictions, perception of incompleteness was not found to be a significant predictor of perceived quality of the site.

Regression was used to test Hypotheses 4 and 5. In testing H4 (perceived quality affects trust), a model with site quality as independent variable was significant (F=141.562, p=.000), and explained 63.9% of variance in trust in the on-line store. Perceived quality of the site was found to be a significant predictor of trust in the on-line store (β =.654, t=7.050, p=.000).

In testing H5 (trust affects intention to purchase), a model with trust as the independent variable was significant (F=159.337, p=.000), and explained 49.9% of the variance in intention to purchase from the on-line store.

Table 1 summarizes the findings and indications of support by the data.

H	Expectation	Result
H_1	Perceived Site Quality: Perceived Good Style > Perceived Poor Style	**Supported**
H_2	Perceived Site Quality: Perceived Complete > Perceived Incomplete	**Not Supported**
H_3	Perceived Site Quality: Perceived No Errors > Perceived Errors	**Supported**
H_4	Perceived Site Quality as an antecedent of Trust	**Supported**
H_5	Trust as an antecedent of Intention to Purchase	**Supported**

Table 1. Summary of Findings

ACTUAL FLAWS VS. PERCEPTION OF FLAWS

In addition to the analysis presented above, we also investigated whether it is the perception of the flaw rather than the actual flaw that influences the users' perception of quality of the web site. Furthermore, a test of mediation was used to determine whether the perception of the flaws

mediates the relationship between the flaws and the users' perception of quality.

Participants who were presented with a complete site and perceived it as such reported higher mean scores of perception of quality (3.31) than participants who perceived the site to be incomplete (1.04). Seventy-six participants out of a possible eighty who were presented with an incomplete site perceived it as incomplete. The mean score for perception of quality was 1.87. Interestingly, the mean scores for perceived quality for the incomplete site perceived as such are higher than the mean scores for perceived quality of the complete site perceived as incomplete.

Participants who were presented with a site without errors and perceived it as such reported higher mean scores of perception of quality (3.40) than participants who perceived the site to include errors (2.13). Participants who were presented with a site with errors and perceived the errors reported lower mean scores for perceived quality (1.73) than participants who were presented with the web site with errors but did not perceive them (3.67).

Participants who were presented with a good style site and who perceived it as such reported higher mean scores of perception of quality (3.80) than participants who perceived a poor style (1.61). Participants presented with a poor style site and perceiving it as poor reported lower scores on perceived quality (1.68) than those who did not perceive a poor style (3.25).

From the analysis above, what appears to matter is the participants' perception of some flaw rather than the actual occurrence of it. In all instances, whether or not the flaw was present it was the perception of the flaw that seems to have lowered the scores on perception of quality.

The next section provides the results of analysis of how the perception of flaws may mediate the relationship between the main factors and users' perception of the web site's quality. As per Baron and Kenny (1986), to test for mediation it is necessary to estimate the three following regression equations: (1) the mediator on the independent variable, (2) the dependent variable on the independent variable, and (3) the dependent variable on both the independent variable and on the mediator.

To test the effect of the factors and the perception of flaws on perceived quality of the site, a multiple regression model with perceived quality of the site as the dependent variable was significant ($F=28.36$, $p=.000$), and explained 50.8% of variance in perceived quality of the site. Both the perception of poor style and the perception of errors were found to be significant predictors of perceived quality of the site ($\beta=-.459$, $p=.000$ and $\beta=-.217$, $p=.006$, respectively).

In order to establish mediation, (1) the independent variable must affect the mediator in the first equation, (2) the independent variable must affect the dependent variable in the second equation, and (3) the mediator must affect the dependent variable on the third equation (Baron & Kenny, 1986). The conditions all hold in the predicted direction, and we can state that the perception of the flaws mediates the relationship between the main factors and the dependent variable, perceived quality of the site.

These results illustrate that it is not the presence of a flaw, but rather the perception of the flaw, that affects users' perception of the site's quality. Actual flaws (whether they exist or not) must be perceived as such to affect the site's perceived quality. See Figure 2 for a revised model.

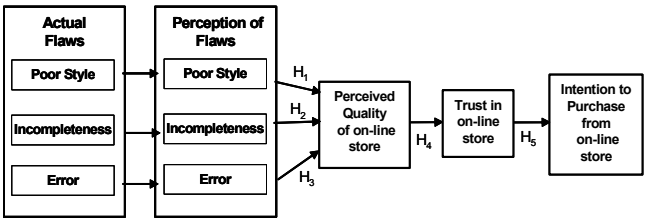


Figure 2. Revised Model

DISCUSSION

This research examined the effects that the perception of errors, incompleteness, and poor style had on users' perceptions of web site quality. More favorable perceptions of quality were reported for sites perceived to be without errors than sites that were perceived to contain errors. As stated in the literature, spelling errors can make users suspect a poor quality of a site.

The perception of poor style also affected users' perception of site quality. More favorable perceptions of quality were reported for users who were presented with good style than for those exposed to poor style.

As predicted, perceived quality of the site was a significant predictor of trust. Users who perceived the site favorably were more likely to trust the site. Consistent with this, users who perceived the quality of the site to be low were less likely to trust the site.

Trust was found to be a significant predictor of purchase intention. Just as predicted in the literature, users who trust the web site are more likely to purchase from the site than users who do not trust the site.

Finally, our results show that the perception of, rather than actual existence of flaws, affects users' perception of site quality. Whether errors, incompleteness, or poor style were actually present did not directly contribute to the users' perception of quality; rather what affected their perception of quality was their perception of the flaw. Because it is the perception of flaws on web sites rather than the actual presence flaws that affects users' quality perceptions it is fundamental for web stores to pay attention to how the features they present are perceived, as opposed to only following generally accepted web site design procedures.

REFERENCES

1. Anderson, C.A. (1991) "How people think about causes: Examination of the typical phenomenal organization of attributions for success and failure," *Social Cognition*, 9, 295-329.
2. Asch, Solomon E. (1946) "Forming impressions of personality," *Journal of Abnormal and Social Psychology*, 41, 258-290.
3. Baron, Reuben M. & Kenny, David A. (1986) "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations," *Journal of Personality and Social Psychology*, 51:6, 1173-1182.
4. Cassel, J. & Bickmore, T. (2000) "External manifestations of trustworthiness in the interface," *Communications of the ACM*, 43(12), 50-56.
5. Cotlier, Moira (2001) "Electronic catalogs: Judging a site by its home page," *Catalog Age*, May 1.
6. Doney, P.M., & Cannon, J.P. (1997) "An examination of the nature of trust in buyer-seller relationships," *Journal of Marketing*, 61, 35-51.
7. Dwyer, F. R., Schurr, P. H., & Oh, S. 1987. Developing buyer-seller relationships. *Journal of Marketing*, 51:2, 11-27.
8. Fiske, S.T. (1980) "Attention and weight in person perception: The impact of negative and extreme behavior," *Journal of Personality and Social Psychology*, 38, 889-906.
9. Fiske, S. & Taylor S.E. (1991) Social Cognition, 2nd edition, New York: Random House.
10. Fogg, B.J. & Tseng, H. (1999) "The elements of computer credibility," Proceedings of Conference on Human Factors and Computing Systems (Pittsburgh, May 15-20), ACM Press, New York, 80-87.
11. Friedman Batya, Kahn, Peter H. & Howe, Daniel C. (2000) "Trust online," *Communications of the ACM*, 43(12), 34-40.
12. Garbarino, E. & Johnson, M. S. (1999) "The different roles of satisfaction, trust, and commitment in customer relationships," *Journal of Marketing*, 63:2, 70-87.
13. Gefen, David (2000) "E-commerce: the role of familiarity and trust," *Omega*, 28(6), 725-737.
14. Goffman, Erving (1959) The Presentation of Self in Everyday Life, New York, NY: Doubleday Anchor, 259.
15. Hanowski, R.J., Kantowitz, S.C., & Kantowitz, B.H. (1994) Driver acceptance of unreliable route guidance information," *Proceedings of the Human Factors Society 38th Annual Meeting*, 1062-1066.
16. Hoffman, D. L., Novak, T. P. & Peralta, M. (1999). "Building Consumer Trust Online," *Communications of the ACM*, 42(4), 80-85.
17. Jarvenpaa, Sirkka L. & Tractinsky, Noam (1999) "Consumer trust in an internet store: A cross-cultural validation," *Journal of Computer-Mediated Communication*, 5(2), http://www.ascusc.org/jcmc /vol7/issue3/jarvenpaa.html.
18. Kantowitz, B.H., Hanowski, R.J., & Kantowitz, S.C. (1997) "Driver acceptance of unreliable traffic information in familiar and unfamiliar settings," *Human Factors*, 39(2), 164-176.
19. Ku, C., Liu, C., Marchewka, J. & Mackie, B. (2002) "A study of consumer's trust in privacy on electronic commerce," *Proceedings of the 12th International Conference on Comparative management*.
20. Liu, Y & Ginther, D (2001). "Managing Impression Formation in Computer-Mediated Communication," *Educause Quarterly,* Nov 3, pp. 50-54.
21. Lynch, Patrick J. & Horton, Sarah (2002) Web Style Guide, www.webstyleguide.com.
22. McKnight, D.H., Choudhury, Vivek & Kacmar, C. (2002) "Developing and validating trust measures for e-commerce: An integrative typology," *Information Systems Research*, 13(3), 334-359.
23. McKnight, D. Harrison, Cummings, L.L. & Chervany, N.L. (1998) "Initial trust formation in new organizational relationships," *Academy of Management Review*, 23(3), 473-490.
24. Molich, Rolf & Nielsen, Jakob (1990) "Improving a human-computer dialogue," *Communications of the ACM*, 33:3, 338-348.
25. Muir, B.M. & Moray, N. (1996) "Trust in Automation: Part II. Experimental studies of trust and human intervention in a process control simulation," *Ergonomics*, 39(3), 429-460.
26. Myers, David G. (1998) Psychology, 5th Edition, New York, NY: Worth.
27. Rosenfeld, Paul, Giacalone, Robert A., & Riordan, Catherine A. (2002) Impression Management: Building and enhancing reputations at work, London, UK: Thomson Learning, 248.
28. Rousseau, Denise M., Sitkin, S.B., Burt, R.S., & Camerer, C. (1998) "Not so different after all: A cross-discipline view of trust," *Academy of Management Review*, 23(3), 393-404.
29. Shneiderman, Ben (2000) "Designing trust into online experiences," *Communications of the ACM*, 43(12), 57-59.
30. Sultan, Fareena & Mooraj, Hussain A (2001) "Designing a trust-based e-business strategy," *Marketing Management*, December, 40-45.

Exploring Website Evaluation Criteria using the Repertory Grid Technique: A Web Designers' Perspective

Felix B Tan
Auckland University of Technology
felix.tan@aut.ac.nz

Lai Lai Tung
Nanyang Technological University
alltung@ntu.edu.sg

ABSTRACT

This study aims to investigate web designers' perceptions of an "effective" website. Twenty web designers were interviewed using Kelly's Repertory Grid Technique in order to elicit factors that they consider important when designing or developing B2C websites. Using grounded theory approach, these elicited data were then classified into 14 meta-categories. The intensive nature of the interviews eventually gave rise to a comprehensive framework that broadens the base of existing web evaluation literature. This framework is based on an adapted Technology Acceptance Model with the 4 dimensions of Perceived Ease of Use, Perceived Usefulness, Perceived Playfulness and Attractiveness.

Keywords

Website evaluation, repertory grid, human-computer interaction.

INTRODUCTION

Research has suggested that poor web-design is turning customers away (Amato-McCoy, 1999). In response, organizations have sought to assess their Web presence through evaluation of their websites and identification of potential problems. However, research on website evaluation to date, has been highly fragmented.

Firstly, it is largely user or customer-focused. Users' views have been studied more extensively because serving users' needs is the primary objective of websites in cyberspace (Bell and Tang, 1998, Dragulanescu, 2002, Whyte et al., 1997). Research has generally placed less emphasis on what web designers consider are important attributes of effective websites.

Secondly, existing research often center their studies on selected aspects of web design which are deemed more important (Aladwani and Palvia, 2001), thereby failing to address website's effectiveness in its entirety. For instance, with the emergence and widespread acceptance of Human-Computer Interaction (HCI) (Hartson, 1998), huge emphasis has been placed on studying websites' usability (Agarwal and Venkatesh, 2002, Cockrell and Jayne, 2002, Hornbaek and Frokjaer, 2001), thus neglecting other qualities such as visual attractiveness.

Finally, most website evaluation research methodology is plagued by one other major drawback: that of having pre-determined structures. Examples include the use of pre-structured questionnaires to collect data (Bell and Tang, 1998, Ranganathan and Ganapathy, 2002) and the use of scripted actions that govern the way participants walkthrough a website in a simulated or laboratory environment (Agarwal and Venkatesh, 2002, Cockrell and Jayne, 2002, Hornbaek and Frokjaer, 2001). Attributes and constructs used in these pre-structured questionnaires are predetermined prior to the start of the study. These methods can limit the scope of information obtained as researchers explore specific selected aspects of web design.

The objective of this study is therefore to investigate website evaluation from the perspective of web designers. It attempts to answer the question "What factors do web designers consider important when designing or developing effective B2C websites?" To circumvent the problem of using pre-determined structures, this research applies an inductive approach, the Repertory Grid Technique (RGT), to elicit a comprehensive set of B2C e-commerce website evaluation constructs and their definitions based on the experiences of web-designers. The application of the RGT has gained some attention recently in the Information Systems (IS) field (Tan and Hunter, 2002, Hunter and Beck, 2000, Whyte et al., 1997) and is comprehensively discussed in Tan & Hunter (2002).

LITERATURE REVIEW

The literature review revealed the use of three common theoretical lenses in the research to date. These lenses have been used to guide researchers in coming up with the evaluation criteria. The first is the Technology Acceptance Model (TAM) (Davis, 1989). Studies that use TAM as a framework or guide to website evaluation include Lee and Lee (2003), Koufaris (2002), Schubert and Dettling (2002) and Benbunan-Fich (2001). These studies use the TAM framework to look at factors affecting the acceptance of websites. However, we find that for these studies, the measures for and criteria affecting PEOU and PU are different even though they deploy the same framework.

The second lens is the application of flow theory to the evaluation of websites, either as a standalone theory (Koufaris, 2002) or used in conjunction with TAM for website evaluation. Flow has been defined as "the holistic sensation that people feel when they act with total involvement" (Csikszentmihalyi, 1975). Many studies also looked at perceived playfulness or perceived enjoyment, as part of the framework for their analyses (Katerattanakul, 2002, Liu and Arnett, 2000). Perceptions of playfulness and shopping enjoyment originated from flow theory, which argues that when a person is in a flow, they shift into a mode of experience and become absorbed in their activity. Therefore, websites that promotes playfulness and shopping enjoyment are likely to be more engaging to the consumers of the website.

A third lens that has been used to examine website evaluation is the Human-computer Interaction (HCI) lens, where the notion of usability is a key theme (Agarwal and Venkatesh, 2002, Palmer, 2002). The theoretical foundation for HCI studies are grounded in psychology and cognitive science. Many studies looked at design features that will help improve the usability of websites, including attractiveness and interactivity (Skadberg and Kimmel, 2003, Lindgaard and Dudek, 2003).

Despite the differences in the theoretical lenses applied to web evaluation studies, there were commonalities in their findings. A case in point is Agarwal and Venkatesh (2002), which examined usability of websites via the HCI lens. The usability categories in that study included ease of use, and "content", which the authors argued was akin to "perceived usefulness". Both perceived ease of use and perceived usefulness are TAM constructs. Additionally, as we have mentioned earlier, there is no consistency on the measures for and criteria affecting similar constructs using the same framework, such as those for PU and PEOU. Finally, for the majority of research that we have reviewed, there is no theoretical framework to guide the criteria selection. One common approach to the selection of criteria is through a review of published academic and/or practice-oriented literature in the general area of website evaluation (Kim et al., 2003). Because the criteria are not derived from theory, these researches are often fragmented: there are no theoretical justifications for the criteria selections and no assurances that the selected criteria are comprehensive and relevant to the measurement of website effectiveness. The Repertory Grid Technique that we will describe in the next section will help in the selection of appropriate criteria and addresses these problems as well as those mentioned earlier on.

RESEARCH METHODOLOGY

Sampling

The sample of intended subjects (web-designers) was drawn from a listing of 1012 software retailers published in the e-Source Directory[1] (2001). Given the intensive nature of the RGT, a relatively small sample size (about 15 to 25 subjects) is often sufficient in eliciting a comprehensive list of constructs for the purpose of a study (Ginsberg, 1989, Dunn et al., 1986, Tan and Hunter, 2002). A modified systematic sampling procedure was applied to the listing, starting with a random record and applying a selection interval of 5. A total of 20 web-designers agreed to participate in this study.

The Repertory Grid Interview Process

Six pilot interviews were conducted with university students with web design experience. As a result, we were able to standardize the RGT interview process and confirm our procedures for the actual interviews. The interview involved 3 steps – element selection, construct elicitation (involving triading and laddering), and then the rating of elements along each elicited construct.

The relevant elements for our study are B2C websites. A minimum of six elements is required in order to provide sufficient triads for use in the second step. Based on Nielsen/NetRatings Singapore Internet Audience Activity Report for April 2000 (Osman, 2002), the websites included are *Yahoo!, MSN, Singapore Telecom, Pacific Internet, AOL websites* and *Lycos*. One week prior to the interviews, we emailed the participants general details of the interview and requested them to surf the six websites to familiarize themselves with the sites. Just before the start of the interview, we confirmed with the participants that they have surfed the 6 stipulated websites. We also gave the participants an option to browse unscripted, the websites for up to 10 minutes at their own workstations before the interview commenced if they wished to do so. At the commencement of the interview, an overview of the study was provided to the participant. To reduce interviewer's bias, all instructions were read from prepared notes to ensure that all subjects received the same set of instructions.

Construct elicitation aims to identify meanings, in the form of bipolar constructs, that subjects attach to the elements (Marsden and Littler, 2000). Two interviewing methods, "*triading*" and "*laddering"*, are employed to achieve this. *Triading* (Kelly, 1955) involves the participant selecting three elements (websites) at random. The participant is then asked to identify, how two of them are similar and different from the third, in terms of what s/he, as a web-designer, consider important when designing or developing websites. The labels for similarity and difference identified form a bipolar construct eg. good navigation – poor navigation. The "*laddering*" method is then used to educe in-depth explanations of the bipolar construct. For example, the respondent might be asked which pole of the construct

[1] This is a directory containing vendor listings of services, software and hardware retailers.

they prefer (good or poor navigation), or how and why they think that particular aspect (Marsden and Littler, 2000) affects the websites. The elicitation process is then repeated to identify more constructs, until the participant cannot add any new constructs to the ones s/he already named earlier.

At the 3rd step of the interview the participant is asked to rate all elements based on the attributes elicited. Each element is rated independently, on a scale of 1 to 7, where 1 represents the construct pole, and 7, the contrast pole. By using a rating scale, the subject is accorded greater freedom when sorting the constructs as they are not forced to take side with either the construct or its contrast pole (Beail, 1985).

To conclude the interview, participants were requested to fill up a demographic sheet and indicate their relevant expertise as web-designers.

Analysis of Data

Using RGT, participants were generally given vast freedom in determining the perceived similarities and differences within each triad, and providing personal interpretations pertaining to the constructs elicited (Hunter and Beck, 2000). This section describes how we analyzed and classified our rich findings into common themes. A walkthrough using examples from our data will be presented. We used a three-layer classification scheme to categorize the data collected, namely **construct class, conceptualization** and **meta-category.** The interpretations and labels we assigned to each of these layers were informed by literature on website evaluation.

In order to facilitate classification of the constructs into conceptualizations, we grouped the constructs into classes. Conceptualizations were formed from the construct classes. Meta-categories of conceptualizations were derived using the grounded theory approach (Strauss, 1987).

RESULTS AND ANALYSIS

Eleven males and nine female web designers participated in this study. They were mostly between 21-30 years of age and had more than 2 years experience in web design. On average, all participants surf the Internet several times a day for up to 5 hours each time.

Forty-six conceptualizations were obtained from the identified construct classes and fourteen meta-categories were derived using grounded theory approach. Table 1 presents a sample of the 14 meta-categories, their underlying conceptualizations, construct classes and a sample of the constructs that make up the categories.

The 14 meta-categories and their definitions are presented in Table 2.

Category	Definition
Graphics Usage	Refers to the purpose for which they are used and extent of usage, including the quality of graphics and how they are being organized
Text Usage	Relates to the purpose of using a text-based interface
Content/ Information	Mainly the scope (wide or specific) and quality of information
Updates	Design considerations with respect to websites that require updating
Layout/ Space Usage	How web space is utilised to present the features and functions across the pages within the website
Presentation of information	Concerns the implications of using colours, fonts and display styles to present information
Headlines	Covers the objective and extent of headline usage
Categorisation of Information	Refers to ways of grouping information on the pages in order to facilitate reading
Navigation	The features used in designing the site that facilitates transition from page to page
Colour use	Concerns the usage and choice of colours used
Visual Appearance	How the website looks and the impact it effects
Advertisements/ Pop-ups/ Animation	The purpose and extent of usage of such features
Downloading Time	Factors in designing that impacts speed of downloading
Establishing Website's Identity	Various methods designers use to portray its unique image

Table 2: Definitions of the Meta-Categories

DISCUSSION

From above, we have identified the set of criteria designers consider when evaluating websites. The findings of this study (i.e., meta-categories, conceptualizations and construct classes) represent a comprehensive list of important considerations web designers should take into account when designing and developing B2C websites.

Additionally, in order to enable the set of criteria to be tested in future studies, and to facilitate an understanding of how our results contribute to increasing user acceptance and website effectiveness, we propose a framework to encompass these criteria at the meta-categories level. This framework is informed by research in Technology Acceptance Model (TAM) (Davis, 1989) and flow theory (Csikszentmihalyi, 1975, Deci and Ryan, 1985). It also incorporates research from the HCI literature, especially in the area of design of websites.

In the proposed research framework, we propose that the set of design factors (the 14 meta-categories of criteria surfaced above) affects perceived usefulness (PU), perceived ease of use (PEOU) and perceived playfulness (PP). The set of causal links in this research model is consistent with Moon and Kim's (2001).

We propose that 2 design factors, content/information and updates, are positively related to PU From the literature, research has indicated that content is a factor determining usability (Agarwal and Venkatesh, 2002), website quality (Aladwani and Palvia, 2001) and website success (Palmer, 2002). Additionally, information usefulness and information service has been cited as factors affecting PU (Lee and Lee, 2003). As information usefulness is similar to content while information service is similar to updates, this provide support for our proposition that the 2 factors, content/information and updates, are positively related to PU.

In terms of factors affecting PEOU, we propose a positive relationship between the following 5 design factors and PEOU: navigation, categorization of information, downloading time, presentation of information and headlines. There is support for the inclusion of these 5 design factors from the literature. In Agarwal and Venkatesh (2002), for instance, ease of use consist of 3 subcategories: (i) goals, having clear and understandable objectives, (ii) structure, referring to the organization of the site and (iii) feedback, provision of information of progress. In Cox and Dale (2002), ease of use refers to the clarity of purpose, design (for usability during navigation) and communication.

As anticipated, there are many design factors that goes to improving attractiveness and interactivity of the websites. We adopt Skadberg and Kimmel's (2003) definition of attractiveness of a Website as the representation's richness and quality and interactivity to refer to the response triggered by the user. In this study, our variable "attractiveness and interactivity" are design factors that will promote richness, quality and response to the website. We categorize the remaining 7 design factors as appropriate graphic and text usage, visual appearance/look, layout and space usage color usage, advertisements/popup/animation and establishing website's identity as factors affecting attractiveness and interactivity of the design.

Lastly, we propose that perceived playfulness is determined by the attractiveness, interactivity and ease of use of the website. Literature on playfulness and flow has supported this contention. Skadberg and Kimmel (2003), for instance, found that attractiveness and interactiveness demonstrated a causal relationship with the flow experience

CONCLUSION

In this study, we translated web-designers' practice into the set of criteria they consider when evaluating websites. The application of the RGT yielded rich and relevant qualitative data from the interviews. The findings of this study (i.e., meta-categories, conceptualizations and construct classes) represent a comprehensive list of important considerations web designers should take into account when designing and developing B2C websites.

REFERENCES

1. Agarwal, R. and Venkatesh, V. (2002) Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability, *Information Systems Research,* 13**,** 168-186.
2. Aladwani, A. M. and Palvia, P. C. (2001) Developing and validating an instrument for measuring user-perceived web quality*, Information & Management,* 39**,** 467-476.
3. Amato-McCoy, D. M. (1999) In RT: Technol Mag Retail Exec, Vol. 18.
4. Beail, N. (1985) In *Repertory grid technique and personal constructs : applications in clinical & educational settings*(Ed, Beail, N.) Brookline Books, Cambridge, MA.
5. Bell, H. and Tang, N. (1998) The Effectiveness of Commercial Internet Web Sites: A users' perspective*, Internet Research,* 8**,** 219-228.
6. Benbunan-Fich, R. (2001) Using Protocal Analysis to Evaluate the usability of a commercial website*, Information & Management,* 39**,** 151-163.
7. Cockrell, B. J. and Jayne, E. A. (2002) How Do I Find an Article? Insights from a Web Usability Study*, The Journal of Academic Librarianship,* 28**,** 122-132.
8. Cox, J. and Dale, B. G. (2002) Key quality factors in Web site design and use: an examination*, International Journal of Quality & Reliability Management,* 19**,** 862-888.
9. Csikszentmihalyi, M. (1975) Play and Intrinsic rewards*, Humanistic Psychology,* 15**,** 41-63.
10. Davis, F. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology*, MIS Quarterly,* 13**,** 319-339.
11. Deci, E. L. and Ryan, R. M. (1985) Intrinsic Motivation and Self-Determination in Human Behaviour*,* Plenum Press, New York.
12. Dragulanescu, N.-G. (2002) Website Quality Evaluations: Criteria and Tools*, International Information & Library Review,* 24**,** 247-254.

13. Dunn, W. N., Cahill, A. G., Dukes, M. J. and Gomsberg, A. (1986) In *Policy Analysis: Perspectives, Concepts, and Methods*(Ed, Dunn, W. N.) JAI Press, Greenwich, CT, pp. 355-375.
14. E-Source Directory (2001) E-Source 2001, Kompass South East Asia Ltd.
15. Ginsberg, A. (1989) Construing the business portfolio: A cognitive model of diversification, *Journal of Management Studies,* 26**,** 417-438.
16. Hartson, H. R. (1998) Human- Computer Interaction: Interdisciplinary roots and trends, *Journal of Systems and Software,* 43**,** 103-118.
17. Hornbaek, K. and Frokjaer, E. (2001) Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces, *Human-Computer Interaction,* 3**,** 293-300.
18. Hunter, M. G. and Beck, J. E. (2000) Using Repertory Grids to Conduct Cross-Cultural Information Systems Research, *Information Systems Research,* 11**,** 93-101.
19. Katerattanakul, P. (2002) Framework of effective web site design for business-to-consumer Internet commerce, *INFOR Journal,* 40**,** 57-70.
20. Kelly, G. A. (1955) The Psychology of Personal Constructs, New York: Norton.
21. Kim, S. E., Shaw, T. and Schneider, H. (2003) Web site design benchmarking within industry groups, *Internet Research: Electronic Networking Applications and Policy,* 13**,** 17-26.
22. Koufaris, M. (2002) Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior, *Information Systems Research,* 13**,** 205-223.
23. Lee, K. C. and Lee, S. (2003) A Cognitive Map Simulation Approach to Adjusting the Design Factors of The Electronic Commerce Web Sites, *Expert Systems with Applications,* 24**,** 1-11.
24. Lindgaard, G. and Dudek, C. (2003) What is this evasive beast we call user satisfaction?, *Interacting with Computers,* 15**,** 429-452.
25. Liu, C. and Arnett, K. P. (2000) Exploring the factors associated with Web site success in the context of electronic commerce, *Information & Management,* 38**,** 23-33.
26. Marsden, D. and Littler, D. (2000) Exploring consumer product construct systems with the repertory grid technique, *Qualitative Market Research,* 3**,** 127-143.
27. Moon, J.-W. and Kim, Y.-G. (2001) Extending the TAM for the World-Wide Web context, *Information & Management,* 38**,** 217-230.
28. Osman, H. (2002) channelnewsasia.com.
29. Palmer, J. W. (2002) Web Site Usability, Design, and Performance Metrics, *Information Systems Research,* 13**,** 151-167.
30. Ranganathan, C. and Ganapathy, S. (2002) Key dimensions of business-to-consumer websites, *Information & Management,* 39**,** 457-465.
31. Schubert, P. and Dettling, W. (2002) In HICSS-35Hawaii.
32. Skadberg, Y. X. and Kimmel, J. R. (2003) Visitors' flow experience while browsing a Web site: its measurement, contributing factors and consequences, *Computers in Human Behavior*.
33. Strauss, A. L. (1987) Qualitative Analysis for social Scientists, Cambridge University Press, New York.
34. Tan, F. B. and Hunter, M. G. (2002) The Repertory Grid Technique: A Method for the Study of Cognition in Information Systems, *MIS Quarterly,* 26**,** 39-57.
35. Whyte, G., Bytheway, A. and Edwards, C. (1997) Understanding user perceptions of information systems success, *Journal of Strategic Information Systems,* 6**,** 35-68.

Meta-Category	Conceptualization	Construct Class	Construct Example
Content/ Information	Scope of Information	Wide variety / General Information	- wide variety of content to attract larger audience
		Specific Information	- information based on user interests
		Specific Corporate Information	- provides information on company and products for corporate website (target specific audience)
	Quality of Information	Breadth and Depth of Information	- should not have so much variety that quality of information suffers (too broad but no quality)
Updates	Information / content / feature update	Frequency of updates	- frequent updates makes website more user-friendly, attract users to visit more frequently
		Characteristics to facilitate frequent updates	- use of pre-defined design that allows frequent updates, changing content only for each update
	Design update	Up-to-date design	- use of new, up-to-date design, graphics always changing

Table 1: Sample Meta-categories, Conceptualizations and Construct Classes

Usability and Efficacy Reactions to Object-Orientation: The Impact of Prior Knowledge

Liping Liu
University of Akron
liping@uakron.edu

ABSTRACT

In this paper, we examine how prior knowledge impacts usability and efficacy reactions to object-oriented techniques. We develop research hypotheses based on the multiconstraint theory of analogical reasoning. We empirically test the hypotheses in an open learning setting. We observed a significant interaction effect: the subjects with prior knowledge on either data or process modeling technique perceived greater difficulty and less confidence in learning object-oriented techniques than novices as well as those who have prior knowledge on both structured techniques. Prior knowledge explained 19% of the variance in both usability and efficacy reactions and, as a common cause, partially explained their correlation.

Keywords

Usability, self-efficacy, object-oriented techniques.

INTRODUCTION

Most current systems analysts were trained in structured techniques. Most information technology (IT) curricula are still teaching structured techniques as the primary topic for systems analysis and design. However, with the recent standardization of the unified modeling language (UML), the trend in software development is moving toward more object-orientation, which is believed to be in many ways different from its structured counterparts and requires a different mindset in modeling business problems. To respond to the radical change, many organizations are faced with the task of retraining their existing analysts as well as new hires.

There exist studies examining how prior knowledge on process models impacts the learning of object-oriented (OO) techniques using objective measures, such as task performance (Agarwal et al., 1996b, Boehm-Davis and Ross, 1992, Morris et al., 1999), cognitive effort (Morris et al., 1999), and cognitive differences (Vessey and Conger, 1994, Lee and Pennington, 1994). In this study we draw attention to a different inquiry — how systems analysts with prior knowledge react to object-orientation. In particular, we consider two trainee reactions — the usability of OO techniques and the self-efficacy of learning OO methodology — as the effectiveness criteria for retraining systems analysts.

Among many trainee reactions, usability (difficulty) is the only one that predicts actual learning such as post-training knowledge and task performance (Warr and Bunce, 1995). Self-efficacy is an antecedent to and consequent of other training outcomes (Gist et al., 1989, Gist et al., 1991), and measures the substantive value of training (Agarwal et al., 1996a). In addition, how systems analysts make behavioral choices is more based on their subjective beliefs rather than objective counterparts. Their after-retraining work attitude and job satisfaction also depends on these subjective beliefs. Therefore, it is important that we examine these trainee reactions to better manage the technology transition for organizations and design effective retraining programs for software designers.

HYPOTHESES DEVELOPMENT

Prior Knowledge: Knowledge is internalized information related to concepts, procedures, and judgments, and a justified personal belief that increases one's capacity to take action. It can be internally represented as IF-THEN rules, mental models, or propositions, exist in one of three progressive forms: declarative knowledge, procedural knowledge, and schemas and scripts, and be classified into six hierarchical levels: recall, comprehension, application, analysis, synthesis, and evaluation.

Usability: Usability refers to the degree to which one believes that using a system is free of effort (Davis, 1989). It captures the cognitive and emotional effort required to master training materials (Warr and Bunce, 1995). There exists extensive research on how to design usable systems for non-technical end users (Adler and Winograd, 1992). However, our knowledge on the usability of development tools is sparse and the results are inconclusive.

Self-Efficacy: Self-efficacy (SE) refers to the degree to which one is confident in performing a specific task (Gist and Mitchell, 1992). It relates to motivational and behavioral concepts such as proactive attitudes, adaptability to new technology, and learning and achievement. In the IT context, computer SE refers to the judgment by an individual of his or her capability to use an information technology (Compeau and Higgins, 1995). Marakas et al. (1998) made a further distinction between computer SE and task-specific SE. Following the same distinction, in this study we define SE as an individual's estimate of his or her capability to perform OO modeling tasks.

Analogical Reasoning

Learning theories all recognize the role of analogical reasoning in learning. Proposition-based theories posit that learning is a process of making proposition-based inferences; incoming information are compared against stored knowledge, represented as propositions, for assessing their similarities, which can then be used to create a new instance in memory or refine existing knowledge. Production-based theories, which assume knowledge is represented as IF-THEN rules, posit that the learner first draws heavily on analogies and examples to understand how the declarative knowledge is applied to problem solving. Then the procedure knowledge is compiled into schemas, scripts, or other abstract knowledge structures so that exercising the knowledge becomes automatic. Schemas and scripts are activated unconsciously based on similarities when interpreting new concepts or events and analogical reasoning is further enhanced (Gick and Holyoak, 1983).

Thus, while learning OO techniques, individuals with prior knowledge tend to draw analogies back to more familiar structured techniques and maps new concepts onto something familiar. When they have surface knowledge on structured techniques, they represent the knowledge as a set of rules (procedural knowledge) and facts (declarative knowledge), and map OO concepts and skills into individual elements in structured techniques. Analogical mappings tend to be made at superficial levels (Gentner, 1988) such as model elements and relationships. However, after possessing deep knowledge, individuals represent it as a more abstract structure, use the structure to evaluate incoming information for relevance, and place OO concepts and skills into the overall structure according to the similarities. Analogical mappings tend to be made at higher levels such as modeling objectives and cognitive modeling tasks.

There are some empirical observations on the use of analogies in learning OO techniques. Nelson et al. (2002) found that procedural developers tend to map "object" to the familiar concepts of "module," "function," or "database record" and map "class" to the concepts of "database table" or "structure" and so on. Detienne (1995) found that procedural programmers structure OO programs by functional similarity and execution order rather than by class memberships. Pennington et al. (1995) found that procedural analysts decompose a problem driven by actions on the data rather than by domain entities. Based on these findings, we developed an extensive list of analogical mappings, including mappings of elements, relationships, overall models, modeling objectives, and cognitive activities. These mappings cover all the essential OO concepts and skills for the rational unified process, the *de facto* industry standard of OO development process.

Holyoak and Thagard (1989, 1995) found that the use of analogy is guided by a number of general constraints that jointly encourage coherence in analogical thinking. They proposed three broad classes of constraints that form the basis of the so-called *multiconstraint theory*. First, the analogy is guided to some extent by direct similarity of the elements involved. Second, the analogy is guided by a pressure to identify consistent structural parallels between roles in the source and target. These first two constraints form a pressure to establish an isomorphism — a set of consistent, one-to-one correspondences — between the elements of the source and target. Third, analogical thinking is guided by what the analogy is intended to achieve. Holyoak and Thagard (1997) further suggested that the multiple constraints — similarity, structure, and purpose — do not operate like rigid rules dictating the interpretation of analogies. Instead they function more like the various pressures that guide an architect engaged in creative design, with some forces converging, others in opposition, and their constant interplay pressing toward some satisfying compromise that is internally coherent.

The multiconstraint theory implies that the ease of analogical reasoning depends on how much the three constraints can be satisfied and how much compromise one has to make. The easier it is to identify the isomorphism of elements and their relationships between the source and target, the easier one feels about performing the analogical reasoning. The easier it is to achieve the reasoning goals, the more favorable one feels about the ease of learning of the target. Therefore, how prior knowledge affects the usability of OO techniques depends on the extent to which OO concepts and skills can be mapped to structured counterparts.

According to empirical analogical mappings, individuals with prior knowledge on both data- and process-oriented techniques can map all essential OO concepts and skills to their familiar ones. In contrast, individuals with knowledge on either data- or process-modeling techniques alone can only map a portion of them while finding many others to be difficult. Thus, when there is knowledge in both data- and process-modeling techniques, there is greater ease of making analogical reasoning, leading to a more favorable perception on the usability of OO techniques:

H1: Individuals with prior knowledge on both data- and process-modeling techniques perceive the usability of OO techniques more favorably than those with knowledge on either data- or process-modeling techniques alone.

Novices have no analogies to make. They approach the learning task by using general problem-solving strategies such as "divide and conquer." They anchor usability to their general beliefs. In contrast, those with prior knowledge will make an adjustment to reflect their experience of analogical reasoning although their judgment still anchors to the general beliefs (Venkatesh, 2000). In particular, for those with prior knowledge on data- or process-modeling techniques, when they find it difficult to coherently map all OO concepts and skills onto those they are familiar with, their perception will be negatively adjusted:

H2: Compared to novices, individuals with prior knowledge on either data- or process-modeling techniques alone perceive the usability of OO techniques less favorably.

The Determinants of Self-Efficacy

Although experience influences efficacy perceptions, it is the cognitive appraisal that ultimately determines SE. Gist and Mitchell (1992) proposed that three types of information cues are involved in forming SE: task requirements analysis, attributional analysis, and resource analysis. Task requirements analysis produces inferences about what it takes to perform at various levels. The attributional analysis involves judgments about why particular performance occurred in the past. The resource analysis examines the availability of specific resources and constraints for performing the task at various levels.

Among the three SE information cues, different cues may be used in assessing SE estimates depending on the assessor's experience and task characteristics. When the task is fairly novel or when it has been observed only, one may invoke in-depth and detailed analysis of task requirements as well as resource constraints as the primary information cue for SE judgments. When the task has been performed personally and frequently in the past, the individual is likely to rely more heavily on his or her interpretation of the causes of pervious performance levels and to use interpretations as the primary determinant of SE. In general, judgments about efficacy become more automatic as experience with a task increases.

Learning OO techniques is a novel task to all trainees. Their experience is at best an observer's. Therefore, trainees will most likely use in-depth analysis of task requirements and resource constraints as the primary information cue for their SE judgments. At the same time, analyzing the skill and effort requirements for performing OO analysis bears a striking similarity to perceiving how easy it is to learn OO techniques. Thus, we have the following three anticipations:

H3: Individuals with prior knowledge on both data- and process-modeling techniques have greater self-efficacy in performing OO analysis than those with knowledge on either data- or process-modeling techniques.

H4: Novices have greater self-efficacy in performing OO analysis than the individuals with prior knowledge on either data- or process-modeling techniques.

H5: Self-efficacy is positively correlated with usability; the more favorably one perceives the usability of OO techniques, the more confident he or she feels about performing OO analysis.

RESEARCH DESIGN

We conceptualize prior knowledge using two variables. We use KDM to represent prior knowledge on data models and KPM to represent prior knowledge on process models. We control each variable at two levels: 0 (absence) and 1 (presence) and follow the 2 × 2 factorial design involving four groups of subjects, where Group A consists of subjects with knowledge on both data and process models; Group B on data models; Group C on process models; and Group D consists of novices who have no prior exposure to either models.

To implement the design, we recruited potential subjects from senior classes at a large Midwest American university. We requested the rosters of all current and previous classes and screened each candidate with respect to his or her prior knowledge on data and process models. After the screening, we selected 131 trainees to participate in this study. We controlled prior knowledge through relevant courses and provided additional pre-training if necessary. For example, the instructors gave five weeks of extensive lectures and exercises on data modeling techniques to Group B and the same amount of preparation on process modeling techniques to Group C. In addition to regular lectures, these subjects were assigned to solve 20 design problems, one exam, and one large, real business project to fulfill their course requirements. The pre-training treatment was meant to provide equivalent coverage of the same topic in industry training and to prepare the subjects for entry-level systems analyst positions.

We conducted the study using an open learning setting, where trainees worked on their own to learn written materials (Warr and Bunce, 1995). After finishing prior knowledge control, we provided each subject with a training material on OO modeling. The material covers UML, OO concepts such as inheritance, encapsulation, and polymorphism, and how to develop use case and class diagrams to model business problems.

After the two-week open learning period, we conducted training evaluation in an examination setting. As a part of examination, we administered a short quiz consisting of 5 screening questions to ensure that the trainees actually read the training materials. A trainee was dropped from the study if he or she did not score at least 4 points. Eventually, we ended up with 72 subjects and 18 in each controlled group. Among them, 41 were males and 31 females. 52% of them majored in Information Systems and 48% in other business areas. All subjects had about the same level of maturity and computer experience.

Training evaluation consists of two parts. First, we gave the trainees a real systems analysis task and asked them to create an OO analysis model as the blueprint for the system to be developed. Then, each subject was asked to respond to a survey regarding his or her efficacy and usability reactions.

Self-Efficacy: To develop a measure for SE, we followed the five-point framework proposed by Marakas et al. (1999); we focused on the subject's perceived ability to perform a specific task while avoiding the ability assessments on cross-domain or general-domain skills. In object modeling, a subject needs to identify objects, attributes, and methods based on data and functional requirements, and discern object relationships based on data navigation and behavior collaboration. Accordingly, we developed seven questions that assess one's estimated ability to perform each specific task.

Usability: We selected three items from Davis (1989) with no modifications: Easy to Learn, Easy to Become Skillful, and Easy to Use. Then we considered the differences between using a system, and learning OO techniques. The most significant difference is that the latter requires a lot more effort in understanding and comprehending concepts and applying them creatively while the former demands more effort in interacting with the system. Therefore, we dropped the two items related to interaction: Controllable and Flexible and modified and expanded the item "Clear and Understandable" into two items that ask whether OO concepts are straightforward and whether it is easy to comprehend them. To capture the cognitive effort aspect of usability (Goldstein and Gilliam, 1990), we added two items that assess how comfortable a subject feels. Finally, we ended up with eight items in the 7-point Likert scale for usability.

RESEARCH RESULTS

To assess the efficacy of scale items, we conducted reliability analyses. The correlations between SE items range from 0.47 to 0.85 and between usability items from 0.45 to 0.78. The Cronbach alphas are respectively 0.93 for SE and 0.92 for usability. The indices are very high compared to the acceptable level 0.7, demonstrating the convergent validity of the items. To ensure that the items for the same construct measure a single trait whereas items for different constructs measure distinct traits, we conducted a principal factor analysis with Varimax rotation. Using the Kaiser eigenvalues criterion, we extracted two factors that collectively explained 69.6% of the variance in all items. The rotated factor matrix shows that all the items cleanly loaded onto the correct latent constructs.

Tables 1 and 2 show the results of testing H1 and H2. As they show, the mean usability of Group D (3.889) is higher than that of Group B (3.017) and the difference is significant at the level $\alpha = 0.01$. Similarly, the mean usability of Group D (3.889) is higher than that of Group C (2.989) and the difference is significant at the level $\alpha = 0.01$. Therefore, Hypothesis H2 is strongly supported by the data. By comparing Group D with Groups B and C combined, we found H2 is even more significantly supported at the level $\alpha = 0.005$. The support for Hypothesis H1 can be similarly analyzed. The mean usability of Group A (4.044) is significantly higher than that of Group B (3.017) at $\alpha = .005$, than that of Group C (2.989) at $\alpha = .005$, and than that of Groups B and C combined (3.003) at $\alpha = .001$. Thus, H1 is strongly supported by the data.

Group	Size	Mean	Std. Dev.	Error
A	18	4.044	.957	.226
B	18	3.017	1.050	.247
C	18	2.989	1.087	.256
B & C	36	3.003	1.053	.176
D	18	3.889	1.049	.247

Table 1: Group Mean Usability

Comparisons	T-Value	DF	Sig.
A vs. B	3.070	34	.002***
A vs. C	3.093	34	.002***
A vs. B & C	3.529	52	.001***
D vs. B	2.493	34	.009***
D vs. C	2.528	34	.008***
D vs. B & C	2.918	52	.003***

Table 2: T-Tests of Usability

Tables 3 and 4 summarized the results of testing Hypotheses H3 and H4. They show that the mean SE of Group A is significantly higher than that of both Groups B and C at the level $\alpha = 0.001$. Thus, H3 is strongly supported. The support for H4 is relatively weaker. The mean SE of Group D is higher than that of both Groups B and C. The difference between Group D and Group C is significant at $\alpha = 0.1$ and between Group D and Groups B and C combined is significant at $\alpha = 0.05$. However, the difference between Groups D and B is not significant at the level $\alpha = 0.1$.

Group	Size	Mean	Std. Dev.	Error
A	18	465.56	85.21	20.09
B	18	335.56	105.34	24.83
C	18	307.78	137.69	32.45
B & C	36	321.67	121.64	20.27
D	18	393.89	172.50	40.66

Table 3: Group Mean Efficacy Indices

Comparisons	T-Value	DF	Sig.
A vs. B	4.071	34	.000***
A vs. C	4.134	34	.000***
A vs. B & C	4.488	52	.000***
D vs. B	1.224	34	.115
D vs. C	1.655	34	.054*
D vs. B & C	1.783	52	.040**

Table 4: T-Test of Self-Efficacy

Although not explicated, H1-H4 jointly predicts an interaction effect of prior knowledge. To validate it, we conducted two ANOVA tests using KDM and KPM as two fixed factors and usability (SE) as the dependent variable. The test results show a strongly significant interaction effect, which is significant at $\alpha = 0.001$ and is able to predict 18.8% of the variance in both usability and SE.

To test H5, we conducted a regression analysis using SE to predict usability. The result shows a Pearson correlation 0.52 with t-value = 5.074, which is significant at $\alpha = 0.001$ in a 2-tailed t-test. Thus, H5 is strongly supported. The regression model is significant at $\alpha = 0.001$ with a F-value = 25.75. If it is correct that SE determines usability (Compeau and Higgins, 1995, Venkatesh, 2000), the result here suggests that SE can predict 26.9% of the variance in usability.

CONCLUSION

Before discussing contributions, we shall note that the use of student trainees may affect external validity. The same concern also affects other similar studies. However, since colleges are still teaching structured techniques, the subjects in this study are representative of the population of at least new graduates, whom organizations often have to re-train to do object-oriented analysis and design. After all, the goal of the research design was to maximize the internal validity and provide a precise control of prior knowledge, which would be difficult in field studies.

This study improves our understanding on the transition from structured to OO techniques, and sheds light on the debate about revolutionary vs. evolutionary theories (Sircar et al., 2001). The existing studies have mixed findings based on objective measures. In contrast, this study suggests that not only the presence of prior knowledge but also the types of the knowledge have different effects. For example, we found that individuals knowing process models perceived greater difficulty and less confidence in learning OO techniques than novices. However, with addition of knowledge on data models, the effect reverses; individuals having knowledge on both data and process models perceive less difficulty and more confidence. Therefore, in terms of usability and efficacy measures, OO techniques represent an evolutionary change from structured ones.

Our findings have a few implications for IT managers. It is commonly believed that usability and self-efficacy predict task performance, job satisfaction, and other work-related behavioral and attitudinal variables. Managers who desire to implement OO techniques should target those individuals with prior knowledge on both data and process models and those with no prior exposure to structured techniques at all; these people are more likely to bring desirable consequence after training or retraining. Also, our findings contradict the common concern that prior knowledge may interfere with the learning of OO techniques. On the contrary, this study found that prior knowledge helps improve trainee reactions to OO techniques, which in turn improve actual learning.

Information Systems is a field full of constant changes. More often than in any other field, IT workers see not just incremental adjustments but fundamental shifts in the way they use technologies. In just the last two decades, we have seen databases evolve from flat files, to hierarchical, to relational, and to OO models, and operating systems from DOS, to Windows, and to Web-based user interfaces. At each turn, IT workers are forced to transfer their existing skills and learn new ones. Whether a transition is successful or not is often measured by their after-transition job performance and satisfaction, which in turn are determined by their efficacy and usability reactions. The current study makes a contribution by introducing the multiconstraint theory to study these phenomena. Future research could apply the theory to other contexts and examine, for example, how prior knowledge on legacy systems predicts trainee reactions to new systems.

REFERENCES

1. Adler, P. S. and Winograd, T. A. (1992) Usability: turning technologies into tools, Oxford University Press, New York, NY.
2. Agarwal, R., Prasad, J. and Zanino, M. C. (1996a) In SIGCPR/SIGMIS '96Denver, CO, pp. 259-267.
3. Agarwal, R., Sinha, A. and Tanniru, M. (1996b) International Journal of Human and Computer Studies, 45, 639-667.
4. Boehm-Davis, D. and Ross, L. (1992) International Journal of Man-Machine Studies, 36, 1-19.
5. Compeau, D. R. and Higgins, C. A. (1995) MIS Quarterly, 19, 189-211.
6. Davis, F. D. (1989) MIS Quarterly, 13, 319-340.
7. Detienne, F. (1995) Human Computer Interaction, 10, 129-169.
8. Gentner, D. (1988) Child Development, 59, 47-59.
9. Gick, M. and Holyoak, K. (1983) Cognitive Psychology, 15, 1-38.
10. Gist, M., Schwoerer, C. and Rosen, B. (1989) Journal of Applied Psychology, 74, 884-491.
11. Gist, M. E. and Mitchell, T. R. (1992) Academy of Management Review, 17, 183-211.
12. Gist, M. E., Stevens, C. K. and Bavetta, A. G. (1991) Personnel Psychology, 44, 837-861.
13. Goldstein, I. L. and Gilliam, P. (1990) American Psychologist, 45, 134-143.
14. Holyoak, K. J. and Thagard, P. (1989) Cognitive Science, 13, 295-355.
15. Holyoak, K. J. and Thagard, P. (1995) Mental leaps: Analogy in creative thought, MIT Press, Cambridge, MA.
16. Holyoak, K. J. and Thagard, P. (1997) American Psychologist, 52, 35-44.
17. Lee, A. and Pennington, N. (1994) International Journal of Human Computer Studies, 40, 577-601.
18. Marakas, G. M., Johnson, R. D. and Yi, M. Y. (1999) Indiana University, Bloomington, IN.
19. Marakas, G. M., Yi, M. Y. and Johnson, R. D. (1998) Information Systems Research, 9, 126-163.
20. Morris, M., Speier, C. and Hoffer, J. (1999) Decision Sciences, 30, 107-135.
21. Nelson, H. J., Armstrong, D. and Ghods, M. (2002) Communications of the ACM, (Forthcoming).
22. Pennington, N., A. L. and Rehder, B. (1995) Human Computer Interaction, 10, 171-226.
23. Sircar, S., Nerur, S. P. and Mahapatra, R. (2001) MIS Quarterly, 25, 457-471.
24. Venkatesh, V. (2000) Information System Research, 11, 342-365.
25. Vessey, I. and Conger, S. (1994) Communications of the ACM, 37, 102-111.
26. Warr, P. and Bunce, D. (1995) Personnel Psychology, 48, 347-375.

Evaluation of the Impacts of Data Model and Query Language on Query Performance

Hock Chuan Chan
National University of Singapore
chanhc@comp.nus.edu.sg

Lian Xiang
National University of Singapore
g0201698@nus.edu.sg

ABSTRACT

It is important to understand how users can utilize database systems more effectively to enhance performance. A major research interest is to evaluate and compare user performance across different data models and query languages. So far, experiments have tested combinations of model plus language. An interesting theoretical and practical question is: how much of the performance difference is caused by the data model itself, and how much by the additional query language syntax? A cognitive model of query processing suggests measurement at two stages. The data model has impact at the first stage, and the model with the query language syntax together has the impact at the second stage. An experiment that compares the objected-oriented and relational models and query languages at the two stages provides fresh results.

Keywords

Data model, query language, user performance, empirical study, query stage

INTRODUCTION

Databases form an integral part of organizational systems. The evaluation and explanation of how users can make effective use of databases is an important area of information systems research, which has seen a steady stream of empirical studies (Aversano et al., 2002; Bowen and Rohde, 2002; Borthick et al., 2001; Chan et al., 1999; Siau et al., 1997; Owei and Navathe, 2001).

Many studies have been done on relative comparison of data models and query languages. For experiment studies on modeling performance, there is one main database variable: the data model. Differences in modeling performance can be readily attributed to the model. For studies on query performance, the main database variable is a combination of a data model and a query language. Studies have typically required subjects to write queries. The process involves a combination of data model and query language knowledge. So far, differences in user query performance have been attributed to the combination of data model and query language. Findings in the literature do not show whether the data model or the query language has more impact on query performance, leaving a lingering doubt on the interpretation and even validity of the findings.

This study addresses this issue in a comparison of the objected-oriented and relational models. It compares the user performance differences because of the impact of data model itself, and also compares the differences because of the additional impact of a query language within a model. Section 2 presents a cognitive model of the query process, which allows us to measure the effect of the model alone and the effect of the model plus query syntax. Section 3 presents the research methodology, followed by the results of the experiment. Lastly the conclusion is given.

A COGNITIVE MODEL OF DATABASE QUERY

This section provides a cognitive perspective on how data model and query language influence query performance. Ogden (1985) proposes a three-stage cognitive model of database query: query formulation stage (stage 0), query translation stage (stage 1), query writing stage (stage 2). The model is illustrated in Figure 1. The query formulation stage is concerned about real world data. An example is "Who are the faculty members in the business school?"

Based on the question from stage 0, users decide what elements of the data model are relevant, and the necessary operations. This is the query translation stage. For example, the output of this stage is "The faculty relation is needed, the column name is to be selected, and there is a condition for school name to be 'business'". This output need not be written down. In the query writing stage, users have to specify the output from stage 1 into the formal syntax of a query language. A simple example in SQL is: "select name from faculty".

There are many other models that involve similar steps in the query process. For example, the model by Mannino (2001) has two steps: from problem statement to database representation, and from the database representation into a database query language statement. Reisner (1977) proposes a process where a user will generate a set of lexical items and also generate a query template, followed by the merging of the lexical items with the template to generate the final query. The correspondence to the query translation and query writing stages are clear. This model is also related to the idea of semantic and articulatory

distances as used in Liao and Palvia (2000). The articulatory distance is about stage 2, where users need to articulate the answers in a formal syntax. The semantic distance is about stage 1.

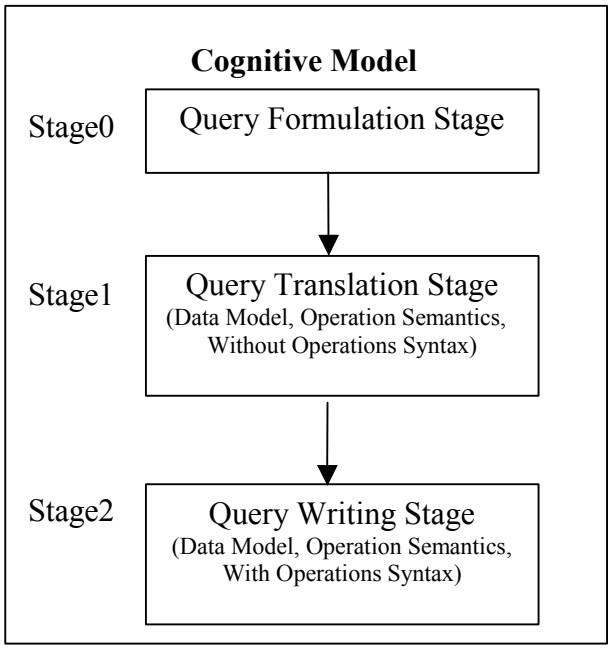


Figure 1. Query Model

Prior experiments on query performance have measured user performance after stage 2. If we can measure user performance after stage 1, and after stage 2, it will be possible to have a better understanding of the relative impact of model alone and the model with the additional language syntax.

RESEARCH MODEL AND METHODOLOGY

Research Model and Variables

Performance is influenced by four major factors: data-model/query language, task, user and system characteristics (Reisner, 1981; Chan et al., 1993). The independent variable is the abstraction level of the data model, set at two levels: the conceptual level where subjects used a version of OO model (O_2) with OQL, and the logical level where subjects used the relational model with SQL. The research model highlighting the comparison within stages (across models with / without query language) and comparison across stages (within the object-oriented model or within the relational model) is shown in Figure 2.

There have already been many empirical studies on the effects of data models and query languages (Liao and Palvia, 2000; Chan et al., 1993; Wu et al., 1994), which suggest that the conceptual level models (OO and ER) will lead to better user performance than the logical level model (relational), at least for the query writing stage. The different performance between the abstraction levels has been attributed to the type and amount of knowledge. For example, at the conceptual level, the objects such as entities and relationships are closer to the real world semantics which users are familiar with. On the other hand, at the logical level, the constructs are relations and primary keys / foreign keys which users are not familiar with. With ideal implementations, a relationship at the conceptual level can be specified quite easily (e.g. employee.department in a typical object-oriented query), compared to the unfamiliar specification of joins at the logical level (e.g. employee.eno. = department.empno in an SQL query). A more detailed description of the abstraction levels can be found in Chan et al. (1993).

So far, there are no studies that measure user query performance at the two different stages. We make the following hypothesis:

H1: Subjects using O_2 / OQL will perform better (in terms of accuracy, time and confidence) than subjects using relational / SQL for the query writing stage.

H2: Subjects using O_2 / OQL will perform better than subjects using relational / SQL for the query translation stage.

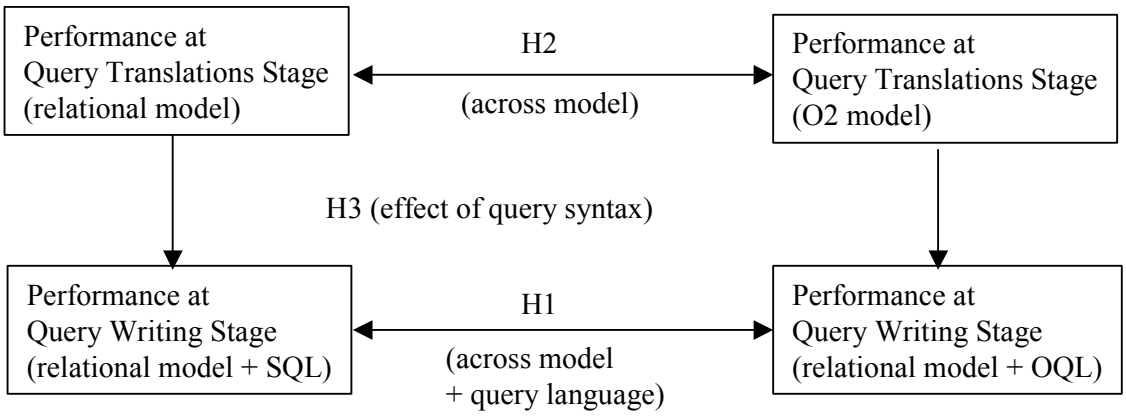


Figure 2. The Research Model (Performance is measured by accuracy, time taken and confidence)

H3: Performance will be better at the query translation stage than at the query writing stage, for both groups.

The third hypothesis compares two different tasks. The difference can be attributed to the effect of the additional query language syntax.

Research Method and Process

A laboratory experiment was conducted to test the hypotheses. Each subject performed eight questions for both stages. These queries covered the basic queries that are commonly made on the relational model. Extraneous factors are controlled through randomization (for individual characteristics) or through standardization across groups (for interface characteristics).

Subjects were trained before they took the query test. The program displayed the questions one by one. They first finished query translation and then the query writing for each query. Each subject was given a relational schema or a diagram of an OO model, on paper. The test materials are in the Appendix. The query answers, the time taken in seconds, and the confidence level for each query were recorded by the computer.

EXPERIMENT RESULT AND DISCUSSION

The mean and standard deviation (in parenthesis) for accuracy, time, and confidence are shown in Table 1. Since the stage 1 data do not follow a normal distribution, non-parametric tests, using SPSS, were used. The Mann-Whitney independent sample test is used to compare between groups. The results show that the OO group is significantly more accurate than the relational group for stage 1 (z=-4.09, p=0.001) and stage 2 (z=-4.66, p=0.001). Time and confidence do not show significant differences. Thus, hypotheses 1 and 2 are both supported for accuracy measure, and not for time and confidence measures. This result corroborates previous studies for stage 2 (Chan et al., 1993; Wu et al., 1994), and provides new evidence for stage 1 differences.

	Relational Model		OO Model	
	Query Translation	Query Writing	Query Translation	Query Writing
Accuracy	4.58 (.43)	3.31 (.53)	4.85 (.29)	4.38 (.56)
Time	50.5 (21.5)	169.5 (47.3)	50.1 (16.4)	146.7 (45.4)
Confidence	4.80 (.53)	4.17 (.72)	4.81 (.36)	4.25 (.77)

Table 1. Mean (Standard Deviation) of Measures

Table 2 shows the results across query stages, using non-parametric Wilcoxon signed ranks test. Both groups show better performance at the query translation stage than at the query writing stage, for all measures of accuracy, time and confidence. Hypothesis 3 is fully supported. This shows that the query language syntax imposes significant additional difficulty to the query process. Furthermore, we find that many subjects with fully correct answers in stage 1 made serious mistakes in stage 2. Thus, even when subjects fully know what they want (the data structures and operations in query translation stage), they have difficulties putting that in the formal syntax required by a query language. These findings apply to both the relational and OO groups.

Data Model	Accuracy	Time	Confidence
Relational Model	z=-3.921[a] p=0.000**	z=-3.920[b] p=0.000**	z=-3.935[a] p=0.000**
OO Model	z=-2.952[a] p=0.003*	z=-3.920[b] p=0.000**	z=-3.525[a] p=0.000**

a Based on positive ranks. b Based on negative ranks.
* Significant at p<0.05 ** Significant at p<0.01

Table2. Non-parametric Test across Stages

Figure 3 illustrates accuracy performance at different stages of the cognitive model. At stage 0, we assume that the subjects can understand the meaning of query questions (and so a value of 5 is given). At this point, we are able to return to the questions posed earlier.

1. How much of the overall drop in performance (from the ideal top score) can we attribute to the data model alone, and how much to the particular query language within a model? At stage 1, performance shows a slight drop from stage 0 (9% for relational model, and 3% for OO model). At stage 2, performance drops by a very large amount (28% for SQL, and 10% for OQL) compared to stage 1. These numbers indicate the relative difficulties imposed on the users by the data model, and by the query language (additional to the model). The syntactical requirements of SQL with relational model and OQL with OO model cause about 3 times the difficulties caused by the data model alone. What we see here is that users basically do know what they want (and they can even perform the operations mentally to identify the right data values, on a small data set), but they have difficulties expressing them in a formal query language.
2. The OO model leads to better query results than the relational model, supporting findings in the literature. How much of this difference can be attributed to the models, and how much to the languages? This study shows that models alone cause a small 0.27 (out of 5) difference in accuracy. But at stage 2, when the query languages have been added to the data model, the difference is much bigger: 1.07. Thus, only about one third of the overall difference across models/query languages can be attributed to the models, and the other two thirds to the languages. This leads to the third question.

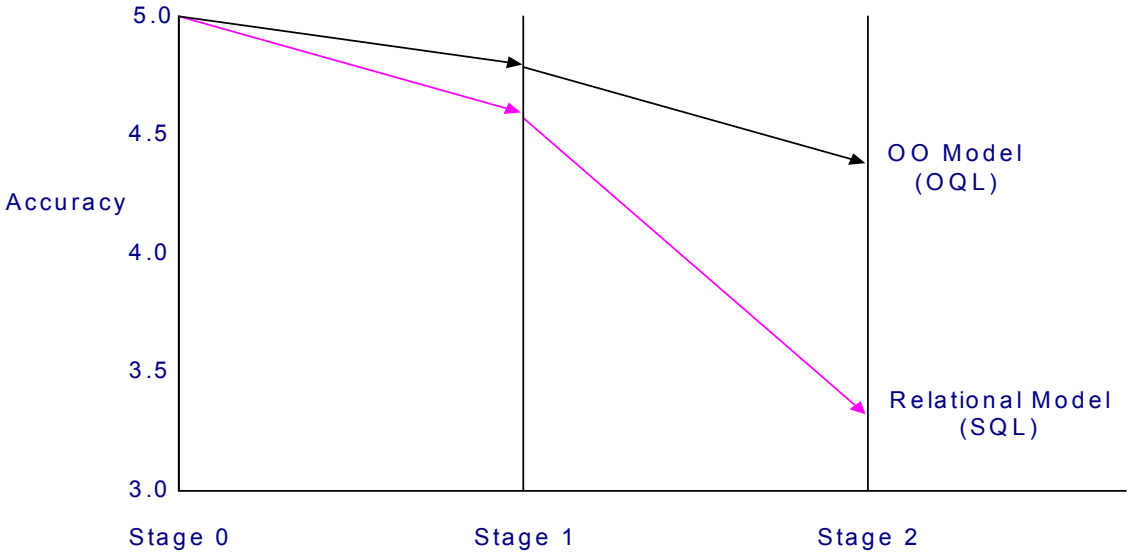


Figure 3. Query Performance at Each Query Stage

3. This study and others in the literature show a consistent finding that the OO/ER models are better than the relational model for query performance. But one doubt that is raised from figure 3 is: instead of using SQL, can we get a better relational language such that the overall query performance across models + languages will show no difference? We note that the relational model's performance at stage 1 is higher than the OO model's performance at stage 2. If a good language with little syntax difficulty can be found for the relational model, it could be possible that the overall query writing performance will show no difference. This is a challenge for researchers to develop a more user friendly relational textual query language.

CONCLUSION AND FUTURE WORK

Our experiment illustrates a finer approach to measure user query performance, based on a 3-stage cognitive model of query processing. By measuring query performance at different stages of the query process, we demonstrated the impacts of data models alone, for the object-oriented and the relational models, and the additional impacts of the query languages. The study shows that a higher abstraction level model leads to higher user performance for both query stages. The study also shows that the data model itself has a relatively small impact (about a third), and the query language has the remaining two thirds. It shows that generally users do know what they want, but have difficulty expressing it in a formal query language.

ACKNOWLEDGEMENT

We thank all the reviewers for their valuable comments.

REFERENCES

1. Aversano, L., G. Canfora, A. De Lucia, S. Stefanucci (2002). "Understanding SQL through iconic interfaces," *Proceedings 26th Annual International Computer Software and Applications Conference*, 703-708.
2. Borthick, A.F., P. L. Bowen, D. R. Jones and M. H. K. Tse (2001). "The effects of information request ambiguity and construct incongruence on query development," *Decision Support Systems*, 32(1) 3-25.
3. Bowen, P. L., F. H. Rohde (2002). "Further evidence of the effects of normalization on end-user query errors: an experimental evaluation," *International Journal of Accounting Information Systems*, 3(4) 255-290.
4. Chan, H.C., Wei, K.K. and Siau, K.L. (1993). "User-Database Interface: The Effect of Abstraction Levels on Query Performance," *MIS Quarterly*, 17(4) 441-464.
5. Chan, H. C., Tan, B. C. Y. and K. K. Wei (1999). "Three Important Determinants of User Performance for Database Retrieval," *International Journal of Human-Computer Studies*, 51(5), 895-918.
6. Liao, C. and P. C. Palvia, (2000), "The Impact of Data Models and Task Complexity on End-User Performance: an Experimental Investigation", *International Journal of Human-Computer Studies,* 52(5) 831-845.
7. Mannino, M.V. (2001). *Database Application*

Development and Design, McGraw-Hill Company, Inc.
8. Ogden, W.C. (1985). "Implications of a Cognitive Model of Database Query: Comparison of a Natural Language, a Formal Language, and Direct Manipulation Interface", *ACM SIGCHI Bulletin,* 18(2) 51-54.
9. Owei, V., S. B. Navathe (2001) "Enriching the conceptual basis for query formulation through relationship semantics in databases," *Information Systems*, 26(6) 445-475.
10. Reisner, P. (1977). "Use of psychological experimentation as an aid to development of a query language," *IEEE Transactions on Software Engineering*, SE-3(3), 218-229.
11. Reisner, P. (1981). "Human Factors Studies of Database Query Languages: A Survey and Assessment." *ACM Computing Surveys.* 13(1) 13-31.
12. Siau, K., Y. Wand, I. Benbasat, (1997). "The relative importance of structural constraints and surface semantics in information modeling," *Information Systems*, 22(2/3) 155-170.
13. Wu, C.Z., Chan, H.C., Teo, H.H. and Wei, K.K. (1994). "An Experimental Study of Object-Oriented Query Language and Relational Query Language for Novice Users," *Journal of Database Management,* 5(4) 16-27.

APPENDIX: DATABASE AND QUERIES FOR THE EXPERIMENT

This appendix contains the relational schema and the OO model, and the set of questions used in the experiments.

Query Questions:

1. Show the department name and city.
2. Show the engineers' name and professions.
3. Show the names of employees who head any project.
4. Show the names of employees who work in the sales department.
5. Show the names of employees who work in the same department as Jack.
6. Show the names of employees with higher salaries than Jack's.
7. List the names of managers who manage more than one department.
8. List the names of engineers who do not head any project.

Data Models:

Employee (<u>number</u>, name, salary)

Engineer (<u>number</u>, profession)

Manager (<u>number</u>, rank)

Department (<u>number</u>, name, city)

Project (<u>number</u>, name)

Work (<u>employee number</u>, <u>department number</u>, date)

Figure A1. The Relational Schema

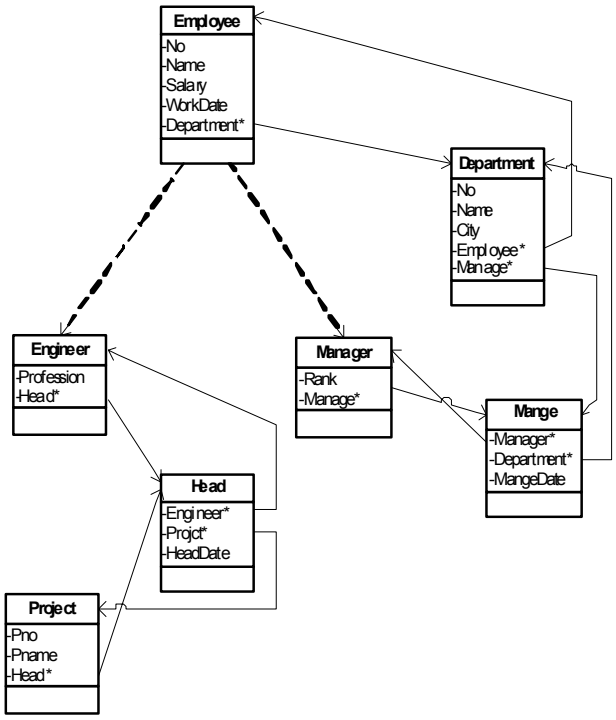


Figure A2. The Object-Oriented Data Model

End User Query Performance: The Interaction of User Characteristics and Information Request Ambiguity

Paul L. Bowen
University of Queensland
p.bowen@business.uq.edu.au

Fiona H. Rohde
University of Queensland
f.rohde@business.uq.edu.au

Chiu Yueh Alice Wu
University of Queensland
a.wu@business.uq.edu.au

ABSTRACT

This paper investigates the effects of personality characteristics on individuals' abilities to resolve ambiguity in an information retrieval environment. In particular, this research examines the effects on query performance of the interaction of personality characteristics (as measured using the NEO PI-R) with information requests that contained extraneous, syntactic, or both extraneous and syntactic ambiguities. The results indicate that ambiguity affected performance. The results also show that various personality dimensions significantly affect end-users' abilities to compose accurate queries. Neuroticism, agreeableness, openness to experience, and conscientiousness affected the number of errors made in the query formulations. Conscientiousness affected the length of time taken to compose the queries and neuroticism affected the confidence end users had in the accuracy of their queries. In addition, the results indicated that, while the personality dimensions affected performance, there was no interaction between the personality dimensions and ambiguity.

Keywords

Personality, Information Retrieval, NEO PI-R, Ambiguity, Performance.

INTRODUCTION

In today's highly competitive business environments organizations are encouraging managers and other end users to query information repositories themselves. Frequently the queries these end users compose are to satisfy information requests, posed in natural language, they receive from stakeholders. Being in natural language, these information requests often contain ambiguities.

This paper investigates the effects of personality characteristics on individuals' abilities to resolve ambiguities in an information retrieval environment. In particular, this research examines the effects on query performance of the interaction of personality characteristics with information requests that contained extraneous, syntactic, or both extraneous and syntactic ambiguities. The personality dimensions examined are neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. These dimensions were measured using the revised NEO Personality Inventory (NEO PI-R).

THEORETICAL FOUNDATIONS AND HYPOTHESIS DEVELOPMENT

Ambiguity and Information Retrieval

Within the realm of information retrieval, a person receives an information request, interprets the information request, and prepares a query to retrieve the required information from a data repository, i.e., a database, data mart, or data warehouse. The presence of ambiguity in an information request is likely to lead to multiple valid interpretations of that information request. Because of the multiplicity of valid interpretations, the information retrieved may not be the information desired by the person initially making the request. Use of potentially inappropriate information can have significant negative ramifications on business decision-making processes.

Walton (1996) identified six ambiguity types: lexical, syntactical, inflective, pragmatic, emphatic, and suggestive. Axelsen et al. (2001) expanded Walton's taxonomy to include a seventh type of ambiguity, extraneous ambiguity. Their results indicate that syntactic and extraneous ambiguity strongly affect people's ability to correctly translate information requests into queries that extract the information desired by the requestor.

SYNTACTIC AMBIGUITY

Syntactic ambiguity, i.e., structural or grammatical ambiguity, often results in recipients being unclear or mistaken as to the subject or the object of a sentence. One of the most common forms of syntactic ambiguity is the use of indefinite pronouns where the pronoun's antecedent is not clear.

EXTRANEOUS AMBIGUITY

Extraneous ambiguity arises when information is included that is not necessary. Some extraneous communications are clearly not relevant to the task at hand and may even be misleading. Axelsen et al. (2001) found that excess information impairs people's ability to recognize critical elements of an information request. The extraneous information could, however, confuse other recipients and cause them to misinterpret the information request, e.g., by expanding the scope of the query.

This research extends the work undertaken by Axelsen et al. (2001) by examining whether some personality types can resolve the syntactic and extraneous ambiguity better than other personality types.

Personality

Personality refers to the cognitive and affective structures maintained by individuals to facilitate their adjustments to the events, people, and situations they encounter (Gough, 1976). Personality variables such as locus of control, ambiguity tolerance, cognitive behaviors, and attitude affect individual's ability to articulate and evaluate designated tasks and ultimately impact on MIS success (Zmud, 1979). When applied within the context of specific occupations and organizations, personality variables are also significant predictors of job performance (Day and Silverman, 1989; George, 1992). The Five Factor model categorizes personality traits into five major dimensions: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). This model of personality has become the dominant basis for investigating the effects of personality traits (Goldberg, 1993) and has been used in studies of job performance (Barrick and Mount 1991, 1993), career success (Judge et al., 1999), job satisfaction and work adjustment (Tenopyr, 1993).

Ambiguity, Personality, and Information Retrieval

NEUROTICISM, AMBIGUITY, AND INFORMATION RETRIEVAL

Once an information request has been received, to formulate an accurate query, end-users interpret the components of the request relative to the tables and attributes in the data structure. When individuals undertake more demanding attentional tasks, higher levels of neuroticism are associated with worse task performance (Szymura and Wodniecka, 2003). Introducing ambiguity into an information request makes the query formulation task more demanding. This increased stress invokes a negative emotional response in persons with higher levels of neuroticism and negatively affects their performance. Thus:

> H1: End users with higher levels of neuroticism faced with ambiguity in information requests perform worse[1] when formulating queries than end users with lower levels of neuroticism.

EXTRAVERSION, AMBIGUITY, AND INFORMATION RETRIEVAL

The six facets of extraversion indicate that persons scoring highly on extraversion tend to be more outgoing, high-spirited, active, excitement seeking, and cheerful. The task of composing queries for particular information requires little use of these exuberant traits. To perform the task well and to resolve the ambiguity relies on people's ability to focus on concepts and ideas. Individuals exhibiting high levels of extraversion may find the task of composing a query more difficult and stressful, as they are required to suppress their enthusiasm for life and focus more intently on the task at hand. This increased stress is likely to have a negative effect on their performance. This analysis leads to the following hypothesis:

> H2: End users with higher levels of extraversion faced with ambiguity in information requests perform worse when formulating queries than end users with lower levels of extraversion.

OPENNESS, AMBIGUITY, AND INFORMATION RETRIEVAL

Traits in the openness to experience dimension reflect the process of using cognition, intelligence, and contemplativeness together with unconventionality (Judge et. al 1999). Individuals with low levels of openness to experience are more conventional and prefer familiar and recognizable items. Conversely, individuals with higher levels of openness to experience have the ability to expand potential innovation, have positive attitudes towards learning, and are more motivated (Barrick and Mount 1991). These individuals are also more willing to embrace novel ideas "as well as experience emotions more keenly" (Costa and McRae, 1992, pp15).

Determining and extracting the information required from an information system requires creative mapping of real world ideas and concepts to a database structures (Wand and Weber, 1990). Individuals with higher level of openness to experience possess traits of flexibility and creativity and hence, should find the task less demanding. When individuals are also confronted with ambiguity, the traits of flexibility and creativity allow such individuals to better resolve the ambiguity and ultimately perform better than individuals who do not possess such traits. This analysis leads to the following hypothesis:

> H3: End users with higher levels of openness to experience faced with ambiguity in information requests perform better[2] when formulating queries than end users with lower levels of openness to experience.

AGREEABLENESS, AMBIGUITY, AND INFORMATION RETRIEVAL

Individuals with high levels of agreeableness are compassionate and cooperative whereas individuals with low levels of agreeableness tend to be non-compliant, critical, sceptical, and more competitive. Individuals exhibiting a low level of agreeableness are better able to recognise, articulate, and evaluate the information necessary to make accurate analyses. Individuals with higher levels of agreeableness, i.e., straightforwardness, ingenuousness, and modesty, tend to misinterpret and overlook relevant information. The process of query

[1] Perform worse is operationalized as making more errors, taking more time, and being less confident when formulating queries.

[2] Perform better refers to make less errors, take less time, and be more confident when end users formulate queries.

composition requires that essential information is recognised in the information request.

Following the execution of each query, end-users must evaluate, on an objective and logical level, the accuracy and relevance of the results generated. The difficulty of these tasks increases with excess information or structural request ambiguities. Individuals who are less agreeable with greater scepticism and possess critical thinking skills tend to be better able to recognise the presence of ambiguities and to be better equipped to resolve them. This analysis leads to the following hypothesis:

> H4: End users with higher levels of agreeableness faced with ambiguity in information requests perform worse when formulating queries than end users with lower levels of agreeableness.

CONSCIENTIOUSNESS, AMBIGUITY, AND INFORMATION RETRIEVAL

The process of composing queries from information requests is iterative. When presented with excess information or syntactical ambiguity in an information request, end users with higher levels of conscientiousness are able to carefully, logically, and persistently work through the request. Because they are more diligent when constructing queries, they are likely to produce more accurate queries and to be more confident in their query results. This discussion leads to the following hypothesis:

> H5: End users with higher levels of conscientiousness faced with ambiguity in information requests perform better[3] when formulating database queries than end users with lower levels of conscientiousness.

METHOD

Research design, participants, and data collection

In a laboratory experiment, 75 participants composed and executed queries in SQL for an Oracle database. All participants received a set of instructions containing the scenario, the details of tasks to be performed, the data dictionary, and the entity-relationship diagram. Participants were assigned to one of four groups according to their GPA in such a manner as to make the IS experience and training of the groups as equivalent as possible. The groups were then randomly assigned to a treatment sequence.

The ambiguity treatment was manipulated via the information requests. Each of the four groups received information requests in all four formulations: clear, extraneous, syntactic, and both extraneous and clear. Similarly, for each information request there were four different formulations (one for each different type of ambiguity). This design required participants in all four groups to experience each type of ambiguity.

The participants had two hours to construct, as accurately as possible, appropriate queries for as many of the twelve information requests as they could. After each query attempt was executed, the system displayed the SQL result. Participants could revise their queries as many times as they wished. When they indicated that they were satisfied with the result they obtained for a particular request, participants were prompted to specify their confidence that the query results were correct. After indicating their confidence levels, participants proceeded to the next information request.

Operationalizing the Variables

The dependent variables were the number of semantic errors in each query, the time taken to compose each query, and the participant's confidence in each of their queries. The independent variables were the types of ambiguities present and each participant's scores on each of the five dimensions of the NEO personality inventory. To ensure that the effects of task complexity and end-user ability had been taken into account, two more variables, query complexity and grade point average (GPA), were used as covariates in the statistical analysis. The information requests were generally of increasing complexity and, thus, information request (query) number was used as a covariate.

RESULTS

Effects of Personality and Ambiguity on Semantic Errors Made By End Users During Query Composition

None of the interactions between the five personality dimensions and the four types of query formulations (clear and the three ambiguous) were significant. That is, individuals with various levels of the five different personality dimensions were neither more nor less successful in resolving ambiguities contained within the information requests. The results of an analysis of covariance (ANCOVA) (Table 1, Panel A) indicate, however, a significant association between four of the five personality dimensions and number of query errors made. In particular, the results indicate that neuroticism, openness, and agreeableness significantly affected the number of semantic errors. Conscientiousness had a marginal affect on the number of semantic errors. The parameter estimates show the direction of the effects of each of the personality dimensions on the number of semantic errors.

[3] Perform better refers to make less errors and be more confident when end users formulate queries. They are, however, likely to take more time composing their queries, e.g., refining their query formulations when earlier query formulations produce results the user deems unreasonable.

Source	R^2	*df*	Mean Square	F Value	Pr > F	Parameter Estimate
Model	0.1501	10	348.99	10.65	0.0001	
Error		603	32.76			
Ambiguity		3	338.72	10.34	0.0001	
Neuroticism		1	134.75	4.11	0.0430	-0.0417
Extraversion		1	9.05	0.25	0.6203	-0.0116
Openness		1	155.52	4.75	0.0297	-0.0545
Agreeableness		1	171.17	5.23	0.0226	0.0626
Conscientiousness		1	94.10	2.87	0.0906	0.0399
Query Number		1	1957.83	59.77	0.0001	0.6664
GPA		1	165.18	5.04	0.0001	-0.7710

Panel A. Effect of Model on Number of Semantic Errors Made During Query Composition

Source	R^2	*df*	Mean Square	F Value	Pr > F	Parameter Estimate
Model	0.0757	10	270.57	4.49	0.0001	
Error		603	54.80			
Ambiguity		3	420.76	7.68	0.0001	
Neuroticism		1	5.89	0.11	0.7432	0.0087
Extraversion		1	10.98	0.20	0.6545	-0.0135
Openness		1	52.66	0.96	0.3274	-0.0317
Agreeableness		1	0.00	0.00	0.9951	0.0002
Conscientiousness		1	292.53	5.34	0.0212	0.0703
Query Number		1	200.84	3.66	0.0560	0.2134
GPA		1	990.63	18.08	0.0001	-1.8882

Panel B. Effect of Model on Amount of Time Taken During Query Composition

Source	R^2	*df*	Mean Square	F Value	Pr > F	Parameter Estimate
Model	0.0805	10	10.98	5.28	0.0001	
Error		603	2.08			
Ambiguity		3	6.60	3.17	0.0238	
Neuroticism		1	18.86	9.07	0.0027	-0.0156
Extraversion		1	0.44	0.21	0.6470	-0.0027
Openness		1	0.43	0.21	0.6496	0.0029
Agreeableness		1	0.63	0.30	0.5825	-0.0038
Conscientiousness		1	1.25	0.60	0.4387	0.0046
Query Number		1	63.53	30.56	0.0001	-0.1200
GPA		1	5.76	2.77	0.0965	0.1440

Panel C. Effect of Model End User Confidence During Query Composition

TABLE 1. Effect of Types of Personality and Ambiguity on Performance

Effects of Personality and Ambiguity on Time Taken By End Users During Query Composition

None of the interactions between the five personality dimensions and the four types of query formulations (clear and the three ambiguous) were significant. That is, individuals with various levels of the five different personality dimensions neither took more or less time to construct the queries from information requests containing different types of ambiguity. The results of an analysis of covariance (ANCOVA), reported in Table 1 Panel B, indicate, however, a significant association between conscientiousness and the time taken for the query composition task. The parameter estimate shows that, as predicted, persons exhibiting higher levels of conscientiousness took longer to complete each query.

Effects of Personality and Ambiguity on End User Confidence During Query Composition

None of the interactions between the five personality dimensions and the four types of query formulations (clear and the three ambiguous) were significant. That is, individuals with various levels of the five different personality dimensions were neither more nor less confident in the queries they produced from information requests containing various types of ambiguity. The results of an analysis of covariance (ANCOVA), reported

in Table 1 Panel C, indicate, however, a significant association between one of the five personality dimensions and number of query errors made. In particular, the results indicate that neuroticism ($F_{1,603}$=9.07, p=0.0027, two-tail test) significantly affected the confidence that end users had in the accuracy of their queries. The parameter estimate shows that, as predicted, persons exhibiting higher levels of neuroticism were less confident in the accuracy of their queries.

CONTRIBUTIONS, LIMITATIONS, AND FUTURE RESEARCH

The results show that various personality dimensions significantly affect end-users' abilities to compose accurate queries. Neuroticism, agreeableness, openness to experience, and conscientiousness affected the number of query errors. Conscientiousness affected the length of time taken to compose the queries and neuroticism affected the confidence the end users had in the accuracy of their queries. The results indicate that, while the personality dimensions affected performance, the various subsets within each dimension did not vary significantly in their ability to resolve ambiguities.

These results have important implications for improving managerial end-user query performance. By analyzing the personality dimensions that lead to more accurate queries, training programs can be developed to help persons with other levels of that personality dimensions learn how to produce more accurate queries. This research would allow organizations to better match these short-term organizational needs with appropriate personality types. This matching should decrease the learning curve for contract personnel, as they would be better suited to querying, data mining, or other information retrieval tasks that require interaction with information repositories.

The usual caveats associated with laboratory experiments using student participants limit the generalizability of the results. Future research is needed to improve end-users' abilities to extract the information they need.

REFERENCES

1. Axelsen, M., Borthick, A.F., Bowen, P.L. (2001). "A Model for and the Effects of Information Request ambiguity and end user query performance" *Proceedings of the International Conference on Information Systems*, December 2001, 537-542.
2. Barrick, M.R., and Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
3. Barrick, M. R., and Mount, M.K. (1993). "Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance". *Journal of applied Psychology*, 78, 111-118
4. Costa, P.T., Jr. and McCrae R.R. (1992) *Revised NEO Personality Inventory and NEO Five-Factor Inventory Professional Manual,* Psychological Assessment Resources Inc., Florida.
5. Costa, P.T., Jr., McCrae, R.R., and Dye, D.A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences*, 12, 887- 898.
6. Costa, P.T, Jr. and McCrae, R.R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, 64, 21-50.
7. Day, D.V., and Silverman, S.B. (1989) "Personality and job performance: Evidence of Incremental Validity", *Personnel Psychology*, 42, 25-36.
8. George, J.M. (1992). "The Role of Personality in Organisational Life: Issues and Evidence", *Journal of Management*, 18, 2, 185-213.
9. Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26 -34.
10. Gough, H. (1976). Personality and Personality Assessment. In *Handbook of Industrial and Organisational Psychology,* edited by M. D. Dunnette, Chicago, Ill, Rand-McNally.
11. Judge, T.A, Higgins, C.A, Thoresen C.J., Barrick, M.R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, 52, 621-651.
12. Szymura, B. and Wodniecka, Z. (2003). "What really bothers neurotics? In search for factors impairing attentional performance" *Personality and individual differences*, 34, 1, 109-126.
13. Tenopyr, M. (1993). Construct validation needs in vocational behavior theories. Special Issue: The theory of work adjustment. *Journal of vocational Behavior*, 43, 84-89.
14. Walton, D. (1996). *Fallacies arising from ambiguity*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
15. Wand, Y and Weber, R. (1990) An Ontological Model of an Information System, *IEEE Transactions on Software Engineering*, 16, 1282-1292.
16. Zmud, R.W. (1979) Individual Differences and MIS Success: A Review of the Empirical Literature, *Management Science*, 25, 10, 966-979.

Development of a Framework for Trust in Mobile Commerce

Keng Siau
University of Nebraska-Lincoln
ksiau@unl.edu

Hong Sheng
University of Nebraska-Lincoln
hsheng@unlnotes.unl.edu

Fiona Nah
University of Nebraska-Lincoln
fnah@unl.edu

ABSTRACT

Mobile commerce represents a significant development in e-commerce. Despite the potential of mobile commerce, trust is a major obstacle in its adoption and development. The focus of this research is to develop a framework to identify the factors influencing trust in mobile commerce and to explain the development of such trust using a means-ends objective network. We utilized the Value-Focused Thinking approach to interview subjects in order to identify their fundamental and means objectives concerning trust in mobile commerce and to construct a means-ends objective network. A trust framework is developed from the means-ends objective network. As one of the first research on trust in mobile commerce, the framework developed in this study provides valuable information for researchers and practitioners, and serves as a conceptual foundation for future research in mobile commerce.

Keywords

Trust, mobile commerce, value-focused thinking.

INTRODUCTION

Advances in wireless technology have stimulated rapid developments in electronic commerce (e-commerce) via the use of mobile devices. E-commerce transactions conducted through radio-based wireless devices are called mobile commerce (also known as m-commerce or mobile e-commerce). Mobile commerce can extend current Internet sales channels into more immediate and personalized mobile environment. However, one of the most daunting challenges to ensure wide diffusion of mobile commerce concerns trust in mobile commerce. Lack of consumer trust is the most significant long-term barrier for e-commerce (Keen, 1997), as well as for mobile commerce. Although mobile devices are more convenient for "anytime shopping", it has some unique features and characteristics that hinder the development of consumer trust.

To become a viable means of doing business, mobile commerce must overcome the problem of user distrust. An in-depth understanding of the factors that constitute and can bring about consumer trust in mobile commerce is essential.

LITERATURE REVIEW

Trust plays a crucial role in commercial relationships (Nah and Davis, 2002). Trust has been studied in various disciplines ranging from social psychology to decision making. Recently, trust has been studied extensively in the e-commerce context (e.g., Jarvenpaa *et al*., 1999; McKnight *et al.*, 2002a, 2002b; Shankar *et al*., 2002; Lee and Turban, 2001).

Table 1 provides a summary of frameworks and models regarding trust in e-commerce and m-commerce.

BUILDING A MEANS-ENDS OBJECTIVE NETWORK USING VALUE-FOCUSED THINKING APPROACH

To understand the values leading to trust in mobile commerce, we used the Keeney's (1992) Value-Focused Thinking (VFT) approach to help identify these values. Value refers to the principles for evaluating the desirability of possible alternatives or consequences. Values that are of concern are made explicit by the identification of objectives (Keeney, 1992), where an objective is defined as a statement of something that one desires to achieve. There are two types of objectives: fundamental objectives and means objectives. Fundamental objectives are concerned with the ends that decision makers value in a specific decision context whereas means objectives are methods to achieve the ends.

The process of Value-Focused Thinking involves four steps (Keeney, 1992), as presented in Figure 1.

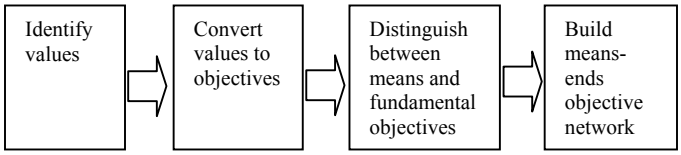


Figure 1: Procedures of Value-Focused Thinking

After interviewing subjects using the value-focused thinking approach, we derived the means and fundamental objectives, and then organized them into a means-ends objective network (see Figure 2).

Literature	Findings
Ambrose and Johnson (1998)	• Seller characteristics (ability, benevolence, integrity) • Buyer's propensity to trust
Ba, Whinston and Zhang (1999)	• Information asymmetry influences consumers' knowledge of product quality • Trusted third party (i.e., certification authorities) help to build trust
Belanger, Hiller and Smith (2002)	• Third party privacy seals • Privacy statement • Third party security seals • Security features
Cheung and Lee (2000)	• Trustworthiness of Internet vendor (perceived security control, perceived privacy control, perceived competence, perceived integrity) • External environment (third party recognition, legal framework) • Propensity to trust
Friedman, Kahn and Howe (2000)	• Reliability and security of technology • Anonymity of transaction information • Performance history and reputation of website
Fung and Lee (1999)	• Company reputation • Web-site interface • Information quality
Gefen (2000)	• Familiarity (with e-commerce vendor) • Disposition to trust
Gefen, Karahanna and Straub (2003)	• Calculative-based • Institution-based structural assurance • Institution-based situational normality • Knowledge-based familiarity • Perceived ease of use
Jarvenpaa *et al*. (1999)	• Perceived size • Perceived reputation
Kim and Prabhakar (2000)	• Trustor's propensity-to-trust • Word-of-mouth referrals • Institutional characteristics
Lee and Turban (2001)	• Trustworthiness of Internet merchant (ability, integrity, benevolence) • Trustworthiness of Internet shopping medium (technical competence, reliability, medium understanding) • Contextual factors (effectiveness of third party certification, effectiveness of security infrastructure) • Individual trust propensity • Other factors
McKnight, Choudhury and Kacmar (2002a)	• Perceived vendor reputation • Perceived site quality • Structural assurance of the web
McKnight, Choudhury and Kacmar (2002b)	• Disposition to trust (faith in humanity, trusting stance) • Institution-based trust (situational normality: general, competence, integrity and benevolence, structural assurance) • Trusting beliefs (competence beliefs, benevolence beliefs, and integrity beliefs) • Trusting intentions (willingness to depend, subjective probability of depending)
Nah and Davis (2002)	• Content of website • Design of website • External certifications and references
Pavlou and Ba (2000)	• Seller's reputation • Appropriate feedback mechanisms
Shankar, Urban and Sultan (2002)	• Website characteristics (e.g., navigation and user friendliness, advice, error free) • User characteristics (e.g., Internet savvy, past Internet shopping behavior, feeling of control) • Other characteristics (e.g., online medium, trustworthiness of firm, perceived size of firm)
Siau and Shen (2003)	• Trust in mobile commerce can be differentiated into two categories: trust in mobile technology and trust in mobile vendors.

Table 1: Summary of Trust Frameworks and Models

CONSTRUCTING THE FRAMEWORK ON TRUST IN MOBILE COMMERCE

The objectives we have obtained from interviewing mobile commerce users provide a comprehensive list of antecedents of trust in mobile commerce. In addition, the links between objectives depicted in the means-ends objective network suggest the causal relationships between the means and fundamental objectives. We classified the various objectives in the means-ends objective network into categories, and proposed a conceptual framework that outlines the variables influencing trust building in mobile commerce. This framework is shown in Figure 3.

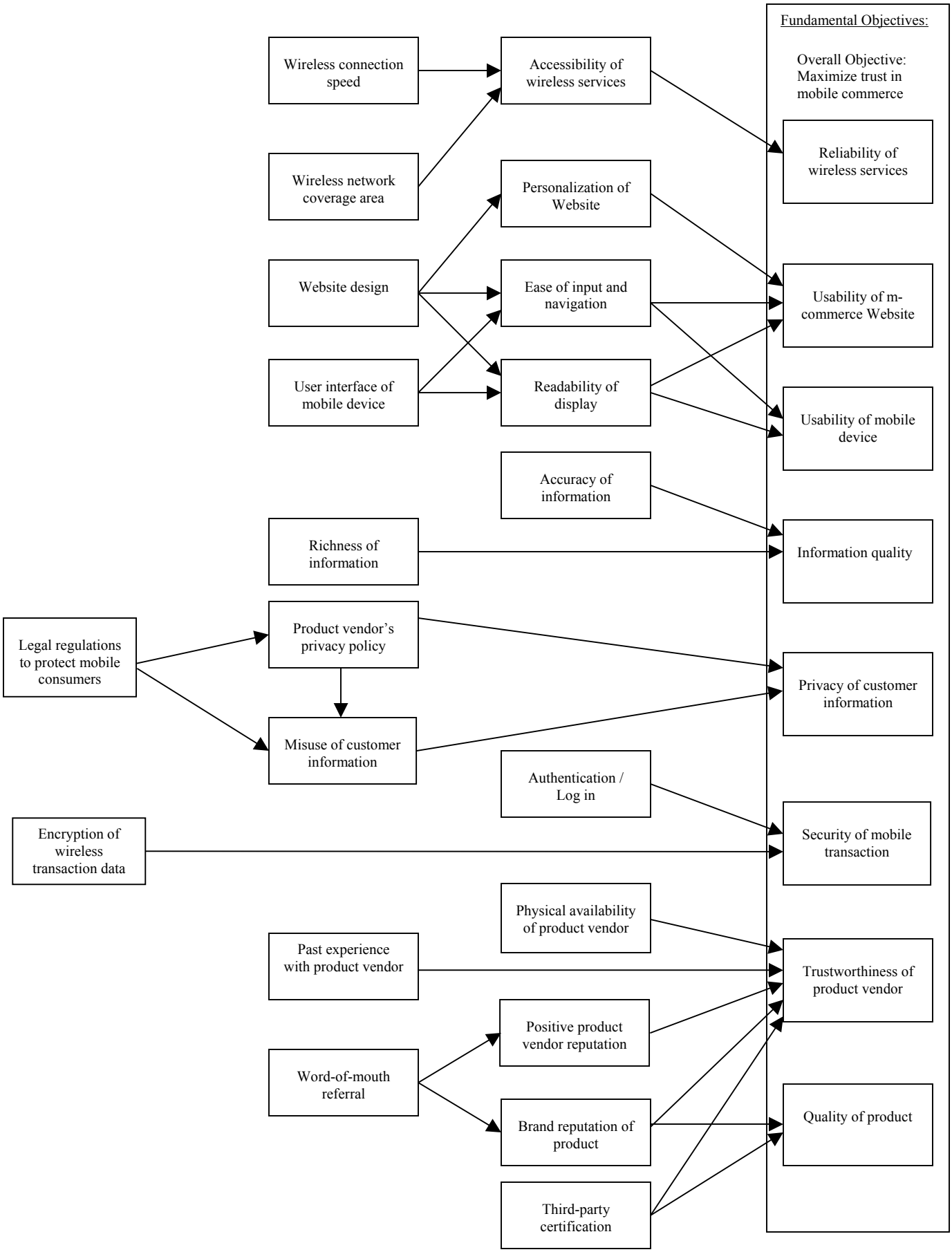


Figure 2: Means-Ends Objective Network

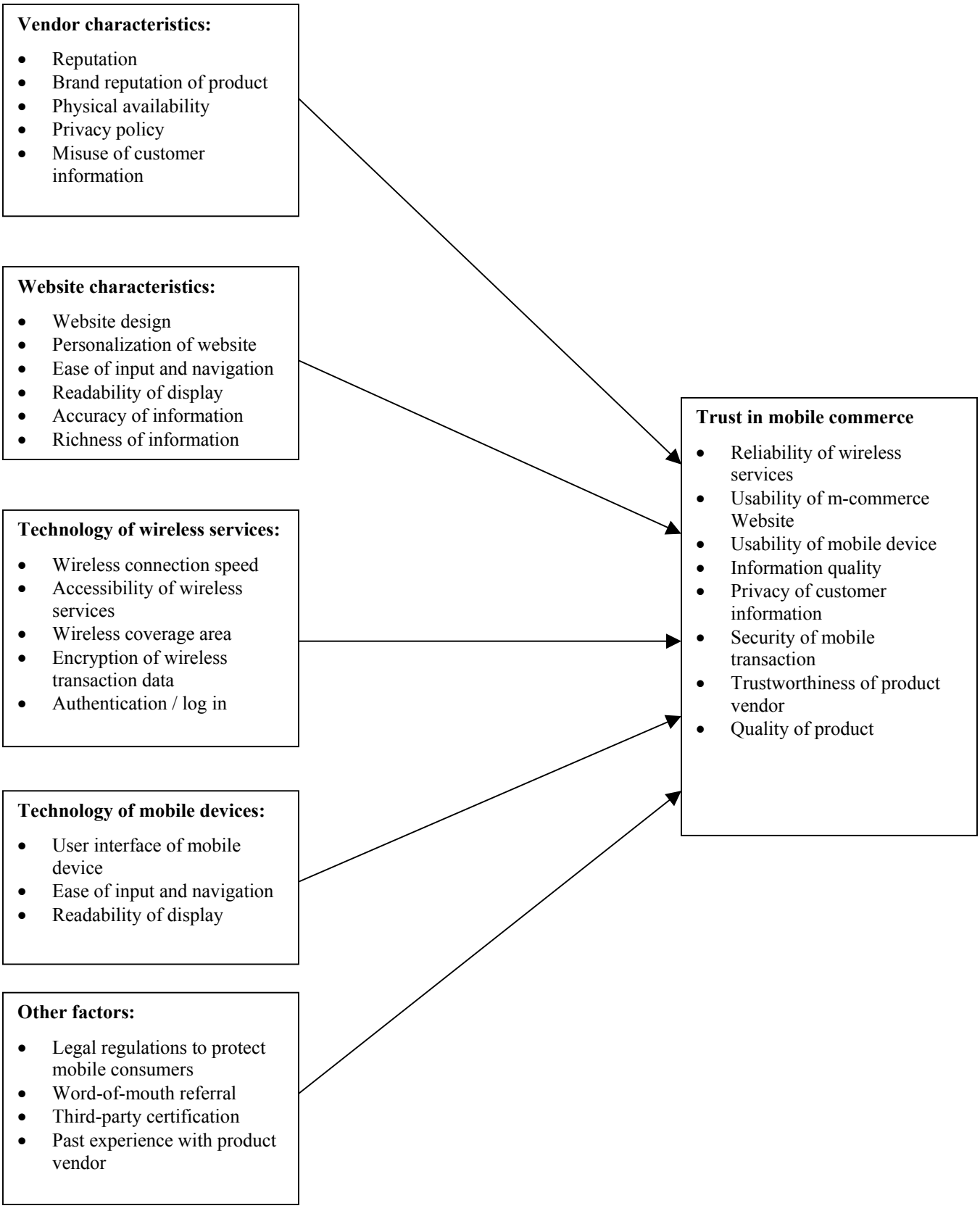


Figure 3: Proposed Framework for Trust in Mobile Commerce

DISCUSSION AND CONCLUSION

This study uses the Value-Focused Thinking approach to identify factors influencing trust in mobile commerce and presents a means-ends objective network to depict the relationships among those factors. Based on the means-ends objective network, a framework for trust in mobile commerce is developed.

This framework validates antecedents of trust in e-commerce in a somewhat different context – the mobile commerce context. Although some of the trust factors identified in this study have been presented in the e-commerce literature, our framework identifies new antecedents that are unique to trust in mobile commerce. For example, technology related factors are considered particularly important in mobile commerce due to the immaturity of mobile technology and the unique user-interface of mobile devices. As suggested by the subjects we interviewed, technology is a main barrier of trust in the conduct of mobile commerce. In our framework, there are three categories of technology related factors: technology relating to wireless services, wireless websites, and mobile devices. Some of the antecedents of trust in mobile commerce arise because of the unique interface and the limited features and functions of mobile devices. Other factors highlighted during the interviews include security features in the conduct of mobile commerce.

In conclusion, we believe that mobile commerce has tremendous potential. However, to achieve this potential, the trust issue needs to be more fully understood and directly addressed by vendors and providers of mobile commerce technologies and services. The framework of trust in mobile commerce developed in this research is an important step in this direction.

REFERENCES

1. Ambrose, P.J., and G.J. Johnson (1998). A Trust Model of Buying Behavior in Electronic Retailing, *Proceedings of Americas Conference on Information Systems*, pp. 263-265
2. Ba, S., A.B. Whinston, and H. Zhang (1999). Building Trust in Electronic Market through an Economic Incentive Mechanism, *Proceedings of the Twentieth International Conference on Information Systems*, pp. 208-213
3. Belanger, F., J.S. Hiller, W.J. Smith (2002). Trustworthiness in Electronic Commerce: the role of privacy, security, and site attributes, *Journal of Strategic Information Systems*, 11, pp. 245-270
4. Cheung, C. and M. Lee (2000). Trust in Internet Shopping: a proposed model and measurement instrument, *Proceedings of the Americas Conference on Information Systems*, pp. 681-689
5. Friedman, B., P. Kahn, and D. Howe (2000). Trust Online, *Communications of the ACM*, 43(12), pp. 34-40
6. Fung, R., and M. Lee (1999). EC-Trust (Trust in Electronic Commerce): exploring the antecedent factors, *Proceedings of America Conference of Information Systems*, pp. 517-519
7. Gefen, D. (2000). E-commerce: the role of familiarity and trust, *Omega*, 28(6), pp. 725-737
8. Gefen, D., E. Karahanna, D.W. Straub (2003). Trust and TAM in Online Shopping: an integrated model, *MIS Quarterly*, 27(1), pp. 51-90
9. Jarvenpaa, S.L., N. Tractinsky, L. Saarinen, and M. Vitale (1999). Consumer Trust in an Internet Store: a cross-cultural validation. *Journal of Computer-Mediated Communication*, 5(2), pp. 44-71
10. Keen, P.G.W. (1997). Are you ready for "Trust" Economy, *Computerworld*, 31(16), p. 80
11. Keeney, R.L. (1992). Value-focused Thinking, Cambridge, Massachusetts: Harvard University Press
12. Kim, K. and B. Prabhaker (2000). Initial Trust, Perceived Risk, and the Adoption of Internet Banking, *Proceedings of the Twenty First International Conference on Information Systems*, pp. 537-543
13. Lee, M.K.O. and E. Turban (2001). A Trust Model for Consumer Internet Shopping. *International Journal of Electronic Commerce*, 6(1), pp. 75-91
14. McKnight, D.H., V. Choudhury, and C. Kacmar (2002a). The Impact of Initial Consumer Trust on Intentions to Transact with a Web Site: a trust building model, *Journal of Strategic information Systems*, 11, pp. 297-323
15. McKnignt, D.H, V. Choudhury, and C. Kacmar (2002b). Developing and Validating Trust Measures for E-commerce: an integrative typology, *Information Systems Research*, 13(3), pp. 334-359
16. Nah, F. and S. Davis (2002). HCI Research Issues in Electronic Commerce. *Journal of Electronic Commerce Research*, 3(3), pp. 98-113
17. Pavlou, P.A. and S. Ba (2000). Does Online Reputation Matter? An empirical investigation of reputation and trust in online auction markets, *Proceedings of the Americas Conference on Information Systems*, pp. 948-950
18. Shankar, V., G.L, Urban, F. Sultan (2002) Online Trust: a stakeholder perspective, concepts, implications, and future directions, *Journal of Strategic Information Systems*, 11, pp. 325-344
19. Siau, K. and Shen, Z. (2003). Building Consumer Trust in Mobile Commerce, *Communications of the ACM*, 46(4), pp. 91-93

A Study of Task Characteristics and User Intention to Use Handheld Devices for Mobile Commerce

Xiaowen Fang
DePaul University
xfang@cti.depaul.edu

Susy Chan
DePaul University
schan@cti.depaul.edu

Jacek Brzezinski
DePaul University
jbrzezinski@cti.depaul.edu

Shuang Xu
DePaul University
shuangxu@yahoo.com

ABSTRACT

Interface design and the selection of appropriate tasks for small-screen mobile applications are issues critical for mobile commerce. Our earlier research has identified five major task factors that may influence user intention to use handheld devices for wireless applications. These factors are: perceived ease of use, perceived usefulness, perceived playfulness, perceived task complexity, and perceived security. We followed up with a questionnaire-based empirical study to validate the relative impact of these proposed task factors on user intention to use handheld devices for mobile commerce. This paper confirms significant correlations between the task factors and user intention. However, only three of the five factors -- perceived usefulness, perceived security, and perceived playfulness -- are important to user intention, explaining 30% of the variations in a multiple regression model. This study makes a unique contribution to HCI and MIS research by providing empirical evidence of user perception of task characteristics for mobile commerce.

Keywords

Task characteristics, mobile commerce, wireless handheld devices, usability study.

INTRODUCTION

The convergence of mobile Internet and wireless communication technology has promised users anytime, anywhere access of information for work and personal communication. Such opportunities include mobile services that support m-commerce transactions and process facilitations for managing personal activities, mobile office, and mobile operations. However, small screen display, limited bandwidth, and the multiple functionality of handheld devices hinder such access.

Researchers suggest that interface developers need to consider the interaction among the interface design of user tasks, form factors, and application objectives (Chan, Fang, Brzezinski, Zhou, Xu and Lam, 2002). M-commerce assumes that users primarily access the Internet or wireless applications either on the move, or while stationary but away from office or home. Since mobile users can spare only limited time and cognitive resources for performing a task, the choice of mobile applications is important. Anckar and D'Incau (2002) suggest that services emphasizing mobile values, meeting time-critical and spontaneous needs are more suitable for wireless devices. In designing tasks for m-commerce applications, it is essential to determine which tasks are suitable for wireless applications (Chan and Fang, 2003).

The objective of this paper is two-fold. First, it intends to validate the relationship between five task factors and user intention to use handheld devices for wireless applications. The five factors are: perceived ease of use, perceived usefulness, perceived playfulness, perceived task complexity, and perceived security (Xu, Fang, Chan and Brzezinski, 2003). Second, the study intends to investigate the causal relationships between the five proposed task factors and user intention to perform tasks on wireless handheld devices. This study makes a unique contribution to usability studies for small-screen mobile devices and mobile commerce by empirically examining task factors and their relative importance on user intention in performing mobile tasks.

BACKGROUND LITERATURE

Usability research in mobile commerce is a new area. Based on a study of 19 novice wireless phone users who were closely tracked for the first 6 weeks after service acquisition, Palen and Salzman (2002) describe the wireless telephony system as having four socio-technical components: hardware, software, "netware," and "bizware." They indicate that each of these four components must have a user-friendly design. Their research recommends a systems-level usability approach. Perry, O'Hare, Sellen, Brown and Harper (2001) present a study of mobile workers that highlights remote "anywhere, anytime" access of information and individuals. They identify four key factors in mobile work: the role of planning, working in "dead time," accessing remote technology and informational resources, and monitoring the activities of remote colleagues. In a study aiming to understand how mobile web access affects home Internet usage, McClard and Somers (2000) have investigated the household integration of tablet computers and the defined user requirements for similar devices. They suggest that an Internet appliance intended for general web access and text-based communication must have three characteristics: 1) Software must contain features that are perceived as useful; 2) The device must

be highly portable and comfortable to use; 3) The screen and keyboard of the device must be large enough to be usable. McClard and Somers also identified the top three preferred tablet features: 1) surfing the Web, 2) Internet availability anywhere in the home, and 3) email. These studies have yielded some general guidelines regarding appropriate tasks for wireless applications.

User adoption of technology applications focuses on tasks and its fit with selected technologies. An essential goal of m-commerce is to identify mobile values for individual users. Anckar and D'Incau (2002) present a framework that differentiates between the values offered by wireless Internet technology (wireless values) and the values arising from the mobile use of the technology (mobile values). Convenience, cost savings, and cell phones best represent wireless values. Services that deliver strong mobile values would make m-commerce a dominant channel. These services meet the following five types of user needs: 1) time-critical needs and arrangements; 2) spontaneous needs and decisions, such as auctions, email, and news; 3) entertainment needs; 4) efficiency needs and ambitions; and 5) mobility related needs. A consumer survey based on this framework reveals that email, routine bank services, and booking theatre tickets are among the most desired mobile services (Anckar and D'Incau, 2002). Desired mobile services also include restaurant reservations, calendaring and alert services, and access to news sources. Fewer than 30% of the respondents were interested in services involving transactions, such as online purchasing. Since this survey did not require participants to have experience with mobile services, discrepancies might exist between user's expectations and responses based on task experience.

These research findings suggest that an understanding of user preferences and perceived values of mobile tasks is essential for improving the usability of mobile tasks. Towards this goal, we have conducted an exploratory study that involved interviews of 37 participants upon completion of an observation experiment (Xu et al., 2003). Stemming from these interviews are five factors that would affect the preference of tasks to be performed on wireless handheld devices. These factors are perceived ease of use, perceived usefulness, perceived playfulness, task complexity, and perceived security.

HYPOTHESES

In our current research, we undertook a questionnaire-based empirical study to validate the five hypotheses proposed earlier (Xu et al., 2003):

- Hypothesis 1: Higher perceived ease of use of a task performed on wireless devices will result in higher user intention to perform the task.
- Hypothesis 2: Higher perceived usefulness of a task performed on wireless devices will result in higher user intention to perform the task.
- Hypothesis 3: Higher perceived playfulness of a task performed on wireless devices will result in higher user intention to perform the task.
- Hypothesis 4: Higher perceived complexity of a task performed on wireless devices will result in lower user intention to perform the task.
- Hypothesis 5: Higher perceived security of a task performed on wireless devices will result in higher user intention to perform the task.

METHODS

Participants

The research team recruited 101 participants for the experiment. The majority of these participants were working adults. Some were alumni of a Midwest university while others were still enrolled at the time of their participation. The participants represented a very diverse group. Their ages ranged from about 20 to 50 and there was a wide representation of different cultures and races. Their experiences with handheld devices also varied. Of the participants: 97% had used wireless phones before and 75.2% used wireless phones on a daily basis; 22.8% had used Pocket PCs before and 8.9% used the device on a daily basis. For Palm Pilot, 19.8% of participants used Palm Pilot daily, 13.9% weekly, 4% monthly, 26.7% rarely, and 35.6% had never used this type of device previously, so the participants included both novice and experienced users of Palm Pilot.

Tasks

Participants were asked to evaluate task characteristics of each of the twelve tasks in five aspects: perceived ease of use, perceived usefulness, perceived playfulness, task complexity, and perceived security. At the end of the experiment, participants rated their intention to perform each task in the future using a wireless handheld device. To ensure that the resulting model was generalizable, task selection considered two requirements: 1) the tasks must be realistic and cover a wide range of possible mobile applications on handheld devices and 2) the tasks must have diverse characteristics. Accordingly, twelve tasks were selected for this study: 1) managing an address book, 2) sending/receiving email, 3) checking flight information, 4) purchasing movie tickets, 5) reading news, 6) banking online, 7) playing games, 8) checking weather information, 9) purchasing books, 10) purchasing clothes, 11) sending short messages, and 12) trading stocks. These tasks were identified by users as appropriate for mobile commerce tasks based on two prior research studies (Anchar and D'Incau, 2002; Xu et al.,

2003). As a set, these twelve tasks covered a wide range of mobile applications and were real tasks performed on wireless handheld devices.

Independent and dependent variables

There were five independent variables describing task characteristics: perceived ease of use (Davis, 1989; Venkatesh & Davis, 1996), perceived usefulness (Davis, 1989; Venkatesh & Davis, 1996), perceived playfulness Venkatesh, 1999, 2000), task complexity (Campbell, 1988), and perceived security (Xu et al., 2003). In order to establish the validity of the measurements, most of the definitions and questions of these independent variables were derived from well-established studies as referenced.

The dependent variable was the intention to use a wireless handheld device for performing the selected tasks. It was measured by one question in the final questionnaire: "Assuming that you have access to a wireless handheld device, assign a score of intention to use to each task to indicate to what extent you intend to use the handheld to perform this task." A score of zero meant no intention to use the handheld devices while seven implied the highest level of intention.

Procedure

Each participant was given one experiment packet (48 pages in length) containing all task scripts and questionnaires by email, mail, or in person. The experiment packet was organized in the following order: 1) A brief instruction to the study; 2) Pre-experiment questionnaire; 3) Twelve sets of task scripts and task characteristics questionnaire in a randomized order; 4) Final questionnaire about user intention. For each of the twelve tasks, an identical task characteristics questionnaire was provided following the task scripts. The participant was asked to evaluate all task scripts and complete all questionnaires page by page. Upon completing all the task scripts, the participant was asked to rate his/her intention to perform each of the twelve tasks on handheld devices in the final questionnaire. The participant was instructed that it was not necessary to complete the entire survey in one sitting but they should complete an individual task and the corresponding characteristics questionnaire in one sitting. This instruction was intended to enable participants to focus on each task.

RESULTS AND DISCUSSION

Internal Consistency of the Instrument

The high Crombach's α values for perceived ease of use (3 items, α=.085), perceived usefulness (3 items, α=.90), perceived playfulness (2 items, α=.071), and task complexity (2 items, α=.075) imply that the measurements of these variables are reliable and valid. However, a Crombach's α value of 0.19 for perceived security (2 items) is unacceptable and suggests that the statements to measure this variable are unreliable. For the subsequent analysis, we only included data for one item "I feel secure to perform this task on the handheld computer" to measure the perceived security of a task.

Correlation Analyses

Hypotheses 1 to 5 postulate causal relationships between each of the five factors (perceived ease of use, perceived usefulness, perceived playfulness, task complexity, perceived security) and user intention to perform a task on wireless handheld devices. Correlation analyses were performed to test these five hypotheses. Table 1 presents the correlation matrix.

The perceived ease of use has a significant and positive correlation ($r = 0.35448$, $p < 0.0001$) with user intention to perform a task on wireless handheld devices. This correlation supports the first hypothesis stating that higher perceived ease of use leads to higher user intention.

The significant and positive correlation ($r = 0.44705$, $p < 0.0001$) between the perceived usefulness and user intention suggests that higher perceived usefulness results in higher user intention. Hypothesis 2 is supported. These results are in agreement with the prior research findings. The technology acceptance model (TAM) proposed by Davis (1989) is one of the most widely used models of IT adoption. According to TAM, an individual's IT adoption is influenced by the perceived usefulness and perceived ease of use. The perceived ease of use influences the user intention indirectly through the perceived usefulness. These two perceptions help in shaping the user's attitude towards usage and intention to use.

Perceived playfulness has a significant and positive correlation ($r = 0.38177$, $p < 0.0001$) with user intention to perform a task on wireless handheld devices. This correlation implies that higher perceived playfulness results in higher user intention and thus supports hypothesis 3. Among the five types of user needs for wireless services identified by Anckar and D'Incau (2002), one is entertainment need. Prior research on the effectiveness of game-based training has also shown that manipulating the level of perceived enjoyment has a significant impact on user behavior in technology adoption (Venkatesh, 1999, 000). Because applications on wireless handheld devices can be accessed anytime and anywhere, users may use these applications to entertain themselves in their free time. Therefore, the perceived playfulness of applications on handheld devices becomes more important than that of ordinary applications.

	Perceived Ease of Use	Perceived Usefulness	Perceived Playfulness	Task Complexity	Perceived Security	User Intention
Perceived Ease of Use	1.00000	0.56436*	0.68731*	-0.87140*	0.37441*	0.35448*
Perceived Usefulness	0.56436*	1.00000	0.57551*	-0.48093*	0.33643*	0.44705*
Perceived Playfulness	0.68731*	0.57551*	1.00000	-0.57538*	0.33729*	0.38177*
Task Complexity	-0.87140*	-0.48093*	-0.57538*	1.00000	-0.32748*	-0.29761*
Perceived Security	0.37441*	0.33643	0.33729*	-0.32748*	1.00000	0.42549*
User Intention	0.35448*	0.44705*	0.38177*	-0.29761*	0.42549*	1.00000

$p<0.0001$

Table 1. Correlation Coefficients

The task complexity has a significant and negative correlation ($r = -0.29761$, $p < 0.0001$) with user intention. This finding favors hypothesis 4 stating that higher task complexity leads to lower user intention. Because handheld devices are used on the move in many cases, tasks performed on these devices would become either secondary tasks or primary tasks with other secondary tasks. In either scenario, different tasks will compete for human cognitive resources. Because of the user's mobility, the interaction between the user and the mobile device is usually very brief. This implies that users will have much less capability of handling complex tasks on handheld devices under mobile conditions than under stationary conditions.

The significant and positive correlation ($r = 0.42549$, $p < 0.0001$) between the perceived security and user intention indicates that higher perceived security results in higher user intention, as postulated in hypothesis 5. Security has been a major concern in e-commerce. It becomes a more sensitive issue in mobile applications due to doubts about the security and the inconsistent reliability of wireless connections. It is not surprising that the perceived security affected user intention to perform a task.

In addition, it is noted that the perceived ease of use is highly correlated with task complexity ($r = -0.87140$, $p < 0.0001$). This correlation suggests that the perceived ease of use and task complexity (as defined in this study) may be measuring the same construct from two directions. A possible explanation might be that participants of this study could not distinguish ease of use (i.e. the nature of the interface) from task complexity (i.e. the nature of the task). This finding implies that user perception of task complexity could not be separated from the perception of the interface design. Therefore, it may be better to combine them into one variable: perceived ease of use. The perceived ease of use may be sufficient to cover the aspect of task complexity. However, task complexity is deemed an important determinant of task characteristics for e-commerce applications (Wells, Sarker, Urbaczewski and Sarker, 2002). As defined by Wood (1986), task complexity is determined by the number of distinct acts and the dynamic changes in the relationships between task inputs and task products. This definition may help users to better differentiate ease of use from task complexity.

Multiple Regression Analysis

We performed a stepwise regression analysis to further investigate the relative impacts of five factors on user intention. The regression results indicate that three of the five variables (perceived usefulness, perceived security, and perceived playfulness) entered the regression model at the 0.05 level and the r-square of the model is 0.2951. In other words, this model can explain only 30% of the total variance of user intention. It has limited power for predicting user intention to perform a task on handheld devices. There are two possible reasons for this limited regression model:

- User intention may be much more complicated than it was expected. Additional factors may influence user intention to use handheld devices, such as user's prior experience with wireless handheld devices and with the mobile tasks.
- The regression analysis has been affected by the correlations among the independent variables.

In either case, further studies are needed to explore a more feasible and accurate regression model. Nevertheless, the regression model does strongly suggest the relative importance of the proposed factors. Perceived usefulness is the first variable that entered the regression model and thus the most important factor impacting user intention. It accounts for about 20% of the variance ($p<.0001$). This means that users heavily rely on their perception of the usefulness of a task when deciding their intention to use a wireless device. Perceived security is the second most important factor. It accounts for about 8.5% of the total variance ($p<.0001$). Perceived

playfulness is the third most important factor. It accounts for about 0.99% of the total variance ($p<.0001$).

The above findings suggest that perceived usefulness, perceived security, and perceived playfulness outweigh perceived ease of use as a determinant of user intention to perform a task on handheld devices. Several prior studies reported similar findings about the perceived ease of use for IS adoption in a work environment (Legris, Ingham and Collerette, 2002) and in an e-commerce context (Koufaris, 2002). As usability is essential to mobile commerce (Chan and Fang, 2003), further research is needed to better understand the dimensions of ease of use to aid interface design for small-screen applications.

CONCLUSIONS

In this study, we conducted a questionnaire-based empirical study to investigate the factors affecting user intention to use wireless handheld devices. The main contribution of this study was the finding of the relative importance of factors affecting user intention to use handheld devices. Key findings from our preliminary results show:

- Perceived usefulness, perceived security, perceived playfulness, and perceived ease of use correlated positively with user intention.
- Perceived usefulness, perceived security, and perceived playfulness were the three most important factors affecting user intention. Together they accounted for 30% of the total variance in user intention. This finding implies that mobile commerce applications should be mapped closely to user needs. Interface design and task selection should consider security and motivational factors.
- The regression model resulted from this study could only predict user intention with 30% accuracy. Further study is needed to identify additional factors that contribute to a more explanatory model.
- The relationship between ease of use and task complexity needs further clarification in order to aid interface design for complex tasks.

REFERENCES

1. Anckar, B. and D'Incau, D. (2002). Value creation in mobile commerce: Findings from a consumer survey, Journal of Information Technology Theory & Application, 4, 1, 43-64.
2. Campbell, D. J. (1988). Task complexity: a review and analysis, *Academy of Management Review*, 13 1, 40-52.
3. Chan, S., Fang, X., Brzezinski, J., Zhou, Y., Xu, S. and Lam, J. (2002). Usability for mobile commerce across multiple form factors, *Journal of Electronic Commerce Research*, 3, 2, 187-199.
4. Chan, S. and Fang, X. (2003). Mobile commerce and usability. In K. Siau & E. Lim, (Eds.), Advances in Mobile Commerce Technologies, Idea Group Publishing, 235-257.
5. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 13, 3, 319-339.
6. Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior, *Information Systems Research*, 13, 2, 205-223.
7. Legris, P., Ingham, J. and Collerette, P. (2002). Why do people use information technology? A critical review of the technology acceptance model, *Information & Management*, 40, 191-204.
8. McClard, A. and Somers, P. (2000). Unleashed: Web tablet integration into the home, *Proceedings of CHI*, 1-6.
9. Palen, L. and Salzman, M. (2002). Beyond the handset: designing for wireless communications usability, *ACM Transactions on Computer-Human Interaction*, 9(2), 125-151
10. Perry, M., O'hara, K., Sellen, A., Brown, B. and Harper, R. (2001). Dealing with mobility: understanding access anytime, anywhere, *ACM Transactions on Computer-Human Interaction*, 8, 4, 323-347.
11. Venkatesh, V. (1999). Creation of favorable user perceptions: exploring the role of intrinsic motivation, *MIS Quarterly*, 23, 2, 239-260.
12. Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model, *Information Systems Research*, 11, 4, 342-365.
13. Venkatesh, V. and Davis, F. (1996). A model of the antecedents of perceived ease of use: Development and test, *Decision Science*, 27, 3, 451-481.
14. Wells, J., Sarker, S., Urbaczewski, A., & Sarker, S. (2002). Studying customer evolutions of electronic commerce applications: A review and adaptation of the task-technology fit perspective. In the Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03).
15. Wood, R. (1986). Task complexity: Definition of the construct. Organizational Behavior and Human Decision Processes, 37(1), 60-8.
16. Xu, S., Fang, X., Chan, S. and Brzezinski, J. (2003). What tasks are suitable for handheld devices? *Proceedings of HCI International 2003, Human Computer Interaction: Theory and Practice,* 2, 333-337.

Post-Adoption Behavior of Mobile Internet Users:

A Model-Based Comparison between Continuers and Discontinuers

Hoyoung Kim
Widerthan.com
hykim@widerthan.com

Jinwoo Kim
Yonsei University
jinwoo@yonsei.ac.kr

ABSTRACT

Many mobile Internet users are not continuing to use mobile Internet services after initial use. This study aims to explore how such users (discontinuers) differ from ongoing users (continuers) in terms of accepting mobile Internet technology. We propose an adoption model for the mobile Internet consisting of seven critical factors. An on-line survey was conducted on the basis of this model to compare continuers and discontinuers. The survey results show that discontinuers are more sensitive to usefulness and social influences in using mobile Internet services, while continuers are more sensitive to ubiquitous connectivity.

Keywords

Mobile Internet, Technology Adoption Model

INTRODUCTION

"Mobile Internet" is defined as wireless access to the Internet through a mobile communication network (e.g., GSM or CDMA) by means of hand-held devices (e.g., mobile phones) (Federal Trade Commission, 2002). Many forecasters, basing their predictions on the uptake of standard mobile Internet phones, suggest that in the near future most Internet access will take place using small, wireless devices that, equipped with browser and wireless connection, provide "anywhere and anytime" access (Buyukkoken et al., 2000).

Despite this optimism, however, mobile Internet services face serious problems in terms of low profits and shallow customer bases (Businessweek, 2002). In order to increase the number of mobile Internet users, and hence profits, we need to understand how those who have stopped using mobile Internet service (discontinuers) are different from those who keep using it (continuers). It has been found that continuers behave differently from discontinuers in terms of accepting new technologies (Parthasarathy and Bhattrcherjee, 1998). Effective strategies for maintaining continuers will, therefore, be different from those for converting discontinuers into continuers (Bruner, 1998). In order to implement different strategies for discontinuers and continuers, we need to identify the differences in post-adoption behavior between the two groups.

Our objective in this study was to identify and compare the critical factors that affect the post-adoption behavior of continuers and discontinuers. In other words, which factors are effective in increasing the intention among discontinuers to use the mobile Internet, and how these factors differ from those that influence continuers?

THEORETICAL BACKGROUND

Prior studies that specifically target the post-adoption behavior of mobile Internet users are scarce. For this reason, we based our mobile Internet adoption model on prior studies of technology-adoption in general in three areas: marketing (e.g., Zeithaml, 1998), technology innovation (e.g., Rogers, 1995), and information systems (Davis, 1989). Our research strategy was first to construct this model, and then to use it to compare the post-adoption behavior of continuers and discontinuers. In order to address our research questions, we have added three more aspects to the results of prior studies. First, among information technologies, mobile Internet services have certain unique characteristics, the most prominent being ubiquitous connectivity: users can use the services regardless of time or place (Dey, 2001). Second, the mobile Internet has been used mostly by individual consumers, rather than by corporate users or organizations (HCI Lab, 2002). Therefore, we have based our adoption model on consumer behavior studies (Dodds and Monroe, 1991). Third, this study investigates empirically the factors critical to post-adoption behavior, comparing continuers and discontinuers by means of structural equation models.

THEORY AND RESEARCH HYPOTHESES

Our model for the adoption of mobile Internet services consists of seven important factors that affect post-adoption behavior (Figure 1).

Continuers and Discontinuers

In general, critical factors that affect the adoption of products or services can differ according to user type (Venkatesh, 2000). In other words, people using a product or service consider different factors to be important, depending on what type of user they are.

Our study classified users of mobile Internet services into two groups: continuers, who keep using mobile Internet

services after their initial adoption, and discontinuers, who have adopted mobile Internet services and later ceased to use them. The two groups may be different in terms of the impact of post-adoption factors on perceived value and behavioral intention.

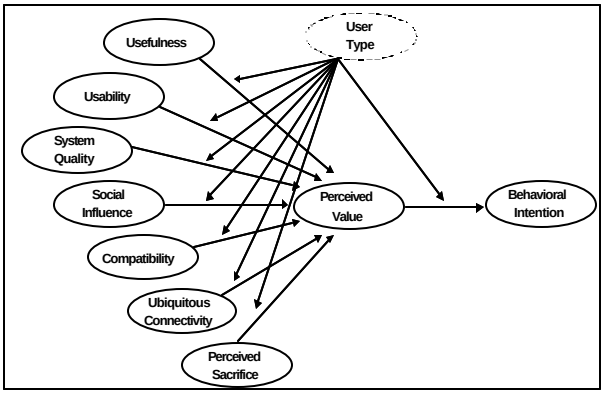


Figure1. Adoption model for Mobile Internet service

Usefulness

Usefulness, a subjective measure, can be defined as how much a user feels a new product is helping in his or her work (Rogers, 1995). When the usefulness of a new product is high, the product is adopted rapidly in the market. Parthasarathy and Bhattrcherjee (1998) suggest that people are more willing to pay for useful on-line services than for less useful one. Karahanna et al. (1999) established that good quality and good function perceived by people allow them to adopt a new system easily in an organizational environment.

We hypothesize that continuers may be less sensitive to the usefulness of a technology, and that usefulness is thus more influential to discontinuers than to continuers. In other words, continuers, who usually are early adopters, keep finding value in the mobile Internet even though it is not as useful in their work as they expected (Roger, 1995). In contrast, we hypothesize that discontinuers are more sensitive to usefulness, and that the perceived value of the mobile Internet drops significantly, for these users, if they find it is not as useful as they expected. Thus we propose as our first hypothesis :

H1: Usefulness influences Perceived Value more heavily for discontinuers than for continuers.

Usability

The perceived value of mobile Internet services will depend in part on their usability. A system's usability is how easy it is to learn and use (Preece et al., 1994) and how comfortable it is to use (Davis 1989). Rogers (1995) suggests that a system is adopted quickly when a user can learn how to use the system easily. Nielsen (1993) proposes that for better usability a system must be efficient to use. Ajzen (1991) finds a high correlation between difficulty of use and choice of services.

We hypothesize that discontinuers are more sensitive to a technology's ease of use, and that usability will therefore have more influence on discontinuers than on continuers. In other words, discontinuers may be quick to stop using the mobile Internet when they experience difficulties. Karahanna (1999) also finds that non-users are affected more sensitively by usability. In contrast, early adopters and continuers keep finding value in the mobile Internet even though they experience difficulties in using it. It is common to see continuers proudly explaining how they overcome the initial difficulty of using a new technology. Thus we propose as our second hypothesis:

H2: Usability influences Perceived Value more heavily for discontinuers than for continuers.

System Quality

System Quality refers to the perceived stability and efficiency of a system (DeLone and McLean, 1992). Speed and stability have been found to be just as important in determining the overall satisfaction of mobile Internet users (HCI Lab, 2002). Thus, the system quality of mobile Internet services may significantly affect both adoption and ongoing usage.

We hypothesize that discontinuers are more sensitive to the overall quality of a technology, and that system quality will be more influential on discontinuers than on continuers. In other words, early adopters and continuers keep finding the value of the mobile Internet even when the system quality is poorer than they expected. In contrast, discontinuers may be quick to stop using the mobile Internet when they experience system malfunctions. Thus we propose as our third hypothesis:

H3: System Quality influences Perceived Value more heavily for discontinuers than for continuers.

Social Influence

Social Influence refers to the degree of interaction among people in their social context (Rice et al., 1990). Social influence helps determine whether technologies are adopted and whether products are purchased (Venkatesh ,1996).

Social influence may also affect the use of mobile Internet services. In fact, since mobile Internet services are a part of the telecommunications industry—an industry specifically designed to facilitate social interactions—social influence may be an even more important factor in service choice than it would otherwise be (Downes and Mui, 1998).

We hypothesize that social influence will be more powerful for discontinuers than for continuers. Continuers tend to keep using a new technology even when there are not many people with whom they can discuss the technology. In contrast, discontinuers have been found to depend more on the experiences of other users, communicated via interpersonal channels, in their assessment of the value of a new technology (Parthasarathy and Bhattacherjee, 1998). Thus we propose as our fourth hypothesis:

H4: Social Influence affects Perceived Value more heavily for discontinuers than for continuers.

Compatibility

Compatibility refers to the conformity of an individual user's background with the services he or she is using (Rogers, 1995). Individual backgrounds might be formed by lifestyle and personal knowledge of frequently used products. Rogers (1995) suggests that a user adopts a new system quickly when the system is compatible with his or her everyday experience.

With their awkward input systems and small screens, mobile Internet services are quite different from traditional stationary Internet services, not only in terms of constrained resources but also in terms of the variety of use contexts (Bhagwat and Tripathi, 1994).

We hypothesize that compatibility will be more influential on discontinuers than on continuers. Continuers tend to keep using a new technology even when it is not quite compatible with their existing lifestyle. In contrast, discontinuers may be more conservative, in terms of requiring a new technology to fit into their existing lifestyle. Thus we propose as our fifth hypothesis:

H5: Compatibility affects Perceived Value more heavily for discontinuers than for continuers.

Ubiquitous Connectivity

Ubiquitous Connectivity is the capacity of mobile Internet services to be used anywhere and at any time (Creativegood, 2000). When a user cannot access the stationary Internet for some reason, s/he can tap into Internet services through a mobile device like a cellular phone.

We hypothesize that ubiquitous connectivity will influence perceived value more strongly for continuers than for discontinuers. Continuers have been found to rely more heavily on external sources of information (Parthasarathy and Bhattrcherjee, 1998), and ubiquitous connectivity is the most frequently mentioned characteristic of the mobile Internet (HCI Lab, 2002). Ubiquitous connectivity is one of the defining features of mobile Internet service, and continuers may be better acquainted with this feature, having used the services more than discontinuers. Thus we propose as our sixth hypothesis:

H6: Ubiquitous Connectivity affects Perceived Value more heavily for continuers than for discontinuers.

Perceived Sacrifice

Perceived Sacrifice refers to both the physical costs and the mental effort that users say they encounter when using a given product or service (Zeithaml, 1988). Consumers' costs and efforts need to be considered when evaluating products or services (Zeithaml, 1988; Dodds and Monroe, 1991).

Compared with other information systems, the mobile Internet involves a higher perceived sacrifice. Mobile Internet services involve several types of monetary costs, such as an information usage fee, while most stationary Internet services are free (HCI Lab, 2002).

We hypothesize that discontinuers will be more sensitive than continuers to the perceived sacrifice involved in use of the mobile Internet. Specifically, discontinuers may consider the high cost of the mobile Internet more important than continuers do, significantly decreasing the overall value they find in the mobile Internet. Thus we propose as our seventh hypothesis:

H7: Perceived Sacrifice affects Perceived Value more heavily for discontinuers than for continuers.

Perceived Value and Behavioral Intention

Perceived value, involving all the benefits users find in the purchase and use of a product or service (Zeithaml and Binter, 2000), affects Behavioral Intention, which is the degree of reported intention to use products or services in the future (Wakefield and Barnes, 1996). For example, Oh (1999) found that when a high value was attributed to a specific service, the behavioral intention to use that service in the future was greater.

We hypothesize that perceived value will affect behavioral intention more heavily for continuers than for discontinuers. Continuers may be more sensitive to the perceived value of the mobile Internet because they tended to be early adopters, who are usually more eager to use such services and more sensitive to a rise in perceived value. Thus we propose as our final hypothesis:

H8: Perceived Value affects Behavioral Intention more heavily for continuers than for discontinuers.

RESEARCH METHODOLOGY

Development of Questionnaire

To test the research constructs of our adoption model, we created a set of questions for a nationwide on-line survey of Korean mobile Internet users. Questions for all the constructs were adapted from related studies (e.g., Rogers 1995; Davis, 1989) in order to increase their content validity

Data Collection

After respondents had finished the online survey, their phone numbers and survey responses were sent to the telecommunication companies for data verification. The telecommunication companies checked whether the phone numbers reported were legitimately registered, and whether the owners of the phone numbers had used the mobile Internet at least once in the past. They also classified respondents as continuers or discontinuers. Those who had used the mobile Internet at least once, but had not used it at all in the past month, were categorized as discontinuers,

while those who had used the mobile Internet more than four times in the past month were classified as continuers. The criterion of more than four uses in a month is an industry standard widely used by telecommunication companies for customer management purposes (HCI Lab 2002). After verification and classification, the telecommunication companies returned the survey data to the authors, with phone numbers deleted to protect the privacy of survey respondents. In effect, we were able to increase the validity of survey data without significantly undermining the privacy of survey respondents. Finally, 1,789 respondents were classified as continuers and 1,770 as discontinuers. There is no statistically significant difference between the two groups, and the two groups were treated as two homogenous groups.

Measurement Validation

To ensure construct validity, explorative factor analysis was performed. Among the seven factors originally proposed, only six were extracted as independent variables; these were referred to as Usefulness, Usability, System Quality, Social Influence, Ubiquitous Connectivity, and Perceived Sacrifice. The two questions for the Compatibility construct were divided into Usefulness and Usability, respectively. Therefore, we dropped the compatibility construct in our further analysis and folded the two questions into the Usefulness and Usability constructs. Two factors were extracted as dependent variables: Perceived Value and Behavior Intention. The variances explained were 59.4% for independent variables and 82.7% for dependent variables, and Eigenvalues of extracted factors were all above 1. Cronbach's Alphas for all factors except Perceived Sacrifice exceeded the cutoff point of 0.7. The results of the CFA also indicate that all the questions were well converged into their respective constructs except the two compatibility questions. It was found that the square roots of Average Variance Extracted (AVE) were all larger than 0.5 except Perceived Sacrifice, and also larger than their corresponding correlation coefficients, which indicates that the metrics have appropriate discriminant validity (Gefen et al., 2000).

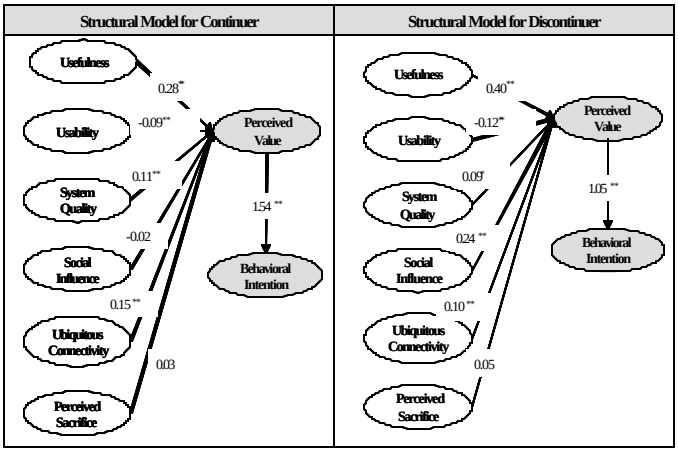


Figure2. Comparison between continuers and discontinuers (*<0.05, **<0.01)

RESULTS

Structural equation modeling (SEM) analysis with LISREL version 8.13 was used to investigate the causal relations among these factors. Figure 2 presents two different causal models, one for continuers and one for discontinuers.

Multi-group analysis using LISREL was conducted for each causal relation in order to compare the strengths of influences for continuers and discontinuers. Chi-squares of the two models were calculated first, and then, after constraining the path coefficient of one variable, the Chi-squares of the two models were compared (Jaccard and Wan 1996). Table 1 summarizes the results of the multi-group analysis.

The impacts of Usefulness and Social Influence on Perceived Value were significantly greater for discontinuers than for continuers at the level of 0.001. Moreover, the impact of Perceived Value on Behavioral Intention was significantly greater for continuers than for discontinuers at the level of 0.001. Finally, the influence of Ubiquitous Connectivity on Perceived Value was greater for continuers than for discontinuers at the level of 0.01. In summary, hypotheses 1, 4, 6 and 8 were supported, whereas hypotheses 2, 3 and 7 were not supported, and finally hypothesis 5 could not be tested in this study.

Type	Chi-square	DF	|Difference|
Unconstraint	4327.94	556	-
Usefulness	4339.11	557	11.17 ***
Usability	4328.28	557	0.34
Sys. Qual.	4327.48	557	0.46
Social Inf.	4356.16	557	28.22 ***
Ubiquitous C.	4331.54	557	3.6**
Per. Sacrifice.	4327.72	557	0.22
Per. Value	4356.41	557	28.47***

Table 1. Sensitivity comparison between continuers and discontinuers (*<0.05, **<0.01, ***<0.001)

DISCUSSION AND IMPLICATIONS

This study attempts to identify important adoption factors that are different for continuers and discontinuers. The results of our online survey indicate that the two groups are different in terms of four adoption factors. First, in terms of the value they perceived in the mobile Internet, discontinuers were more sensitive than continuers to usefulness. Second, social influence was stronger on discontinuers than on continuers. Third, the impact of ubiquitous connectivity was stronger for continuers than for discontinuers. Finally, the impact of perceived value on behavioral intention was stronger for continuers than for discontinuers.

This study has several limitations. First, it suffers from a methodological limitation, in that it relies on an on-line survey. Second, the questions for Compatibility failed to be

converged into a single factor, and the Perceived Sacrifice construct could not meet the reliability and discriminant validity criteria. Third, the criterion that distinguishes continuers from discontinuers was set at more than four times in the month prior to the survey period. Fourth, we assumed that factors important in technology adoption generally will be also important to a comparison of continuers and discontinuers of mobile Internet services. Finally, the online survey was conducted in Korea, mostly with Korean mobile Internet users.

In spite of these limitations, the study has several interesting theoretical and practical implications. First, it views post-adoption behaviors from the perspective of the individual user. Second, the study suggests a model-based comparison method for analyzing post-adoption behavior. Third, this study provides empirical bases for marketing strategies targeted specifically at continuers/discontinuers.

REFERENCES

1. Ajzen, I. (1991) The Theory of Planned Behavior, *Organizational Behavior and Human Decision Processes*, 50, 179-211.
2. Bhagwat, P. and Tripathi, S. K. (1994) Mobile Computing, *Proceedings of Networks* '94, 3-12.
3. Bruner, R. E. (1998) Net Results: Web Marketing that Works, Cybernautics and Usweb Corporation.
4. Businessweek (2002) What Ails Wireless?, available at http://www.businessweek.com/
5. Buyukkoken, O., Garcia-Molina, H., Paepcke, A., (2000), Focused Web searching with PDAs, *Computer Networks*, 22, 213-230.
6. Creativegood (2000) The Wireless Customer Experience, available at http://www.creativegood.com.
7. Davis, F. D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Supportance of Information Technology, *MIS Quarterly*, 13,3, 319-340.
8. DeLone, W. H., and McLean, E. R. (1992) Information systems success, *Information Systems Research*, 3, 60-95.
9. Dey, A. K. (2001) Understanding and Using Context, *Personal and Ubiquitous Computing*, 5,1,4-7.
10. Dodds, W. B. and Monroe, K. B. (1991) Effects of Price, Brand and Price Information on Buyers' Product Evaluation, *Journal of Marketing Research*, 28, 307-319.
11. Downes, L., and Mui, M. (1998) Unleashing the Killer App: digital strategies for market dominance, Harvard Business School Press, Boston, MA.
12. Federal Trade Commission. (2002) The Mobile Wireless Web, Data Services and Beyond. From http://www.ftc.gov/bcp/workshops/wireless/.
13. Gefen D, Straub D.W and Boudreau M.C (2000) Structural Equation Modeling and Regression: Guidelines for Research Practice, *Communication of the AIS*, 4, 7, 1-77.
14. HCI Lab (2002) The Third Mobile Internet User Survey, available at http://www.mbiz.or.kr/survey
15. Jaccard, J., and Wan, C. K. (1996) LISREL Approaches to Interaction Effects in Multiple Regression, Sage University Paper series on Quantitative Applications in the Social Sciences (07-117), Thousand Oaks, CA: Sage.
16. Karahanna, E., Straub, D. W., and Chervany, N. L. (1999) information technology adoption across time, *MIS Quarterly*, 23,2, 183-213.
17. Nielsen, J. (1993) Usability Engineering, New York: Academic Press.
18. Oh, H. (1999) Service quality, customer satisfaction, and customer value. *International Journal of Hospitality Management*, 18, 1, 67-82.
19. Parthasarathy, M. and Bhattrcherjee A. (1998) Understanding Post-Adoption Behavior in the Context of Online Services, *Information Systems Research*, 9,4, 362-379.
20. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994.) Human-Computer Interaction, Boston, MA: Addison-Wesley.
21. Rice, R. E., Grand, A. E., Schmitz, J., and Torobin, J. (1990) Individual and Network Influences on the Adoption and Perceived Outcomes of Electronic Messaging, *Social Networks*, 12,1, 27-55.
22. Rogers, E.M. (1995) Diffusion of Innovations, 4th Edition, New York: Free Press.
23. Sheth, J. N., Newman, B. I., and Gross, B. L. (1991) Why We Buy What We Buy: A Theory of Consumption Values, *Journal of Business Research*, 22, 159–170.
24. Venkatesh, A. (1996) Computers and Other Interactive Technologies for the Home, *Communications of the ACM*, 39,12, 47-54.
25. Venkatesh, V. (2000) Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Supportance Model, *Information Systems Research*, 11,4, 342-365.
26. Zeithaml, V. A. (1988) Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence, *Journal of Marketing*, 52, 2-22.
27. Zeithaml, V. A. and Bitner, M. J. (2000) Services Marketing, Integrating Customer Focus Across the Firm, 2nd Edition, Irwin/McGraw-Hill, Boston.

Finding Common Ground Among HCI Reference Disciplines

Dennis Galletta
University of Pittsburgh
galletta@katz.pitt.edu

Jonathan Lazar
Towson University
jlazar@towson.edu

Judith S. Olson
University of Michigan
jsolson@umich.edu

Dov Te'eni
Tel-Aviv University
teeni@tau.ac.il

Marilyn Mantei Tremaine
New Jersey Institute of Technology
tremaine@njit.edu

Jane Webster
Queen's University
jwebster@business.queensu.ca

ABSTRACT

Five panelists provide an interesting set of contrasting points of view of the HCI field from four distinct disciplines: Business, Computer Science, Information Science, and Psychology. Panelists are asked to respond to six questions in their presentations that address what their particular field offers that is unique, what seems to be quite similar, the effects of the overlaps, and advice for the future. Many of the panelists represent multiple fields, providing a unique opportunity to address the issues of overlap.

Keywords

HCI Field, Research, Business, Computer Science, Information Science, Psychology

INTRODUCTION

Over the last three or four decades, researchers began to investigate the area of human factors in interacting with artifacts in our world, and computer systems quickly became a natural target for such investigation. The emergent field called "human/computer interaction" (HCI) was so well-received by so many researchers that the dams that exist between fields could not hold back a large flood of work that has been completed and much that is upstream, yet to be done. The water metaphor also applies to the ubiquitous application and importance of the work; computers are all around us and are found in many sizes, shapes, and capacities.

Several fields might inform our understanding of these paradoxically troublesome yet valuable devices. Understanding the human requires background in psychology, understanding the computer requires background in computer and information sciences, and understanding the context requires understanding of business and information systems. These three complex areas interact in even more complex ways, providing adequate fodder for investigation by disparate teams for many years to come.

It would be useful at this juncture to stop and examine the rivers of work that feed our wide and winding stream in HCI. What is unique about each source of knowledge? What do we have in common? Where are there overlaps, and are they synergistic and beneficial or redundant and contradictory? This panel will explore these general questions, and will represent their fields in a rare opportunity for such interaction.

QUESTIONS FOR THE PANEL

1. What does your particular source field offer to HCI that is unique to your field? Please feel free to showcase important work in your field that would not be likely to come from other fields.
2. What does your source field offer to HCI that seems to be very much like what is offered by other fields? Please provide some examines to illustrate your points.
3. Which of those overlaps are beneficial, providing a sort of triangulation for our results?
4. Which overlaps are redundant and contradictory, hindering our progress?
5. What trends do you see? Do you see HCI drifting towards any particular field? Is it spreading wider or contracting?
6. What do you recommend for the future? That is, should territories for each source field be "staked out?" Should we stay the course, working hard to navigate the sometimes conflicting currents? Provide a plan that you believe would be useful.

PANELISTS' POSITION STATEMENTS

Panelists provided initial statements that will provide an understanding of their general philosophies. The panel session will provide thorough discussion of these and other issues.

Jonathan Lazar

I am a professor in a computer science department, but my background is in information systems. I also work with sociologists and cognitive psychologists. So while I can actually represent a number of different communities, for the purposes of this panel, I am representing computer science. I do believe that computer science is the root of all Human-Computer Interaction, as Human-Computer Interaction originated in the computer science field in the early 1980s. As HCI has grown and blossomed in different research communities, this has been very valuable, because it can give us multiple views of a problem, and multiple views will always yield a better outcome. To limit our research to only one field is to limit our understanding of a problem. For instance, library scientists have been studying the optimal methods for organizing, searching, and finding information for over 50 years. Why re-invent the wheel when we can learn from other fields? I think that HCI researchers in computer science can offer fresh views of how humans interact with technology. For instance, at CS-based HCI conferences, there are frequently presentations on new interaction methods, everything from pen-based input to desks that can interact with you, or even entire rooms that can interact with you. Without this fresh approach, we can sometimes become limited in only examining what we currently have, and what we currently use. In general, computer scientists tend to build and refine the tools, and then, the other research communities examine how users actually use those tools, how organizations implement those tools, and how these tools can change society.

I see the growth of Human-Computer Interaction reflected in a topic of study near and dear to my heart, the topic of web accessibility. Designing assistive technology was primarily a computer science topic, with help from areas such as industrial engineering and rehabilitation engineering, health science, speech pathology, and special education. Since the implementation of Section 508 in the United States, and similar laws in other countries, the topic of accessibility has taken center stage. More software tools are being built to help build accessible software products. The topic of accessibility is changing from a CS-focused topic to a topic that is explored in multiple research communities, including sociology, information systems, and policy sciences. Computer scientists first build the software, hardware, and other tools. As more researchers from different communities get involved, we can fine-tune the software and hardware products.

There are some challenges to getting the communities talking with each other. Computer scientists and information systems use different terminology, attend different conferences, and do not have frequent opportunities to interact. The end goals of the research are different. The publication outlets are different. Computer scientists are generally concerned with topics such as building better tools or interfaces.

Information systems researchers are interested in understanding usage patterns, organizational uses of technology, and user acceptance of technology. But there are common areas of interest. I have seen CS researchers focus on usage patterns, and I have seen almost all research communities study the problem of long download times. In fact, you never know where useful research will come from. I have recently been studying user frustration, and I have been reading the work of medical school researchers that focus on muscle movement and blood pressure as it relates to frustration.

All research communities have something useful to add to the study of human-computer interaction. I do not think that each community should stake out their territory and claim their corner of the research world. I think that we need to have more inter-disciplinary efforts. I think that the future of human-computer interaction education and research is an inter-disciplinary department, or a College of Information Studies, Information Technology, or my ideal title, "The College of User Interaction." The mission statement for the College of User Interaction should be "We Help People Improve Their Lives Through Technology." Is that a mission that we can all agree on?

Judith S. Olson

I represent three approaches: Business, Psychology, and Information. Given that others on the panel are from Business Schools, I will concentrate on the Psychology and Information approaches.

Psychology has long been about how people interact with the world, from the neural level to individual psychology to social and organizational psychology. From the early 1970s in individual psychology there were investigations of the psychology of programming, followed by how people interact with personal computers and now mobile computing. One of the most influential contributions in HCI was the 1983 book, "The Psychology of Human Computer Interaction," by Card, Moran and Newell. In this book, the authors compiled the major phenomena in psychology that explained how people interact with various computing devices, from Fitts' Law about motor movement and size of targets to the cognition involved in forming goals, unpacking sub-goals, choosing among alternative methods and enacting those methods. They developed a comprehensive model of individual behavior on computers, called GOMS, which has now 20 years of extensions. At the social level, researchers in HCI adopted various laboratory tasks, like the prisoner's dilemma and Arctic Survival tasks, to investigate the ways in which new communication tools like IM, email, video conferencing, etc. changed the dynamic of social interaction. Organizational psychologists looked at how people's contributions to their organization changed with the advent of email and other tools. Many of the early people doing HCI in companies were trained in psychology.

Schools of Information are growing from various roots, most commonly from schools of Library Science. Librarians have long been taught to be user centered, and it is no surprise that they are interested in the digital world and making it easier for people to find the information they seek. So it's natural for schools of information to house programs in HCI. Some of the HCI work coming out of schools of information focus on information retrieval; others on what's going on in the head of the information seeker, whether they really know what they're after. Schools of information, however, have grown to encompass much more than what used to be Information Science. They cover the pricing of information (a non-consumable), the effect of digital information on society (like social capital, the digital divide), alternatives to copyright policy issues (including copyleft and copymiddle!), and the problems associated with archiving digital information (e.g. whether any of it will be readable in 50 years).

Simply put, people doing HCI from the psychological perspective are mostly concerned with more interactive computing and the designs' effects on cognition, decision making, affect etc. People doing HCI from the information perspective are mostly concerned with information access, navigation, and visualization rather than the more interactive applications. Business schools are more concerned with the productivity paradox, the adoption of technology in organizations, and the HCI of e-commerce and issues of stickiness and trust. Although it is natural for each to focus on those aspects, there is no reason we should be restricted to continue in that path and not branch out. Let many flowers bloom!

Dov Te'eni

Having studied and since then taught in a B-school, let me define my source field as MIS, and at the same time, note that much of my reading and some of my publications are in journals that lay outside the management domain. Furthermore, I teach an HCI course in the B-school that overlaps with those taught in CS but I doubt that it would be judged appropriate if it were identical to the CS course. This duality runs through my arguments below.

HCI in the MIS context should explicitly consider the task, and, moreover, the task as seen by the user in the organizational context, such as scheduling a meeting, searching for information or making a decision. This is in contrast to important HCI studies outside MIS that are concerned with human-computer interaction across tasks, e.g., ergonomic design of input devices or appropriate combinations of color on visual displays. It is also in contrast to HCI research that looks at lower level tasks, which I refer to as interaction tasks that are dependent on the specific input output design, e.g., moving a text string from one location to another. Vessey and Galletta's Cognitive Fit Theory (CFT) is a prime demonstration of a contribution that comes from this perspective in which task is central to HCI. I look for the impact of CFT on CS and find it similar, but certainly not identical, to Norman's notion of naturalness. Why have these two concepts not enriched one another? Are there still unnecessary dams between the two fields?

Perhaps it would be useful to think of two layers in HCI research, teaching and practice: the upper task level inherent in a particular domain such as MIS or education, and the lower interaction level, which cuts across domains. Applying HCI at the higher level occurs at the analysis stage of systems development. At the design stage there is substantial overlap due to common interfaces so that guidelines on arranging a screen layout and presenting graphics effectively should be taught and practiced with little distinction (I hesitate to say none) between the domains of application. I anticipate more HCI research in different domains, in other words, I believe HCI is spreading wider.

Concentrating on HCI in MIS, I have some suggestions for the future. Visualize the two layers, on the upper layer are the HCI research efforts within domains such as education, management, information science etc. On the lower layer is the CHI / human factors field. Now draw a vertical line between the two layers from management down to CHI. At the upper intersection is our SIGHCI. It has the all-important role of spreading the word in MIS. Without this close tie between HCI and management in terms of research, publications and teaching we will not be able to impact MIS. But at the same time we also need to maintain the tie with the lower level by building on and adapting research, by participating in academic activities such as SIGCHI, by collaborating in research and, hopefully, by feeding back our research results and impacting the lower layer.

Marilyn Mantei Tremaine

(Subtitle: Mixing Oil and Water: Paradigmatic Differences in HCI)

When I matriculated with my Ph.D., I had job offers from Computer-Science and Information Systems. I chose to enter IS on the advice of Herb Simon who felt a business school was eclectic enough to allow the new field of Human-Computer Interaction. After 8 years I returned to Computer Science and spent 10 years in a CS Department. During this interlude I watched and helped the HCI field grow, primarily through ACM SIGCHI. In my most recent re-incarnation, I find myself chairing an IS Department in a small technical university in Eastern United States, NJIT. Both in my graduate education and in my academic seasoning, I have worked in both the IS and CS / Psychology paradigms and have experience with the differences between both. Most recently, I served as technical program chair for ACM Group 2003 and saved at least 3 good papers from extinction that were scored low merely because of the large paradigm differences between the two camps. What are these differences and what impacts do they make on how research is being done?

I am going to distinguish between the two approaches by calling the IS one, HCI, and the CS / Psychology / Design one, CHI, since this second name was chosen by the community when they created ACM SIGCHI.

A key difference between the two paradigms is their focus. CHI focuses on design and design improvements. Thus, much of the work is in the invention of new interfaces. This invention is purported to be theory-based design, but the theory often follows the design. CHI also has a large computer-science component that examines methods for building classes of interfaces and also a developing graphics design component. In contrast, HCI focuses on the evaluation and impact of the interface design. As such, it takes existing design and examines what aspects of the design are relevant to a user's choice, a user's performance or a user's subsequent behavior with the application. Both HCI and CHI overlap in their research endeavors and HCI can be found inventing new software technology and CHI evaluating designs, but, in general, the goals and methods are quite different for each group.

CHI will tend to use looser qualitative methods and a low number of subjects for data capture whereas HCI will require tight validation of its experimental methods and a larger N. In contrast, at the very detailed design end of CHI, very tightly controlled experiments will be run. This leads to conflict and confusion amongst the research papers in both fields. CHI people are criticized for not running studies based on underlying theory and HCI people are criticized for running studies that are irrelevant. In both cases there are elements of research where the wrong method is applied to the wrong problem and the criticisms are valid. In others, the research question being asked requires methods that are unknown by the opposing research camp.

Jane Webster

My background is in organizational behavior and information systems from a business school. My remarks concerning HCI are from the perspective of MIS. Historically, HCI research and teaching have taken place in computer science, psychology, and MIS departments. Computer scientists focus more on developing technologies for the computer interface, psychologists concentrate more on individual issues, while MIS researchers attempt to truly bring together the technology with the individual in their studies. More recently, with the development of web-based applications and so-called customer relationship management systems, marketers have also started to conduct research in this area.

The interaction between the computer and the human has been a core element of MIS since its inception as a discipline (e.g., Mason & Mitroff, 1973) and regularly shows up in summaries of research areas within MIS (e.g., Swanson & Ramiller, 1993). Nonetheless, it appears to me that HCI remains on the periphery of what many researchers consider MIS.

Why is this? MIS researchers are distinctively positioned to address HCI issues as they focus both on people and information technologies. Nevertheless, most HCI research occurs in computer science, rather than in MIS. Further, most undergraduate HCI courses take place in computer science departments, with a smattering in psychology and MIS departments.

Why do computer scientists dominate this area of research and teaching? This may be because computer scientists are uniquely situated to develop the latest HCI tools and techniques and therefore get the attention of practice. These researchers thrive on developing the latest tools and need to continue to do so. However, they don't follow their tool development to market in terms of research; that is, they generally do not conduct large-sample tests on their HCI technologies. Quite the opposite, these technologies are "moving targets" that they are continually extending and improving.

I will put forth the perhaps controversial proposition that computer scientists' research often hampers development of the HCI field for several reasons. First, computer scientists often seem to "come up" with reasons for their technologies, post-hoc. For instance, they may draw on "common knowledge" of human behaviors or even theories from psychology to "justify" applications for their tools. However, I will demonstrate how this might do more disservice to the field (than not suggesting these applications at all). Second, they generally do not test out their assumptions regarding appropriate applications through controlled studies such as experiments or through large-scale field studies. In contrast, they often demonstrate their technologies to like-minded computer scientists. Further, the developers themselves and their colleagues frequently act as the "testers" for these applications. I will use recent HCI articles from computer science to illustrate these arguments.

This void in the large-scale study of actual technologies with employees represents a substantial opportunity for MIS researchers. Further, with the continuing development of Web-based applications, I believe that HCI research and teaching will continue to grow in MIS departments. Optimistic signs of this have been recent special issues of Information Systems Research on web-based applications and forthcoming HCI publications in MIS outlets. Therefore, I hope to see HCI move from the periphery of MIS to a more central role in the near future.

REFERENCES

1. Mason, Richard O. and Ian I. Mitroff. (1973). "A Program for Research on Management Information Systems." Management Science 19(5): 475-487.
2. Swanson, E. Burton, and Neil C. Ramiller. (1993). "Information Systems Research Thematics: Submissions to a New Journal, 1987-1992." Information Systems Research 4(4): 299-330.

Is Relevance Relevant? Investigating Coherence in Knowledge Sharing Environments

Andrew Gemino
Simon Fraser University
gemino@sfu.ca

Drew Parker
Simon Fraser University
drew@sfu.ca

Adrienne Olnick Kutzschan
Queen's University
aok@business.queensu.ca

ABSTRACT

This paper focuses on the impact of relevant backgrounds on computer-mediated knowledge sharing and individual knowledge acquisition. An experiment is described based on the coherence principle from the Cognitive Theory of Multi-Media Learning. Results suggest groups using visual chat scored higher in retention and understanding than individuals working alone. In addition, participants using visual chat with relevant backgrounds obtained higher levels of understanding than participants using no relevance or irrelevant backgrounds. These results support the coherence principle in the cognitive theory of multimedia learning and suggest new directions in the design and evaluation of knowledge sharing environments.

Keywords

Knowledge sharing, collaborative learning, computer mediated communication, knowledge acquisition.

INTRODUCTION

Advances in information technology (IT), including synchronous person-to-person communication, widen choices for business communication (Smith et. al., 2003). Companies are using IT to facilitate knowledge sharing by supporting activities such as electronic meetings, international team development, and electronic forums (Brazelton and Gorry, 2003; Chai et. al, 2003). While collaborative systems exist, a positive link between collaborative technology and knowledge sharing has not been established (Kock and Davison, 2003). This has led to a call for research in this area: "To what degree does the application of IT to knowledge transfer increase the knowledge transferred among individuals" (Alavi and Leidner, 2001, p. 139).

This paper focuses on two questions related to this discussion: 1) do group members supported by collaborative technology acquire more knowledge than individuals without the opportunity for collaboration, and 2) does the collaborative environment make a difference in the level of knowledge acquired? Much depends on how the technology can be used to aid individual knowledge acquisition. The track record of application designers has been weak (Landauer, 1995). Norman (1990) suggests the overriding issue may be technology-centered design as opposed to a user-centered approach. Mayer (2001) supports this suggestion and notes we are in early stages in understanding how to use technology effectively.

Purpose Statement

This paper presents results from an experiment manipulating the design of a human-computer interface to evaluate the effectiveness of a collaborative knowledge sharing environment. Focus is placed on one element of the environment: the background. The coherence principle from the cognitive theory of multimedia learning (Mayer 2001) is used to hypothesize that a background in a computer mediated collaborative environment containing relevant information will lead to a higher level of individual understanding than a background with either no relevant or irrelevant information. The dependent variable is the level of understanding as developed by a group member. An experiment with a control group and three treatments is described. Results from this experiment support the coherence principle and provide two results: 1) that distributed groups can be more effective in knowledge sharing than working individually and 2) that the environment for knowledge sharing can make a difference in the level of understanding.

BACKGROUND

Knowledge sharing is defined here as a process of communicating explicit representations of knowledge (diagrams, documents, e-mails) among a group with the purpose of fostering understanding. A literature review revealed an array of studies and perspectives. In considering our focus on computer-mediated environments to support knowledge sharing, three issues were apparent 1) how individuals acquire knowledge, 2) how this knowledge acquisition might be supported collaboratively, and finally 3) how this collaboration might be supported by technology. Frameworks considered from each of these areas are summarized in Table 1 below.

The research areas described above might initially suggest disparate perspectives on knowledge sharing. On deeper inspection, however, the models share much common

ground. The disparity is more the result of a different focus than different perspectives on knowledge transfer. For example, if we assume the "task dimension" in the Dennis et. al (1988) GSS model is knowledge sharing, then all three models suggest sharing knowledge is a process, with potentially measurable outputs. The difference between the models is in the inputs recognized in the process. The Alavi and Leidner (2001) knowledge transfer model recognizes different knowledge types (explicit, tacit) and group memory (semantic and episodic) as relevant inputs. The knowledge acquisition model (Mayer, 1989) would view group memory as residing within "individual characteristics". The knowledge types would be considered "content" which is represented to individuals using various "presentation methods" (verbal, visual or multimedia). In turn, these presentation methods can be viewed as functions supported by the "technology" dimension of the Dennis et. al. (1988) GSS model. The GSS mode recognizes the influence of organization and context explicitly, but this is also recognized by Alavi and Leidner (2001) who state the important influence of the organization in the knowledge sharing process when they note: "*This view of organizations as knowledge systems represents both the cognitive and social nature of organizational knowledge and its embodiment in the individual cognition and practices as well as the collective (i.e. organizational) practices and cultures*" Alavi and Leidner (2001, p. 115).

Research Area	Primary Refer.	Analysis Focus	Inputs Recognized
Individual Knowledge Acquisition	Mayer (1989)	Individual	Content Presentation Method Individual Characteristics
Knowledge Transfer	Alavi and Leidner (2001)	Individual/ Group	Knowledge types Group Memory
Computer Mediated Group Support (GSS)	Dennis et. al (1988)	Group	Group Task Context Technology

Table 1. Summary of Background Research

Assumptions

Our consideration of literature supporting knowledge sharing environments suggests a process model and sets of inputs to consider in knowledge sharing. These models do not provide, however, a theory explaining how knowledge acquisition might occur. Before introducing theory we make three assumptions guiding our choices in theory development. The first assumption is that knowledge is a justified belief (Nonaka, 1994). In taking this approach we accept the constructivist approach (Chai et. al., 2003) and choose to separate knowledge held within individuals from information represented and stored externally. The implication is that knowledge can only be held within individuals and suggests the output of knowledge sharing can be measured at an individual level.

The second assumption follows from the constructivist view and recognizes that knowledge presented is not necessarily equal to knowledge gained. Developing knowledge requires individuals to actively engage in selecting, organizing and integrating presented information. Two persons presented with the same material may develop different levels of knowledge depending on what information they paid attention to and how it was integrated into memory.

The final assumption is to suggest that groups can be viewed as perceptions of other individuals. While groups have been studied at both the group and individual level, we assume a "group" as viewed from one member's perspective can be different from the same "group" viewed for a different member's perspective. It is our view that perceptions of the group and of the people within the group can be assessed at an individual level.

These assumptions enable us to focus attention on the individual. We will view information presented to the individual as content, in one presentation format or another, and recognize that individuals can differ in the level of knowledge attained from the viewing the same content. These assumptions provide the basis for suggesting a theory of knowledge sharing focused at the individual level. We therefore suggest the Cognitive Theory of Multimedia Learning (CTML), introduced by Mayer (2001), as a useful theory for understanding knowledge sharing in computer mediated collaborative environments.

COGNITIVE THEORY OF MULTIMEDIA LEARNING

The CTML has been developed through more than a decade of empirical work using a variety of experimental data (Mayer, 1989; 2001). This foundation has been used to compare presentations in science learning (Mayer and Gallini, 1990), multimedia explanations (Lim and Benbasat, 2002) and conceptual modeling in systems analysis (Bodart et. al, 2001).

The theory focuses on the interaction between a learner and presented information. It argues for two pathways in cognition, verbal and visual. While independent, these channels can interact in working memory. When a person views presented material, relevant sensory information is selected through the verbal and visual channels into working memory. This information is organized into visual and verbal models. Linkages between these models can be created in working memory. These two models are then integrated with prior knowledge to develop new understanding. An overview of the theory is provided in Figure 1.

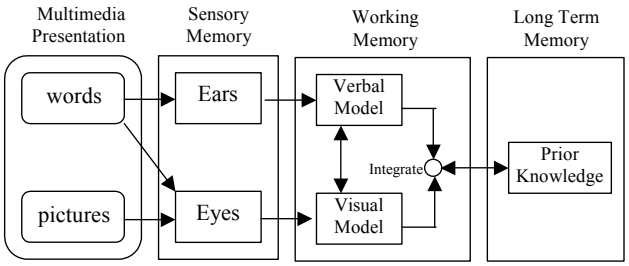


Figure 1. Overview of CTML (Mayer , 2001)

Assuming dual channels exist, it can be argued that not all presentation formats are equally successful in producing learning outcomes. For example, providing a written passage while simultaneously narrating the passage will not be an effective presentation format because both of the presentation methods are utilizing the verbal channel. Since cognitive resources are limited, using the same channel creates a capacity conflict resulting in only a portion of the information reaching the learner. In addition, the ability to create linkages between verbal and visual models is lost as no visual information is provided.

Learning Outcomes

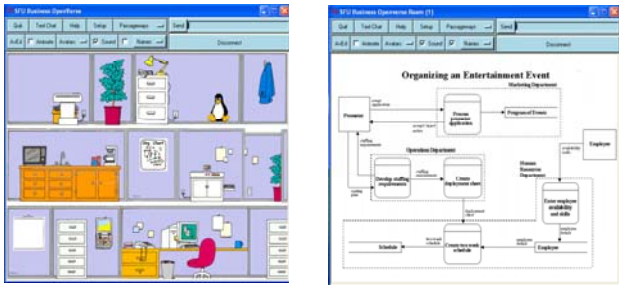
Mayer (2001) suggests three outcomes when presenting material: 1) no learning, 2) rote learning and 3) meaningful learning. These outcomes are based on measures of two variables: retention and transfer. Retention is the comprehension of material being presented. Transfer is the ability to use acquired knowledge to solve new but related problems. For example, if presented with an explanation of how a car's braking system works, a retention question might be "What are the components of a braking system," but a transfer problem might be "How can you make a car stop faster?"

Regarding learning outcomes, no learning occurs when retention and transfer are low. Rote learning occurs when retention is high and transfer measures are low. Meaningful learning occurs when both retention and transfer are high. The high transfer score indicates a high level of understanding of the material

HYPOTHESES

An experiment was developed to test the CTML in a distributed knowledge sharing environment. We focused on a system analysis and design task; specifically, the interpretation of a system analysis diagram. This task is appropriate because: 1) it is an explanative task containing explicit knowledge, 2) it combines pictures and words, 3) it not simple to understand, and 4) it is often accomplished in groups supported by technology.

The technology used to support this task was a synchronous visual chat using peer-to-peer technology to emulate distributed group discussions (Smith et al, 2003). Visual chat is a type of chat where group discussions occur in a "room" often with the use of avatars (small icons representing a participant) and an ability to move in the discussion room. Two examples of the chat rooms used in the experiment are provided in Figure 2.



Cartoon Room **DFD Room**

Figure 2. Visual Chat Rooms used in Experiment

Three treatment groups and a control group were compared. The control group worked individually with no support from technology. The first treatment group (plain room) was provided with a visual chat environment featuring a white background. The second treatment (cartoon room) was provided a chat environment with an irrelevant cartoon background The third treatment (DFD room) was provided a background with a relevant dataflow diagram embedded into it.

In a study of collaborative technology, Alavi (1994) argued that a group support system "enhances the effectiveness of collaborative learning … by increasing group process gains and decreasing group process losses.' (p 163). Recognized process gains included more information being generated, potential for synergies, group motivation and more effective evaluation of information. Based on a body of research in collaborative learning, a specific hypothesis was formulated:

> H1: The control group, who work alone, will score lower in both retention and transfer than the treatment groups (who work in groups).

The Coherence Principle

The background may add to the chat experience, however, it can also distract. The coherence principle, derived from the CTML, suggests that uninformative and irrelevant information distracts from the potential for understanding and reduces the coherence of the message. Irrelevant information must be filtered. This filtering uses valuable cognitive resources, providing less for knowledge development. If irrelevant information is selected into working memory, cognitive effort is wasted on integrating unrelated words and images. The coherence principle, along with the definition of meaningful learning, enables us to develop our second hypothesis:

> H2: A Visual chat environment with a relevant background will produce higher learning outcomes (higher transfer scores) than visual chat environments with either irrelevant or not relevant backgrounds.

METHOD

Participants included 101 undergraduate students in management information systems familiar with dataflow diagramming. The experiment took place in a computing lab with 30 stations. Twenty-nine participants served as a control group. They worked alone (no groups) with exactly the same materials as other treatments. Remaining participants were randomly assigned to a group with 3, or if necessary, 4 members. Each group was then randomly assigned to one of three treatments. An average of 15 people and 5 different groups worked simultaneously in each session.

Group members were spread throughout the lab to eliminate face-to-face discussion. Participants started with a pretest. Next was a short introduction to the chat tool, OpenVerse (www.openverse.com), which was used in the experiment. OpenVerse is an open source application that runs on variety of platforms and provides the ability to embed GIF (Graphics Image Format) files into the background, a function required for this experiment. To familiarize group members with the group and the chat tool application, participants were then asked to use OpenVerse to select a group name. To increase task participation, a monetary incentive was provided to the group with the most creative name in each session.

After completing the group naming exercise, participants were asked to answer the multiple-choice questions in Part I using OpenVerse as a discussion forum. In Part II, participants were asked to turn off the computer monitors and answer questions individually. Participants were given 2.5 minutes to answer each transfer question.

The only difference between treatment groups was the background image used. The first treatment group used a plain white background for chatting. The second treatment group used a cartoon as an irrelevant background (Figure 2), and the third treatment group used the relevant Data Flow Diagram (Figure 2). In addition to the background, each participant was also provided with an "avatar" and a text chat toolbar to view the discussion.

Paper-based materials included a pretest survey, a case description, the DFD diagram and two tasks. The pretest gathered participants' demographics and experience regarding computer skills, discussion tools and Data Flow Diagrams. In Part I, a one page written case with an accompanying DFD was given to each participant. Along with 12 yes/no/uncertain questions These questions familiarized participants with the diagram and measured the level of retention. In Part II, the chat discussion was stopped and case materials and computing resources were taken away. Participants then answered four open-ended transfer questions. These questions were used to measure the level of transfer, which is a measure of the understanding developed from viewing the case materials (Mayer, 2001).

RESULTS

Two ANOVA analyses were performed on the data, one for each dependent measure. The means and standard deviations of the dependent measures (retention and transfer) across the three treatment groups are provided in Table 2. The F statistic and p-value results of the ANOVA tests for each dependent variable are provided in the final column of Table 3. A post hoc analysis using least square differences (LSD) are provided in Table 3 to show comparisons across treatment groups.

Measure	Treatment Groups				
	Control Group n=29 Means (SD)	Plain Room n=21 Means (SD)	Cartoo Room n=27 Means (SD)	DFD Room n=24 Means (SD)	Sig. n=101 F Stat (p-val)
Retention	8.07 (1.44)	9.10 (.54)	9.22 (.42)	8.79 (.41)	9.83*** (0.000)
Transfer	9.00 (1.89)	10.43 (2.36)	10.04 (3.38)	11.92 (2.67)	5.47** (0.004)
** significant at the .01 level, *** significant at the .001 level					

Table 2. Means, Std. Dev. And ANOVA Results

Dependent Variable	(I) Treat	(J) Treatment	Mean Difference (I-J)	Std. Error	p-value
Retention	Control Group	Plain Room	-1.03	0.25	0.000***
		Cartoon	-1.15	0.23	0.000***
		DFD Room	-0.72	0.24	0.003**
	DFD Room	Control	0.72	0.24	0.003**
		Plain Room	-0.30	0.26	0.242
		Cartoon	-0.43	0.24	0.079
Transfer	Control Group	Plain Room	-1.43	0.75	0.061
		Cartoon	-1.04	0.70	0.144
		DFD Room	-2.92	0.73	0.000***
	DFD Room	Control	2.92	0.73	0.000***
		Plain Room	1.49	0.79	0.061
		Cartoon	1.88	0.74	0.012*
* significant at the .05 level, ** at the .01 level, *** at the .001 level					

Table 3: Post Hoc Comparisons (Least Square Difference)

In the post hoc comparisons, retention scores showed significant differences between control group and the three treatment groups. This results supports hypothesis H1 and suggests the level of retention was significantly lower in the control group (operating individually) than in the treatment groups (operating as groups). Furthermore, no significant differences were found between the three treatment groups in regards to retention. Since the same materials were used in all treatments, the results suggest treatments provided relatively similar levels of content. In other words, regardless of which treatment was provided, individuals were able to generate similar retention scores.

Transfer measures showed differences in the anticipated direction. In post hoc comparison between treatment groups, group members in the DFD room scored higher on transfer than any of the treatment groups, and significantly higher than the control group and "cartoon" room participants. These results suggest that although the retention scores across treatments were the same, the added coherence provided by embedding the DFD into the chat tool enabled participants to develop a higher level of understanding. This result suggests a relevant background can have measurable positive effects on viewer understanding. Chat tools may be an exciting step forward in collaborative knowledge sharing, however the performance of group members can be affected by features in the environment. Careful thought should therefore be placed on delivering environments that are effective and engaging.

CONCLUSION

This experiment has provided two interesting results. The results suggest collaboration in a computer mediated collaborative environment can be more effective than working alone. Additionally, the environment in which groups work matters. Groups working with the relevant information in the background developed a mental model that was facilitated by coherence. This suggests the potential to improve the design of distributed group interfaces to improve knowledge sharing within organizations. Relevant backgrounds are preferable to the plain white or "interesting" backgrounds currently provided by most visual chat environments.

The results also provide support for the Cognitive Theory of Multimedia Learning and the design principles for multimedia messages that the theory suggests. The results suggest the coherence principle is of direct relevance to both the users and designers of visual chat environments and reaffirm the importance of diagrams and visual information in knowledge sharing.

ACKNOWLEDGMENTS

This research was supported from a grant from the Social Science and Humanities Research Council of Canada. The authors also wish to express their appreciation to the three reviewers for their insights and helpful comments.

REFERENCES

1. Alavi, Maryam (1994), "Computer Mediated Collaborative Environment", MIS Quarterly, Vol. 18, 2, p. 159-174.
2. Alavi, Maryam, and Leidner, Dorothy, E. (2001), "Research Commentary: Technology Mediated Learning – A Call for Greater Depth and Breadth", Information Systems Research, Vo. 12, No. 1, March 2001, p. 1-10.
3. Bodart F, Patel A, Sim M, Weber R. Should optional properties be used in conceptual modeling? A theory and three empirical tests. Info Systems Research 2001 (December); 12(4): 384-405.
4. Brazelton, Jessica and Gorry G. Anthony, "Creating a knowledge-sharing community: if you build it, will they come?" Communications of the ACM. Vol 46, Issue 2, Feb. 2003 p. 23 - 25
5. Chai, Kah-Hin, Gregory, M. and Shi, Y. "Bridging islands of knowledge: a framework of knowledge sharing mechanisms", Int. J. of Technology Management, Vol. 25, No. 8 , 2003, p. 703-727
6. Dennis, A.R., George, J.F., Jessup, L.M., Nunamaker, J.F. and Vogel, D.R., Information Technology to Support Electronic Meetings, MIS Quarterly (22:4), December, 1988, pp. 591-624.
7. Kock, N.; Davison, R "Can lean media support knowledge sharing? Investigating a hidden advantage of process improvement", IEE Transactions on Engineering Management, Vol. 50 (2), May 2003, p. 151-163.
8. Landauer, Thomas K., The Trouble with Computers: Usefulness, Usability and Productivity, MIT Press, Cambridge, Massachusetts, 1995.
9. Lim KH, Benbasat I. "The influence of multimedia on improving the comprehension of organizational information". JMIS 2002 (Summer); 19(1): 99-127.
10. Mayer, Richard E., "Models for Understanding", Review of Educational Research, Spring 1989, Vol. 59, No. 1, pp 43-64.
11. Mayer, Richard E. and J.K. Gallini, "When is an Illustration Worth Ten Thousand Words?", Journal of Educational Psychology, 1990, Vol. 82, no.4,pp. 715-726.
12. Mayer, Richard E., Multimedia Learning, Cambridge University Press, New York, 2001.
13. Nonaka, I., "A Dynamic Theory of Organizational Knowledge Creation", Organization Science, (5:1), Feb. 1994, p. 14-37.
14. Norman, Donald A., The Design of Everyday Things, Doubleday, New York, 1990
15. O'Dell, C. and Grayson, C.. If only we knew what we know: Identification and transfer of internal best practices. California Management Review, 40(3): 1998, p.154--174.
16. Smith, Helen, Clippinger, J. and Konsynski, B., "Riding the Wave: Discovering the Value of P2P Technologies", Communication of the AIS, January, 2003, Volume 11, Article 4, p. 1-27.

Interactivity and Control: The Case of Dynamic Maps for Navigation in Hypertext

Dov Te'eni
Tel-Aviv University
teeni@tau.ac.il

Hadar Ronen
Bar-Ilan University
Hadar_r@bigfoot.com

ABSTRACT

Rich information environments such as online tutorials and web-books pose considerable difficulties for users, of which the most notable is being 'lost in hypertext'. If these environments are to become commonplace, they must be designed to relieve users of these difficulties. In this paper we study the effects of dynamic navigational maps on orientation and search performance. We designed a conceptual map that tracks the user's position vis-à-vis the content of the web-book and the history of the user's visits. We show how these maps improve search performance significantly in terms of efficiency (number of clicks) but only weakly in terms of time or accuracy. We call for more research on how to enhance user control in complex information environments.

Keywords

Web-book, e-book, orientation, disorientation, navigation, dynamic maps, feedback, search efficiency.

INTRODUCTION

Finding information on the Web and within a specific website is today a common, yet increasingly difficult, activity. In particular, the task of interacting with a specific website is becoming difficult because of the growing size and complexity of websites and their diversity in terms of types of content and structure. The most frequently reported difficulty is 'getting lost' in hypertext, making it urgent and important to seek effective navigational aids to facilitate better orientation when traveling the website. Indeed, the phenomenon of "Lost in hypertext", i.e., losing your sense of location and direction (Conklin, 1987) seems to be spreading as the population of users grows. Furthermore, the aggregate cost of lost productivity is enormous due to the vast number of users now encountering disorientation problems.

Our research aims at a better understanding of the role navigational aids can play in facilitating more efficient interaction. More specifically, we look at common navigation devices, i.e., navigational maps, and use their interactive features to provide the user with dynamic feedback on the user's progression within the site. In effect, the feedback provided by these interactive maps enhances the user's sense of control (reduces the disorientation). The underlying theory is therefore one of interactivity and control, and the use of navigational maps is seen to be a special case of the broader call for enhanced control in the face of complex interactions users face nowadays. Thus, websites create new navigational problems of getting lost in hypertext, on the one hand, but offer new opportunities for overcoming these problems with dynamic feedback, on the other hand. This research concentrates on the impact of dynamic maps on the user's behavior and sense of orientation in websites that function as professional books e.g., online manuals, learning materials and professional papers. We refer to such websites as web-books.

Static and dynamic diagrams in web-books

Printed professional books such as learning materials and reference manuals frequently use diagrams to explain concepts and structures, e.g., a class diagram of an information system. Diagrams are also used to help provide the 'overall picture' of the book, i.e., help organize the materials. In this latter sense, diagrams functions like maps, helping the reader navigate in the conceptual world of the book by providing the reader with a stable context for understanding the more detailed arguments made throughout the book. Thus, we regard these diagrams as conceptual maps that provide the reader with context information.

Such diagrams, when they include concepts and relationships among concepts, are often referred to as cognitive maps. Furthermore, the overall map is often exploded (zoomed in) to produce more detailed maps of more limited scope. The more detailed maps are used to provide a narrower context for the more detailed materials thus creating a hierarchy of diagrams. For example, you may expect a very general diagram outlining the entire structure of the book in the introductory chapter and more detailed diagrams in the subsequent chapters. All in all, diagrams are useful devices for depicting the structure of the book as it relates to the content and thus helping the reader to navigate the book.

Clearly, diagrams can also be used in web-books and other online documents in the same fashion. Furthermore, in online documents they can serve as sensitive maps that incorporate links to specific locations in the text, turning

the map into an effective navigational device by enabling direct access to the target location. Indeed, several researchers have proposed cognitive maps as navigational aids to hypertext (Dillon et al., 1993; Nielsen, 2000; Germonprez and Zigurs, 2003).

The discussion so far has been about static maps. Below, after we develop the notion of orientation, we return to the design of dynamic maps, which we claim are more effective in enhancing orientation and search performance.

Maps are useful for both target-led searches and browsing. Although we emphasize search performance in the experiment described below, we also look at the effect of maps on browsing. Several distinct research streams suggest that finding information involves a duality of interaction modes: searching (or querying) that is focused on an explicit search target and browsing that gathers information whilst scanning with only a rough idea, if at all, of a target (Byrne et al., 1999; Choo et al., 2000). Moreover, both modes require control over the interaction process, although the consequences of poor control, such as low precision or discomfort, may differ from one mode to another (Te'eni and Feldman, 2001).

Dynamic maps combine the powerful spatial features of static navigational maps discussed above with the ability to provide timely context information (Zizi and Beaudouin-Lafon, 1995). For instance, site maps that register the paths taken and provide them in an auxiliary window seem to be useful in helping plan your next move (Zaphiris, Shneiderman and Norman 1999). We believe the same idea holds true for conceptual maps (e.g., cognitive maps) that are supplemented with context information dynamically, which according to the findings noted above should promote orientation. Moreover, the placement of the context information (feedback) on the navigational map is an advantage in comparison to presenting the paths on auxiliary windows because it requires less cognitive effort, as in the principle of direct manipulation.

We summarize this discussion on the impact of dynamic maps in the following hypotheses 1& 2.

Hypothesis 1. Searching with dynamic navigational maps, compared with static maps, will lead to higher orientation.

Overall, we expect improved orientation to lead to better search performance. Improved orientation, by definition, should reduce errors in navigation, should therefore also result in fewer unnecessary steps (clicks) and, correspondingly, a shorter search time. Indeed, Nielsen (2002) studied Site Map usability, and found that on average, users visited 0.3 erroneous destinations for each task that asked them to go to a page linked directly from the site map. (Erroneous clicks increased to 1.1 per task when we asked users to perform tasks that required going to pages that were two clicks from the site map, rather than directly linked).

Hypothesis 2. Higher orientation will lead to improved search performance

2.1 Higher orientation will lead to higher search accuracy.

2.2 Higher orientation will lead to faster searches.

2.3 Higher orientation will lead to shorter searches.

EXPERIMENT

This section describes the web-book and its two versions (static and dynamic) on which the experiment was conducted, the experimental procedure, the pilot study and the measurements.

The web-book is a professionally edited article about organizational communication that also appears in print (see details in Te'eni, 2001). The 100-page article was reorganized in a hierarchical-networked structure.

As can be seen in Figure 1, the interface is divided into 3 main frames: in the left column is the TOC, in the upper right area is the navigation diagram, and in the lower right area is the text display along with the forward / backward arrows. Each level in the TOC can be expanded or contracted. Selecting an item through any of the three navigation aids updates both the text area and the current pointer in the TOC. The reader can travel to the topics noted on the navigation diagram and zoom in and out of more detailed diagrams.

The difference between the static and dynamic maps is that the latter includes the current location in a red wire-frame box (WFB) and the previous locations visited painted in gray. For the purpose of this publication and in order to depict these additional features in black and white print, we annotated the gray areas in the online system with diagonal lines and the redrew the WFB with thicker lines.

Procedure

Fifty students in an Information Studies program participated in the experiment, and were randomly distributed to one of the two experimental groups: one group received the static navigational map and the other received the dynamic navigational map.

Tasks

Each subject was asked to perform 8 search tasks. The first two were taken as exercises (this is in addition to the short tutorial they received). The last two were rendered unfit for analysis because of the low proportion of people finishing the tasks successfully. Our analysis concentrated on the four middle tasks, which are detailed in Table 1. Each task required the subject to read the question and find the answer in the text.

What is the name of communication goal #3?
Read about communication impacts and name the person who developed the theory of communicative action!
With reference to Proposition 2B: a) what is the name of the proposition? b) on the impact of which strategy does it hypothesize?
What type of communication complexity affects the strategy of affectivity?

Table 1. The four search tasks in the experiment.

Dependent variables

1. Reaction time – the amount of time a user took to perform each task. It is calculated by measuring the time taken to get from the current page to the target page. The measurement was calculated according to the log file mechanism.
2. Efficiency – a count of the clicks a user made to reach the target page minus the optimal number of clicks for that task. The measurement was based on the log file mechanism.
3. Accuracy - the percent of correct answers to the search tasks.
4. Accurate representation of the web-book – the percent of correct answers to question about the representations of the Web-book's structure as modeled by the user.
5. Orientation - the aggregation of four Likert type questions about the user's sense of orientation (see below).
6. Diagram interface usage – the number of tasks in which the user clicked the diagram links. This variable was used as a manipulation check. It was based on the log file mechanism.
7. Total navigational aids usage – total number of clicks on diagram, TOC menu and navigation arrows. The measurement was based on the log file mechanism.

RESULTS

Descriptive statistics

The average experiment time was 1033 seconds (17 minutes and 13 seconds). The shortest time was 617 seconds, and the longest 1600 seconds. The static-map group averaged 1065.6 seconds (std 256.05) and the dynamic-map group averaged 999.48 seconds (std 193.83). Table 2 summarizes the data for each group (manipulation) for the dependent variables, averaged over the four tasks.

Manipulation	Reaction Time	Efficiency	Accuracy
Average – Static map	132.24 (122.78 std)	2.11 (1.40 std)	0.66 (0.48 std)
Average– Dynamic map	118.17 (89.04 std)	1.54 (1.42 std)	0.63 (0.49 std)
P-Value	0.3546	0.0049	0.6595

Table 2: Data for the aggregated four tasks

Results on Hypotheses

1. *Searching with dynamic navigational maps will lead to higher orientation than static navigational maps.*
2. *Higher orientation will lead to improved search performance*

These two hypotheses were examined in a path analysis, described below. The only path that was found to be statistically significant was the search efficiency. However, there were no corresponding paths for reaction time and accuracy.

Using linear regression we tested the mediator effect of orientation and concluded that static-dynamic maps affect performance through orientation.

In addition we also examined the subjects' perception of the structure of the Web-book, i.e., their mental model of the book (see dependent variable #4). We found that using an interface with a dynamic diagram, improves the accuracy of mental model representation of knowledge architecture within users. See Table 3.

Manipulation	Correct answers to questions checking the representations of the Web-book structure
Average – Static Diagram	0.40 (0.27 std)
Average– Dynamic Diagram	0.60 (0.25 std)
P Value	0.005

Table 3: Accuracy of web-book structure.

DISCUSSIONS

We looked at web-books – a growing information environment that has been shown to have challenged our cognitive capabilities leaving many of us 'lost in hypertext'. For this environment we sought effective navigational designs to help overcome these disorientation problems. The most important result found in this research is the impact of dynamic maps, in

comparison to static maps, on the user's orientation and on search efficiency. The impacts of dynamics maps on search time and on accuracy, although in the expected direction, were not statistically significant. The lack of correspondence between the efficiency (the number of steps taken) and the reaction time deserves our attention. The exact trace of the steps taken provided a clear comparison between the actual path and the optimal path. Clearly, the users working with a dynamic map were able to come closer to the optimal path. One possible explanation for this was the need to consider the information and use it to plan ahead. This takes time, which may however decrease with experience. In any event, the impact on efficiency is an important result. Our aim in this paper was to demonstrate how navigational aids can be designed to enhance control (orientation) and boost search performance in information environments that have become extremely popular. We claim however that there is an important more general message here – the need to actively seek better control for the user.

Generalization

We began this research with a claim that dynamic maps and navigation in web-books is a special case of interactivity and control. We showed how dynamic maps provide context information (feedback) that enhance orientation and, thereby, search efficiency. Orientation represents control and the lack of it (disorientation) is the loss of control. Dynamic maps are one form of feedback but we can think of other forms that may boost our control over the human-computer interaction. Unlike printed books or non-interactive systems, online hypertext documents are designed to support instant interaction with users and should do so by providing dynamic feedback as in the dynamic maps. Online hypertext may be designed, however, to provide other forms of feedback to enhance user control. For example, a better understanding of the hypertext (the mental model) will likely enhance control. Other forms of dynamic feedback provided in audio or visual cues, could inform the user about her or his surroundings in the hypertext (who are my current neighbors). Again, such information is known to be important to people for their interactions by reducing uncertainties – why not provide it to enhance user's control?

Our aim in this paper was to demonstrate how navigational aids can be designed to enhance control (orientation) and boost search performance in information environments that have become extremely popular. We claim however that there is an important more general message here – the need to actively seek better control for the user.

ACKNOWLEDGMENTS

We thank the three anonymous reviewers for their helpful comments.

REFERENCES

1. Byrne, M.D., B.E. John, N.S. Wehrle, and D.C. Crow (1999) The tangled web we wove: A taxonomy of WWW use, *CHI'99 Proceedings of the conference on Human Factors in Computing*, ACM. Pittsburgh, PA USA, pp. 544-551.
2. Choo, C.W., B. Detlor and D. Turnbull (2000) Working the web: An empirical model of web use *Proceedings of the Hawaii International Conference on Systems Sciences*, Maui, Hawaii.
3. Conklin. J. (1987) Hypertext: An introduction and survey**.** *IEEE Computer*, 20(9): pp. 17-41.
4. Dillon A., McKnight C/ and Richardson J. (1993) Space – The final Chapter *or* Why physical representations are nit semantic Intentions, *Hypertext - a psychological perspective*, E. Horwood publishing NY, pp. 169-188.
5. Germonprez M. and Zigurs I. (2003) Causal factors for web site complexity. Working Paper: *Case Western Reserve University* URL: weatherhead.cwru.edu/WebSiteComplexity.pdf.
6. Nielsen J. (2000) Designing Web Usability: The Practice of Simplicity *New Riders Publishing, Indianapolis*
7. Nielsen J. (2002) Site Map Usability, *AlertBox* (January 2002)
8. Te'eni D. (2001). A Cognitive-Affective Model of Organizational Communication for Designing IT. *MIS Quarterly,* 25 (2), 1-62 (available online at MISQ.org site).
9. Te'eni D. and Feldman R. (2001) Performance and satisfaction in adaptive websites: a laboratory experiment on search tasks within a task-adapted website. *Journal of AIS,* 2(3), 1-28.
10. Zaphiris P, Shneiderman B, Norman KL (1999). Expandable indexes versus sequential menus for searching hierarchies on the World Wide Web. *Computer Interaction Laboratory Technical Report* HCIL-99-15. college Park, MD: University of Maryland
11. Zizi, M & Beaudouin-Lafon, M. (1995). Hypermedia exploration with interactive dynamic maps. *International Journal of Human-Computer Studies*, 43, 441-464.

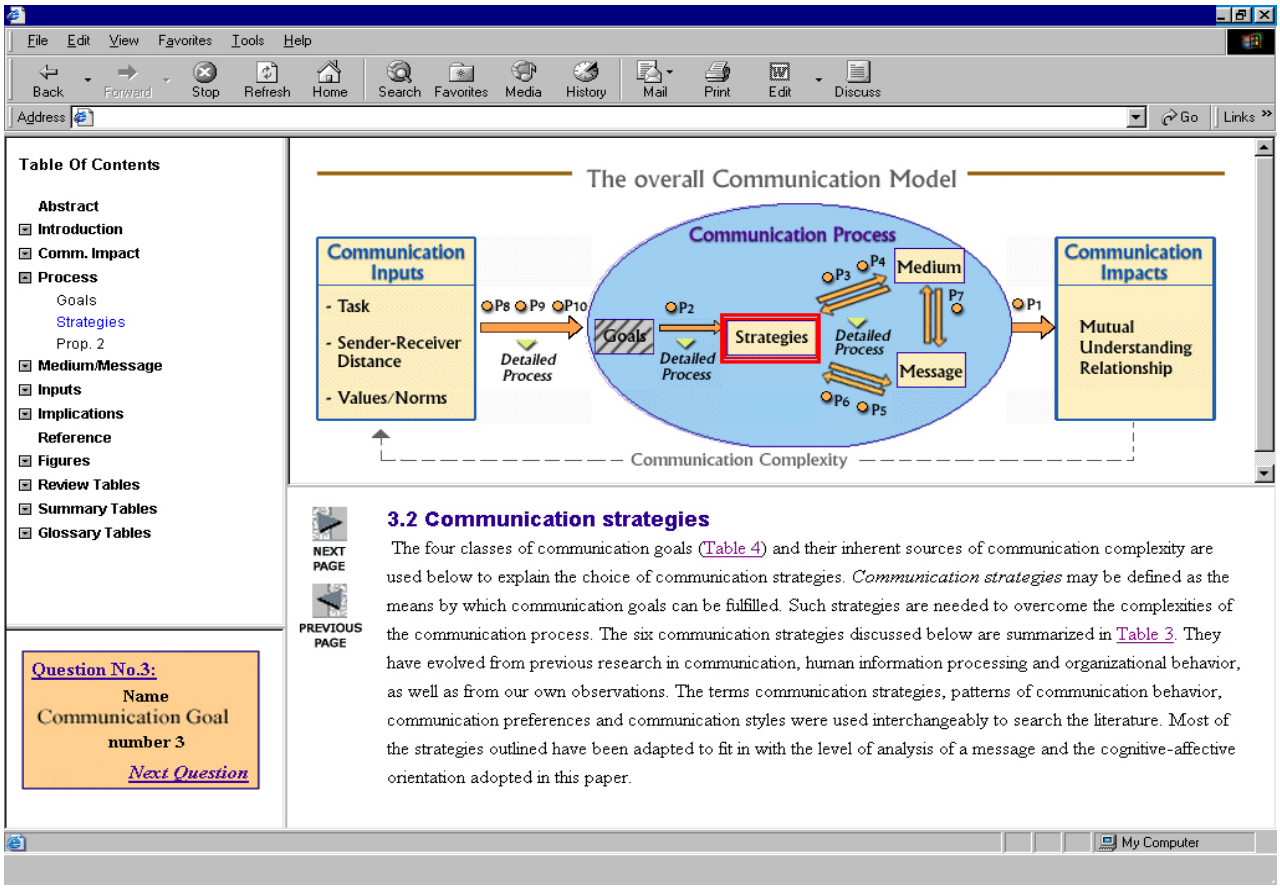
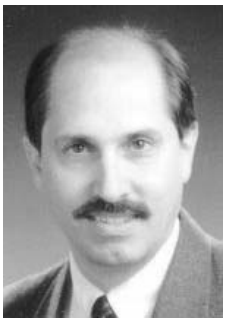


Figure 1. User Interface of the dynamic map with the WFB and history in gray – (Gray areas are shaded with diagonals here, for lack of color-print version)

INTRODUCING THE PANELISTS

Izak Benbasat is CANADA Research Chair in Information Technology Management at the Sauder School of Business, University of British Columbia, Vancouver, Canada. He received his Ph.D. (1974) in Management Information Systems from the University of Minnesota. Professor Benbasat is the past editor-in-chief of *Information Systems Research* and currently is a Senior Editor of the *Journal of the Association for Information Systems*. His research interests include evaluating user interfaces and web-based recommendation agents to facilitate business-to-consumer electronic commerce.

Mary Czerwinski is a Senior Researcher and Manager of the Visualization and Interaction Research group at Microsoft Research. The group is responsible for studying and designing advanced technology and interaction techniques that leverage human capabilities across a wide variety of input and output channels. Mary's primary research areas include spatial cognition, information visualization and task switching. Mary has been an affiliate assistant professor at the Department of Psychology, University of Washington since 1996. She has also held positions at Compaq Computer Corporation, Rice University, Lockheed Engineering and Sciences Corporation, and Bell Communications Research. She received a Ph.D. in cognitive psychology from Indiana University in Bloomington. Mary is active in the field of Human-Computer Interaction, publishing and participating in a wide number of conferences, professional venues and journals. More information about Dr. Czerwinski can be found at http://research.microsoft.com/users/marycz.



Dennis Galletta is an Associate Professor at the Katz Graduate School of Business, University of Pittsburgh. His articles have appeared in journals such as Management Science, Information Systems Research, Decision Sciences, Communications of the ACM, Journal of MIS, Data Base, Information & Management, and Accounting, Management and Information Technologies. He has served on several editorial boards, including MIS Quarterly, Data Base, Information Systems and eBusiness Management, and Cycle Time Research. He is an AIS Fellow, and served as AIS VP of Member Services, ICIS Treasurer, Inaugural AMCIS Conference Chair, AIS Council Americas representative, and AMCIS 2003 program co-chair.

Jonathan Lazar is a faculty member in the Department of Computer and Information Sciences at Towson University. He is the author of the book "User-Centered Web Development," and regularly presents and publishes papers on the topic of web usability. He is also the 2002 winner of the "Excellence in Teaching" award, and the 2003 winner of the "Business and Community Outreach" award, both in the College of Science and Mathematics. At Towson University, Dr. Lazar serves as director of the undergraduate program in computer information systems, and also is the director of the Universal Usability Laboratory.

Judith S. Olson is the Richard W. Pew Professor of Human Computer Interaction at the University of Michigan. She is Professor in the Computer and Information Systems department of the Business School, Professor in the School of Information, and Professor of Psychology. Her research focuses on how groups both get their work done and how they feel about each other (e.g. trust, social capital) when they communicate over various digital media, like email, IM, video, audio, and shared workspaces. She leads a group of researchers working in the laboratory, doing empirical studies of these phenomena using organizational simulations. She is also doing work in the field, watching real groups in corporations and in science struggle with working together at a distance. Judy has published widely in the CHI and CSCW conferences and the major journals both in HCI and Psychology. She led various conference committees and serves on the editorial boards of several major journals. She has served on a number of national panels for the National Science Foundation and the National Research Council. In 2001 she was appointed to the CHI Academy in the first cohort group.

INTRODUCING THE PANELISTS (CONT'D)

Julie Ratner is User Experience Architect at InfoSpace Mobile after HCI positions at Sandia National Laboratories in Albuquerque New Mexico and The MathWorks in Natick, Massachusetts. Working on wireless data applications for mobile phones is quite a change from her Ph.D. research in Instructional Technology and her publications on discount usability methods. For a complete list of research publications, visit http://www.iterativedesign.com/articles.htm. She has presented papers and a tutorial at UPA, CHI, ASEE, Human Factors and the Web Conferences, and Leading Edge Training Technologies.

Radhika Santhanam is a Gatton Research faculty member in the Gatton School of Business and Economics, at the University of Kentucky. Using a variety of research methods, she has investigated ways by which user training and better system design can improve users' understanding and utilization of information systems. From a macro perspective, her research has proposed methods that can help organizations monitor their investments in information systems. These research studies have been published in various journals, including *Information Systems Research*, *MIS Quarterly, Journal of Management Information Systems, Decision Support Systems*, *Information and Organization*, and *European Journal of Information Systems,* among others.

Dov Te'eni is on the faculty of management at Tel-Aviv University, Israel. He is also the chairman of Israel's Higher Education E-learning center. Dov studies several related areas of information systems in the organizational context: human-computer interaction, computer support for decision-making and for communication, and systems design. In addition, he is interested in information systems for non-profit organizations. His research usually combines model building, laboratory and field experiments and the development of prototype systems such as Spider and kMail. He is currently working on the design of adaptive Web Articles and the use of information technology for supporting communication within multi-national corporations. For a full list of publications see http://faculty.biu.ac.il/~teenid/publications.htm

Peter Todd is the Chesapeake and Potomac Professor of Commerce and Associate Dean for Graduate Programs in the McIntire School of Commerce at the University of Virginia. He received his Ph.D. in MIS from the University of British Columbia. His research interests relate to the role of information technology in decision making, the adoption and diffusion of IT innovations and the management of the IT function. He has published in a variety of journals including *Information Systems Research* and *MIS Quarterly.* He currently serves as a Senior Editor for the *MIS Quarterly*.

Marilyn Tremaine is professor in Information Systems department at NJIT. Prior to NJIT, she was professor of Computer and Information Systems at Drexel University and Professor of Computer Science at the University of Toronto. She has also worked as Vice President of Research and Development for a microcomputer software company and Senior Research Scientist at the EDS Center for Applied Research. Her current research focuses on the design and development of multimodal interfaces to be used in collaborative settings. Dr. Tremaine has been active in the ACM SIGCHI community serving as VC of Communications, Chair of the Advisory Council, VC of Finance and currently, SIGCHI Chair. She also chaired CHI'86, CSCW'92, and ASSETS'2000. She serves on six technical journal editorial boards and four-to five conference program committees annually.

Jane Webster received her PhD from NYU and currently is a professor in the School of Business at Queen's University in Canada. She investigates the impacts of technologies to support distributed work, organizational communication, employee recruitment and selection, employee monitoring, and training and learning. She also focuses on human-computer interaction issues as they relate to organizational systems. She serves as a senior editor for MIS Quarterly, and has published in a variety of journals including the *Academy of Management Journal, Communication Research, Journal of Organizational Behavior, MIS Quarterly, and Organization Science*.

INTRODUCING THE AUTHORS OF PEER-REVIEWED ARTICLES

Catherine **Arnott Smith** (MA, MILS, MSIS, PhD) is Assistant Professor at Syracuse University's School of Information Studies. She was a National Library of Medicine predoctoral medical informatics trainee at the University of Pittsburgh's Center for Biomedical Informatics from 1997 to 2002. Her research interests include consumer health, terminologies in medicine and alternative medicine, and problems of knowledge representation in all three domains.

Brian M. **Ashford** is CIO at the U.S. Army Logistics Management College, Fort Lee, Virginia, and Assistant Professor in the School of Continuing Studies at the University of Richmond. Ashford's specialties are strategic systems planning, management of information systems, and the development and deployment of collaborative technology. He received the Ph.D. in information systems from Virginia Commonwealth University. He has been a registered professional engineer since 1988.

Brent **Auernheimer**, Interim Director, Digital Campus, and professor of computer science, CSU Fresno, received his PhD from UC Santa Barbara; was co-founder/chief scientist, Oak Grove Systems; has held visiting appointments at the Software Engineering Institute (SEI), NASA Jet Propulsion Laboratory, and the University of Hawaii; has been a summer research fellow at NASA Kennedy Space Center. His professional interests are software engineering and human-computer interaction.

Paul L. **Bowen**, PhD, CPA, is a Senior Lecturer in the UQ Business School at The University of Queensland. His primary research activities involve the measurement, control, and improvement of information quality. This research encompasses investigating strategies for improving end-user information retrieval performance. Paul's previous experience includes eight years as a systems analyst and project manager at the Oak Ridge National Laboratory.

Jacek **Brzezinski** is an assistant professor at DePaul University, School of CTI. His research interests include artificial intelligence, natural language processing and e-commerce. His doctoral dissertation focused on machine learning-based modeling of natural language information for information classification purposes. Currently, he is teaching graduate courses in E-commerce.

Hock Chuan **Chan** is an associate professor at the Department of Information Systems, National University of Singapore. He obtained his B.A. and M.A. from the University of Cambridge, and Ph.D. in management information systems from the University of British Columbia. His research interests include user-database interactions and information systems.

Susy **Chan** is an associate professor and directs the Center for E-Commerce Research in the School of CTI at DePaul University. She is the founding director of DePaul's pioneering master's and baccalaureate programs in e-commerce technology. The Mobile Commerce Research Lab that she co-leads with Xiaowen Fang studies usability of wireless applications. Her research also focuses on enterprise systems for e-collaboration.

Andrea **Everard** is an Assistant Professor of Management Information Systems at the Lerner College of Business and Economics at the University of Delaware. She received her PhD in MIS from the University of Pittsburgh. Her research interests are human-computer interaction, electronic commerce, trust, ethical issues, and strategic uses of IT.

Xiaowen **Fang** is an assistant professor in School of Computer Science, Telecommunications and Information Systems (CTI) at DePaul University. His research focuses on user-centered design of Web search tools, usability of e-business, usability of mobile commerce, and multi-modal information presentation.

Mark A. **Fuller** (Ph.D., Univ. of Arizona) is Associate Professor and Department Chair of MIS at Washington State University. Dr. Fuller's research interests include virtual teamwork and social influence in computer mediated communications. Past research has appeared in journals such as Information Systems Research, Group Decision and Negotiation, and Decision Support Systems.

Andrew **Gemino** is Assistant Professor of MIS in the Faculty of Business Administration at Simon Fraser University, Vancouver, Canada. He received his Ph.D. from the University of British Columbia. His research interests include evaluating the ability of system analysis techniques to communicate knowledge and the use of technology to support collaborative work environments.

Dennis F. **Galletta**, Associate Professor at the Katz Business School, University of Pittsburgh, has published in journals such as Management Science, ISR, Decision Sciences, CACM, JMIS, and Database. He served on MIS Quarterly's editorial board and several others, is an AIS Fellow, chaired the first AMCIS conference, and was co-program chair of AMCIS 2003.

Traci J. **Hess** is an Assistant Professor of Management Information Systems at Washington State University and received her Ph.D. from Virginia Tech. Her research interests include social presence in human-computer interaction and user evaluations of IS implementations. She has published her research in such journals as Decision Sciences, Decision Support Systems, and Information Resources Management Journal.

INTRODUCING THE AUTHORS OF PEER-REVIEWED ARTICLES (CONT'D)

George M. **Kasper** is Professor of IS at Virginia Commonwealth University. He is the author of numerous articles. *MIT Press* and *Sage Publications* have reprinted his work, and professional groups including SIM have recognized it. Kasper has served as Chairman of ACM's SIGMIS, represents USA ACM to IFIP, and was recently appointed AIS Alternate to CSAB, the lead society within ABET for accrediting CS, IS, and computer engineering programs.

Hoyoung **Kim** is an Assistant Manager of the corporate planning team in Widerthan.com, where he develops the corporate strategy and evaluates mobile Internet service. He received his master degree in MIS and bachelor of the business administration from the Yonsei University. He is interested in post-adoption of the mobile internet and its usability.

Jinwoo **Kim** is Professor of HCI in the School of Business at Yonsei University. He is also working as Director of Human Computer Interaction Lab at Yonsei University. His research interests include navigation structures for interactive systems, cross-cultural value of IT with its impact on usability, and evaluation metrics and framework for digital contents.

Adrienne Olnick **Kutzschan** is 2nd year PhD student in MIS at Queen's University, Kingston, Ontario. Her primary research interests are in the area of human computer interaction and online collaboration. Before resuming studies, she completed an MBA at Simon Fraser University and consulted in IT where she observed first-hand the challenges people face when using technology.

Liping **Liu** is Associate Professor of MIS at University of Akron. His research interests include Electronic Business, Object-Oriented Systems Design, and Uncertainty Reasoning in Artificial Intelligence. His articles have appeared in Decision Support Systems, European Journal of Operational Research, Journal of Risk and Uncertainty, etc. He has served as a guest editor for International Journal of Intelligent Systems and a co-editor for Classic Works on Theory of Evidence.

Paul Benjamin **Lowry** is Assistant Professor of IS in Brigham Young University, and faculty researcher at the Kevin and Deborah Rollins Center for e-Business. His interests include Internet-based collaboration, virtual teams, distributed facilitation, collaborative software, e-Business, group-based HCI, heuristics, and scientometrics. He received Ph.D. in MIS from the University of Arizona.

Ying **Lu** is a doctoral student in Management Information Systems at the University of Wisconsin-Milwaukee. She received her M.S. from the Pennsylvania State University. Her research interests are in the areas of HCI in e-commerce, computer-mediated communications, web personalization, and end-user training and adoption of information systems.

John **Mathew** is a doctoral student in Management Information Systems at Washington State University. His research interests include human computer interaction and knowledge management.

Fiona Fui-Hoon **Nah** is Assistant Professor at University of Nebraska-Lincoln. She received her Ph.D. in MIS from University of British Columbia. Her research interests include the use of knowledge-based systems to support group decision-making, HCI in individual and group settings, and ERP. She serves on the editorial board of six major journals and is a co-founder of AIS SIGHCI.

Drew **Parker** is Associate Professor of MIS in the Faculty of Business Administration at Simon Fraser University, Vancouver, Canada. He received his Ph.D. from the University of Western Ontario. He has offered completely online courses since the mid-1990's and has effectively combined research and practice in the area of learning in distributed knowledge sharing environments.

Peter **Polak** is an assistant professor of computer information systems in Business School at the University of Miami. His research interests are in the areas of human-computer interaction, Internet technologies, electronic commerce, electronic auctions, and telecommunications.

Tom **Roberts** is an Assistant Professor at the School of Business at the University of Kansas. He received his doctorate from Auburn University. He previously taught at the University of Central Florida. His research interests include virtual teams, collaborative software, group support systems, and the behavioral aspects of the information technology profession.

Fiona H. **Rohde**, PhD, is a Senior Lecturer in Business School at The University of Queensland. Her primary research activities involve the area of outsourcing, and also the area of information management and its effect of information retrieval. Fiona worked for KPMG in the Computer Audit Division before joining the school approximately 10 years ago.

Hadar **Ronen** received MA in Information Studies at Bar-Ilan University (2003). She is Certified in Museology studies Haifa University (1998) and has a BA in Creative Arts and general studies from Haifa University (1992). Hadar worked as a graphic designer and currently is an information specialist in a company developing Knowledge Management platform and Business Intelligence tools. Her interests include User Interface design of web sites and e-books, and digital libraries.

INTRODUCING THE AUTHORS OF PEER-REVIEWED ARTICLES (CONT'D)

Hong **Sheng** is a MIS Ph.D. student at University of Nebraska-Lincoln. She received her master degree in MIS from Univeristy of Nebraska-Lincoln and her bachelor degree from Shanghai Jiaotong University, China. Her research interests include mobile commerce, strategic implications of mobile applications in organizations, and HCI.

Keng **Siau** is Associate Professor of MIS at University of Nebraska-Lincoln. He is currently serving as Editor-in-Chief of Journal of Database Management and the Book Series Editor for Advanced Topic in Database Research. His research interests include Systems Analysis & Design, mobile commerce, and HCI. Dr. Siau has published over 150 articles, which include more than 55 refereed journal publications.

Felix B **Tan** is Professor of Information Systems at Auckland University of Technology, New Zealand. He serves as the Editor-in-Chief of the Journal of Global Information Management. His current research interests are in business-IT alignment, e-commerce, global IT, management of IT and cognitive mapping/narrative methods. He has published in MIS Quarterly, Journal of Global Information Management, Journal of Information Technology and Information & Management.

Dov **Te'eni** is on the faculty of management at Tel-Aviv University, Israel. He is also the chairman of Israel's Higher Education E-learning center. Dov studies several related areas of information systems in the organizational context: human-computer interaction, computer support for decision-making and for communication, and systems design. In addition, he is interested in information systems for non-profit organizations.

Tung Lai Lai is Associate Professor and Associate Director for the Information Management Research Centre at the Nanyang Business School in Singapore. She researches E-Commerce, GSS and Virtual Communities. She is an Associate Editor for Electronic Commerce Research and Applications, Journal of Electronic Commerce in Organizations and e-Service journal. She has also served as a consultant for both the private and public sectors in Singapore.

Vance **Wilson** is an Assistant Professor of Management Information Systems at the University of Wisconsin-Milwaukee. His research focuses on organizational aspects of human-computer interaction, including study of computer-mediated communication in healthcare, negotiation, and software-development settings.

Rita **Vick** recently completed her PhD in Communication and Information Sciences at the Univ. of Hawaii. Since obtaining her MBA, she has consulted, owned a business, taught graduate and undergraduate courses, published conference papers, journal articles, is currently working on two book chapters, and is seeking a permanent position in industry or academia. Her research interests include distributed cognition, virtual team decision-making, and human-computer interaction.

Chiu Yueh Alice **Wu** has recently completed a Masters Degree specializing in the area of information systems. Prior to undertaking her Masters she completed a Commerce degree. Alice is currently gaining valuable work experience before considering returning to pursue an academic career.

Lian **Xiang** is a graduate student at the Department of Information Systems, National University of Singapore. She received her B.A in information management and information systems from Wuhan University, China. Her research interests include query languages, and user-database interactions.

Shuang **Xu** is a doctoral student in School of CTI at DePaul University. She received her Bachelor degree in CS from Zhejiang University, and her Master degree in CS from Nanjing University, China. Her research concentration is on Human-Computer Interaction, with a keen interest in usability issues on handheld devices. Her dissertation is focusing on dual-modal information presentation.

NOTES

INDEX OF AUTHORS