Mutual Affects in Computer-mediated Collaborative Learning: Positive Feelings Shared by Collaborators Enhance System Evaluations

Takashi Yamauchi,
Texas A&M University
College Station, TX
takashi-yamauchi@tamu.edu

Takehiko Ohno, Momoko Nakatani, Yoichi Kato,
NTT Cyber Solutions Laboratories
Yokosuka-shi, Japan
ohno.takehiko@lab.ntt.co.jp

Arthur B. Markman
University of Texas
Austin, TX
markman@psy.utexas.edu

Abstract

The authors employ behavioral theories of human motivation and affect and present an explanation for why some computer-mediated collaborative learning is satisfying for a user. In a longitudinal experiment, participants were divided into four groups and solved two open-ended problems together using a video-conference system. Traditional metrics of usability and product acceptance were examined with respect to psychological variables such as personality, background knowledge, and feelings toward group members (mutual affect). The results show that group-level mutual affect is a strong predictor of system acceptability judgments, even after controlling for other pragmatic variables such as opinion convergence. It is proposed that evaluating one's experience with a computer-mediated collaborative system is a sensemaking process and that the variables that modulate this process also influence subjective judgments of usability and acceptability of a system.

Keywords: User satisfaction, user experience, mutual affect

Cultivating positive emotions among collaborators is essential for the success of groupware applications because shared positive affects promote group coordination, common ground, and group awareness—key ingredients for successful online collaboration (Carroll et al., 2006). But what design features are critical to generate positive mutual affects? Do mutual affects influence user experience primarily by elevating pragmatic qualities of group interaction, such as group communication and coordination?

To help improve computer-mediated collaborative learning (e.g., learning collaboratively via video conferencing), researchers have identified important variables, such as group awareness, common ground, shared visual information and teamwork coordination (Carroll et al., 2006). However, to make a "good" collaborative learning system, these pragmatic variables should be supplanted; the product should be not only useful but also *engaging* and *satisfying* for the users (Hassenzahl & Tractinsky, 2006; Norman, 2004). But to make an engaging and satisfying product, it is crucial to know how users come to evaluate their experiences with a collaborative system.

Conceptual frameworks such as information processing, affordance, and cognitive architecture have generated testable hypotheses and guidelines instrumental for single-user product design. However, these pragmatic variables are not entirely feasible in collaborative settings because of added complexity inherent in group interaction (Grudin, 1994).

Here, we propose a conjecture that psychological models of sensemaking can provide a useful framework for user experience in a groupware setting much in the same way that affordance and cognitive architecture helped the evolution of single-user interfaces. We argue that the evaluation of one's experience is basically a sensemaking process, and the variables that intervene this process influence "user experience."

To test our framework, we developed an experimental study, where 29 college students were divided into four groups and solved two problems together in a 2-month period using a video-conference system. We examined subjective metrics of usability and product acceptance with respect to other psychological variables such as personality, background knowledge and group-coherence. The results showed that a positive mutual affect among group members led to increased product acceptance, even after controlling for other pragmatic variables such as opinion convergence and communication effectiveness.

User Experience as a Sensemaking Process

Klein et al. (2006) and Pirolli and Card (2005) provide models of sensemaking. The two models differ in specifics but share some basic properties. Sensemaking consists of dynamic processes of data selection and frame/schema revision. Relevant data are selected according to one's frame (prior knowledge/beliefs/mindsets), the data are interpreted and the frame is revised according to the interpretation. Sensemaking goes through cycles of this data selection/interpretation and frame/schema revision loop. Our central hypothesis is that user experience is a sensemaking process. "Experience" does not come to people unambiguously. Experience is selected, sensed, represented, and interpreted by people (Pirolli & Card, 2005). In this process, affects play critical roles as affect seeps into the evaluation of the data.

Group-level Mutual Affect The importance of affect in product design is well known, but affect in human-computer interaction has pertained to a specific product. We think that group-level mutual affect (e.g., feeling of closeness of group members) can also be an important factor because affects are contagious and affects coming from unrelated sources can be easily fused into the evaluation of a product.

Much research has shown that relatively simple manipulations of inducing a positive affect, such as viewing a comedy film for a few minutes or writing about happy events, influence subsequent decision making of unrelated objects (Clore & Huntsinger, 2007). Schwarz and Clore (1983) present one of the most stunning demonstrations of affect contamination. In their experiment, the researchers interviewed subjects about their general happiness with their lives. Subjects were selected randomly and telephone-interviews were conducted on either a sunny day or a rainy day. Those who had an interview on a sunny day gave higher happiness ratings than those who had an interview on a rainy day. When the link between mood and weather was made clear to subjects, the ratings made on the rainy day went up, indicating that subjects' ratings about happiness were partly due to their erroneous generalization of their unhappy mood on the rainy day.

A similar misattribution is likely to happen in the judgment of usability and acceptability. Usability and acceptability of a product will be judged by pragmatic, hedonic, and aesthetic features of the product (Hassenzahl & Tractinsky, 2006). However, in making an actual judgment, a user will *interpret* his/her memories of experience. In this process, affective experience with group members can contaminate their evaluation (Clore & Huntsinger, 2007).

In the experiment described below, we examined the extent to which mutual affects formed among collaborators influence their usability and acceptability judgments of a video-conference system.

Experiment

In our 2-month-long experimental study, four groups of participants (seven to eight participants per group) met eight times using MeetingPlaza, a multi-party Web conferencing and collaboration system (http://www.meetingplaza.com); each group worked together to solve two different open-ended problems (i.e., how to improve the university and recommendations for freshman job search) [15], and wrote two one-page white paper proposals together as a group using MeetingPlaza.

MeetingPlaza has web-, file-, whiteboard- and application-sharing functions that facilitate collaborative communication. For example, the web-share function allows participants in different locations to view the same web site on their own computers at the same time. The file-sharing and application-sharing functions help people in remote locations to view and manipulate the same file together (e.g., an MS Word file). The participants were encouraged to write papers together using these sharing functions.

Hypothesis. On the basis of the theoretical background discussed previously, we formed the following hypothesis. Group-level mutual affect (e.g., feelings of closeness toward group members) influences subjective judgments of system usability and acceptability. In particular, those who have high positive group-level affect should give high acceptability and usability ratings even when other group-

level variables such as opinion convergence and communication effectiveness are controlled for.

Figure 1. MeetingPlaza.

The rationales for this hypothesis are as follows. Because assessing one's experience with a computer system is a sensemaking process, group-level mutual affect can be easily fused into the evaluation of the conference system, as a user makes a system evaluation based on his/her memory of the experience with a product. Thus, positive mutual affect will be translated into positive product evaluation.

Participants. Thirty-two participants were recruited from the Texas A&M University community. They were assigned randomly to four groups. Three participants chose not to take part in the experiment after the first meeting. Thus, a total of 29 participants completed the two problem solving sessions (Table 1). Participants received a payment of $144 ($12 per hour for a total of 12 hours for their involvement). Bonus payments of $24~$48 were made to group members who produced the best and second-best white papers. In a separate experiment, 47 undergraduate students were recruited from the university psychology subject pool for the evaluation of the white papers submitted by the four groups.

Table 1. Participant information

N=	29	Major: psychology = 16; public health = 2; political science, history, general studies, telecommunication, management, accounting, industrial engineering, electrical engineering, chemical engineering, nursing, nutrition = 1
(Male, Female)	(14, 15)	
Freshman	1	
Junior	5	
Sophomore	6	
Senior	12	
Graduate student	4	
Staff	1	
Average age	21.1	
Note. Three participants dropped after the first meeting and the data from 29 participants were analyzed.		

Note. The participants received a payment of $144. Bonus payments of $24~$48 were made to group members who produced the best and second-best white papers.

Materials. We employed five questionnaires to assess participants' mindsets (implicit beliefs on intelligence, morality and world), personality (neuroticism, extraversion and psychoticism), technological literacy (computer-literacy and Internet-literacy), and expectations (expected ease of use and expected usefulness of the product). These questionnaires, which were given at the orientation meeting, were adopted to isolate the effect of group-level mutual affect as much as possible.
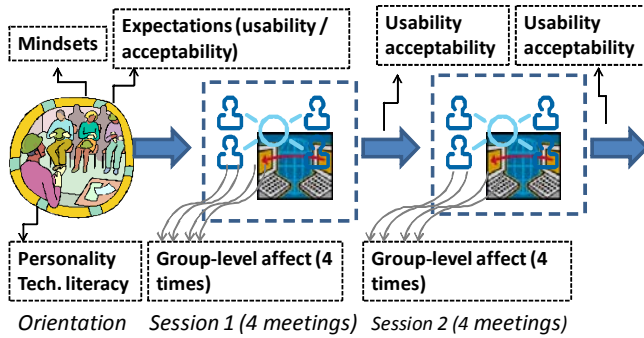
| Mindsets | Expectations (usability / acceptability) | | Usability acceptability | Usability acceptability |

Figure 2.The logistics of the experiment.

Group-level coherence (affect, communication and opinion) was measured by electronic questionnaires given at the end of each meeting. Participants' subjective judgments of system usability and acceptability were collected three times in a two-month period, before using the system (at the orientation meeting), in the middle of using the system (after the fourth meeting), and at the end of the experiment (after the eighth meeting) (Figure 2). Participants' subjective judgments of system usability and acceptability were collected three times in a two-month period, before using the system (at the orientation meeting), in the middle of using the system (after the fourth meeting), and at the end of the experiment (after the eighth meeting) (Figure 2).

Below we explain the questionnaires used in the experiment.

Implicit belief. The implicit belief questionnaire assesses the extent to which people conceptualize intelligence, morality, or the world as a dynamic or fixed construct (Dweck, 1999). This questionnaire was included because people's implicit beliefs are known to influence their goal setting and learning experience.

Personality. Francis et al. (1992) developed an abbreviated version of the Eysenck personality questionnaire (EPQR), which has four dimensions (extraversion, neuroticism, psychoticism, and lie scale) with six questions each. The questionnaire assesses personality of a person with three dimensions, extraversion (high-low tendency to seek external stimulation), neuroticism (high-low level of negative affect), and psychoticism (high-low level of impulsiveness). Following the suggestion by Francis et al. (1992) we did not analyze lie-scale scores in the present experiment.

Technology literacy. The technology literacy questionnaire was developed for this experiment based on the digital literacy questionnaire by Hargittai (2009). Our questionnaire consists of two categories, computer literacy and Internet literacy, and a total of 10 questions. The computer literacy measure has four items related to knowledge about software (e.g., PowerPoint) and programming language (e.g., Java). Internet literacy consists of six items related to common Internet-based activities (e.g., tweeting or on-line shopping).

Expected ease of use and expected usefulness of the product. We modified Davis's Technology Acceptance Model (TAM) (Davis, 1989) and developed questionnaires assessing expected ease of use and expected usefulness of the product (six questions for each). We included the term "expected" because our questionnaires were given shortly after MeetingPlaza was introduced to the participants but before they actually used the system.

Usability. We employed Lewis's Computer System Usability Questionnaire (CSUQ; 19 questions) (Lewis, 1995). The CSUQ consists of three factors, system usefulness, information quality, and interface quality. The pre-usability questionnaire was given at the orientation meeting shortly after participants were introduced to the system but before using the system. The post-usability questionnaire was given twice at the end of session 1 and at the very end of the experiment.

Acceptability. To measure participants' behavioral intention of adopting the video-conference system, we created 10 acceptability questions based on Davis et al., (1989) and Venkatesh and Morris (2003). These questions assessed participants' intention to continue to use MeetingPlaza if the system were made available to them.

Group-level coherence. We measured group-level coherence of individual members with three dimensions, mutual affect (e.g., how close do you feel with each member of your group?), opinion convergence (e.g., how close was your opinion with that of each member of your group?), and communication effectiveness (e.g., how effectively did you communicate with each member of your group?). Every participant rated how he/she felt about each group member at the end of every group meeting (a total of eight meetings), the ratings he/she gave to all group members were averaged over affect, communication effectiveness and opinion convergence dimensions, and these average values were treated as his/her group-level affect, communication effectiveness and opinion convergence (Strauss, 1997).

Procedure. The experiment consisted of four segments: orientation, session 1, session 2, and paper evaluation. Below, we describe the segments in chronological order (Figure 3).

Orientation. The orientation meeting was held in a large classroom. First, participants indicated their implicit beliefs, personality and technology literacy, and then the experimenter introduced MeetingPlaza. At this stage, participants were allowed to view MeetingPlaza, but they were not allowed to use the system. After this brief instruction, participants indicated their expected ease of use and expected usefulness of MeetingPlaza, along with their expected usability of the product (pre-usability) and their intention of using the product in the future (acceptability).

Sessions 1 & 2. Approximately 1 week after the orientation meeting, participants were assigned to four groups, and each group had its first on-line meeting using MeetingPlaza. In this segment, participants received extended instruction and demonstrations of MeetingPlaza functions and tested MeetingPlaza by themselves. Each group met twice a week, and discussed solutions for the assigned problems using MeetingPlaza. In one session,

participants as a group were required to write a one-page white paper describing ways to improve the university; in the other session participants as a group were required to write another white paper describing recommendations for job search for college freshmen. Each group was required to submit a paper at the end of the fourth meeting of each session. Each meeting lasted about 1 hour.
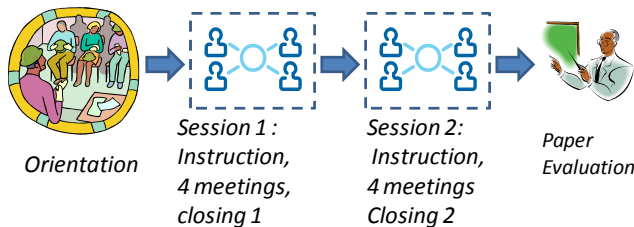

Figure 3. Four segments of the experiment

Session 2 was given 1 week after the end of session 1. The procedure of session 2 was identical to that described in session 1.
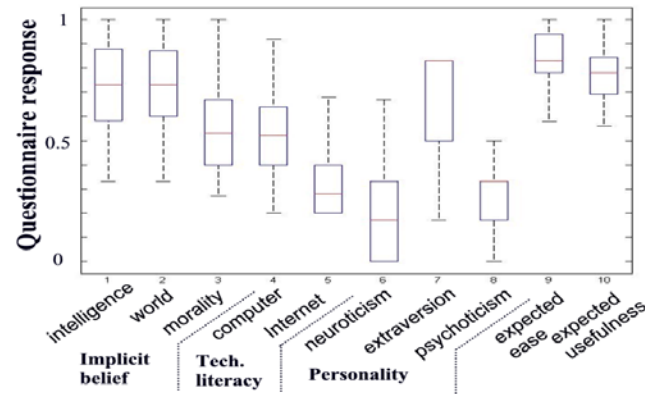
Closings 1 & 2. Two closing meetings, closing 1 and closing 2, were held at the end of sessions 1 and 2, respectively. In closings 1 and 2, participants filled out the usability and product acceptance questionnaires. Closing 2 was the final meeting.

Paper evaluation. In a separate experiment, 47 undergraduate students participated in the paper evaluation experiment (male=23, female=23, unknown=1) and rated the eight papers written by the four groups in six categories (creativity, implementation, coherence, effectiveness, cost, and communication) on a 0–100 scale. They were encouraged to rate the papers in the same way a professor grades their papers in a classroom.

Results

All questionnaire responses were converted to a 0-1 scale such that the direction of the observed scores corresponded to the direction of the psychological dimensions in question. Thus, a high score corresponded to a high degree of the given dimension. This section begins with a summary of questionnaire responses, followed by a description of the longitudinal shifts of usability, product acceptability and group-coherence, and concludes with statistical analyses that examine the relationship between group-coherence and system evaluation.

User profiles. The responses on the 10 dimensions of the questionnaires (Figure 4) show that there were no ceiling or floor effects, except for the responses regarding expected ease of use and expected usefulness. This problem will be discussed in the next paragraph and later in the Results section. ANOVAs (analysis of variance) comparing the four groups in each of the ten user profile dimensions showed that the mean profile scores of the four groups were not statistically different: $F\text{'s}(3, 25) < 2.2$, $p\text{'s} > 0.11$.

Note. The central mark and the edges of a box are the median and the 25th and 75th percentiles, respectively. The whiskers are the most extreme data points.
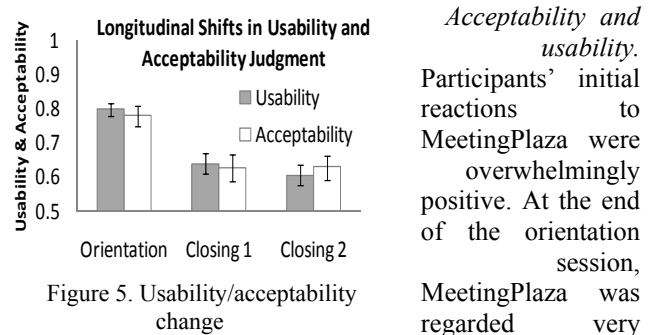
Figure 4. Boxplots for questionnaire responses.

Figure 5. Usability/acceptability change

Acceptability and usability. Participants' initial reactions to MeetingPlaza were overwhelmingly positive. At the end of the orientation session, MeetingPlaza was regarded very favorably (usability, M=0.80, SD=0.11; acceptability, M=0.78, SD=0.17; Figure 5). However, the ratings of MeetingPlaza dropped significantly at closing 1 (usability, M=0.64, SD=0.64; acceptability, M=0.63, SD=0.21) and closing 2 (usability, M=0.61, SD=0.18; acceptability, M=0.63, SD=0.22). Two 3 (sequence: orientation, closing 1, closing 2) x 4 (group: 1, 2, 3, 4) ANOVAs revealed that this drop occurred uniformly in all groups: usability, $F(2, 50)=18.61$, MSE=0.02, p<0.01; acceptability, $F(2, 50)=9.70$, MSE=0.02, p<0.01. Neither the main effect of group nor the interaction between sequence and group was observed in both usability and acceptability measures: $F\text{'s} < 1.5$, p>0.24. These results suggest that MeetingPlaza created a positive impression on the college-age participants but the excitement dropped considerably once the participants used the product for problem solving, indicating that using the collaborative video-conferencing system was much more challenging than anticipated.

Longitudinal shifts of group coherence. The group-coherence scores (mutual affect, opinion convergence, communication effectiveness) all increased as the collaborative sessions progressed (Figure 6). Three sets of linear contrast analyses (shift; beginning, middle, end of the experimental sessions) x (group; 1-4) applied to the three group-coherence measures revealed significant linear upward trends: communication effectiveness, $F(1, 25)=7.72$, MSE=.01, p<.05; opinion convergence, $F(1, 25)=26.2$, MSE=.01, p<.001; mutual affect, $F(1, 25)=35.8$,

MSE=.01, p<.001, suggesting that our online meetings were indeed effective in developing a sense of common ground, better communication, and positive feelings. There was no interaction effect of group and shift: F's<1.0.
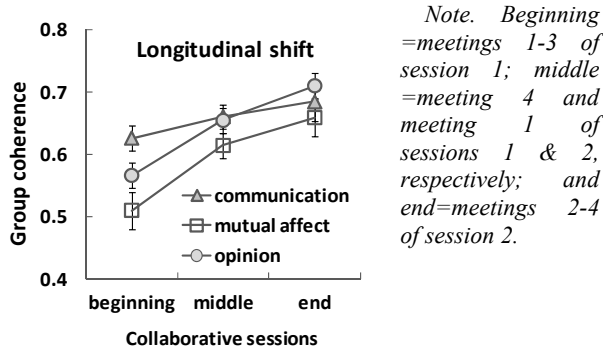

Note. Beginning =meetings 1-3 of session 1; middle =meeting 4 and meeting 1 of sessions 1 & 2, respectively; and end=meetings 2-4 of session 2.

Figure 6. Longitudinal shifts of group coherence

Evaluation of the hypothesis. Regression analyses were employed to investigate the link between group-level affect and system evaluation. Both step-wise regression and regular multiple regression were adopted to ameliorate the problem of multicollinearity. In the step-wise regression procedure, the forward selection method was applied with the entry criterion of 0.1 to ensure that all relevant predictors were included in the regression models.

The step-wise regression analyses using group coherence variables suggest that mutual affects influenced system acceptability scores significantly; β=.52, p<0.01, R^2 =.27; but other group coherence variables—opinion convergence and communication effectiveness—did not influence acceptability scores, p>.1. The impact of the mutual affect variable remained strong on the system acceptability measure even after controlling for the effects of all other personal variables [mindsets (intelligence, morality, world); personality (neuroticism, extraversion, psychoticism), technology literacy (computer-literacy, Internet-literacy)]; β =0.48, p=.05; but not the usability scores; β =0.35, p=.15.

The results from multiple regression analyses were analogous to those found in the step-wise regression analyses. Even after the communicative variables—communication effectiveness and opinion convergence—were forced into the models, the strongest predictors were still mutual affect; the correlation between mutual affect and acceptability was significant after the effects of communication effectiveness and opinion convergence were partialed out (r=0.45, p<0.05).

Cohort effects. The predictor, group-level mutual affect, was evaluated with the data obtained from individual participants. Because MeetingPlaza is a collaboration tool, the impact of this variable should be scrutinized with respect to the properties obtained from each group. For this reason, we employed hierarchical linear regression models and estimated the beta coefficients of the mutual affect variable for each group and investigated if the impact of mutual affects remain robust after controlling for other group-specific properties—the ratio of female and male

participants in each group and the white paper evaluation score that each group received (Table 2).

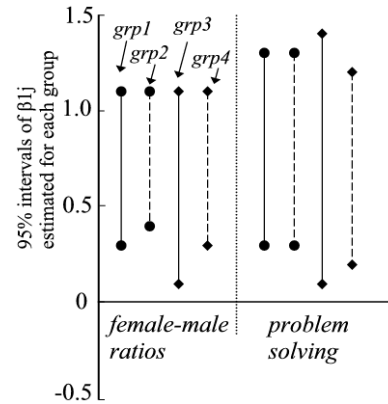Table 2. Hierarchical Linear Regression Model

Individual layer:	*Group layer:*
$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + u_{ij}$	$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$
	$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$

Note. Subscript j denotes group ID and Y_{ij} represents the acceptability score obtained from participant i of group j. β_{0j} and β_{1j} are intercepts and slopes of the regression line of group j, respectively. u's are error terms. γ_{00} is the overall mean of the acceptability scores and γ_{10} is the mean of the slopes of the four groups. W_j represents group-specific values (e.g., either the female-male ratio of group j or the problem solving score of group j) and γ_{11} is the coefficient for predictor W_i.

Note. The problem-solving performance was measured by the average paper evaluation score that each group received in a separate experiment.

Figure 7. 95% high density intervals of β_{1j} estimated for each group

Our hierarchical models had two layers, individual and group (Table 2). The model assumes that the coefficients β_{1j} of mutual affects (X_{ij}) vary for each group and are modulated by a group-specific properties (problem-solving scores or female-male ratios). Note that the "group-level mutual affect" variable X_{ij} was included in the individual layer because the values of this variable were calculated for individual participants. The values of group-specific variable W_j (e.g., problem-solving scores or female-male ratios) were calculated for each group, not for each participant. Because we had only four groups, the intervals of β_{1j} were estimated by a Bayesian method where the coefficients (γ.) in the group layer were treated as non-informative hyper-parameters and posterior distributions were obtained by the Markov Chain Monte Caro algorithm with 1000 iterations (Gelman & Hill, 2007).

The results, which are summarized in Figure 7, suggest that even after the two group-specific properties (female-male ratios and problem-solving scores) were taken into account, the impact of mutual affect remained robust, as the 95% high density intervals of the coefficients β_{1j} were above 0 in all cases, suggesting that the effect of mutual affect occurred on top of the group-specific properties.

Discussion

In computer-mediated collaborative learning, the focus has been to enhance pragmatic functionality of the system. The present study shows that fostering positive emotions among collaborators is no less important. The results suggest that mutual affects shared among collaborators influence the evaluation of product acceptability even after personal variables, such as personality and background knowledge, were taken into account, implying that the influence of mutual affect on a video-conferencing system is far reaching than previously thought. Because the effect of mutual affect was stronger than that of pragmatic variable such as group-level communication and opinion convergence, it is likely that mutual affects were fused into users' experience with the system.

Note that the fact that affective experience can be misattributed does not mean that affect is irrelevant in enhancing the functionality of a collaborative learning system. Positive emotions can unleash creative and flexible thinking (Isen, 2008) and shared feelings have a multiplicative effect on collaborators because emotions are highly contagious.

A considerable progress has been made in the area of affective computing of intelligent tutoring systems, primarily thanks to the pioneering studies by D'Mello, Graesser, and Conati (Conati & Maclaren, 2009; D'Mello et al., 2007). We suggest that similar affect detection technologies help advance group-ware applications. In a large computer-based collaborative situation, it is difficult to assess participants' affective states in real time. A collaborative groupware system that can trace users' affective state can facilitate group participation and learning.

Acknowledgments

References

Carroll, J. M., Rosson, M. B., Convertino, G., & Ganoe, C. H. (2006). Awareness and teamwork in computer-supported collaborations. *nteracting with Computers, 26*, 21-46.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*, 393-399.

Conati, C., & Maclaren, H. (2009). Modeling User Affect from Causes and Effects. In G.-J. H. et al. (Ed.), *UMAP 2009, LNCS 5535* (pp. 4-15). Berlin Heidelberg: Springer-Verlag.

D'Mello, S., Graesser, A., & Picard, R. (2007). Towards an Affect-Sensitive AutoTutor. *IEEE Intelligent Systems, 22*, 53-61.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*, 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982-1003.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality and development*. Ann Arbor, MI: Psychology Press.

Francis, L. J., Brown, L. B., & Philipchalk, R. (1992). The development of an abbreviated form of the revised Eysenck Personality Questionnaire (EPQR-A): Its use among students in England, Canada, the U.S.A. and Australia. *Personality and Individual Differences, 13*, 443-449.

Gelman, A., & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. New York: Cambridge University Press

Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communication of the ACM, 37*(1), 92-105.

Hargittai, E. (2009). An update on survey measures of web-oriented digital literacy. *Social Science Computer Review, 27*(1), 130-137.

Hassenzahl, M., & Tractinsky, N. (2006). User experience – research agenda. *Behaviour & Information Technology, 25*, 91-97.

Isen, A. M. (2008). Some ways in which positive affect influences decision making and problem solving *Handbook of emotions* (pp. 548-573). New York: Guilford Press.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88–92.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaire: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7*, 57-78.

Norman, D. A. (2004). Introduction to this special section on beauty, goodness, and usability. *Human-Computer Interaction, 19*, 311-318.

Pirolli, P., & Card, S. K. (2005, May). *The sensemaking and leverage points for analyst technology as identified through cognitive task analysis*. Paper presented at the International Conference on Intelligence Analysis, McLean, VA.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513–523.

Straus, S. G. (1997). Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. *Human-Computer Interaction, 12*, 227-266.

Venkatesh, V., & Morris, M. G. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*, 425-478.