

[College of Information Science and Technology](#)



Drexel E-Repository and Archive (iDEA)

<http://idea.library.drexel.edu/>

Drexel University Libraries

www.library.drexel.edu

The following item is made available as a courtesy to scholars by the author(s) and Drexel University Library and may contain materials and content, including computer code and tags, artwork, text, graphics, images, and illustrations (Material) which may be protected by copyright law. Unless otherwise noted, the Material is made available for non profit and educational purposes, such as research, teaching and private study. For these limited purposes, you may reproduce (print, download or make copies) the Material without prior permission. All copies must include any copyright notice originally included with the Material. **You must seek permission from the authors or copyright owners for all uses that are not allowed by fair use and other provisions of the U.S. Copyright Law.** The responsibility for making an independent legal assessment and securing any necessary permission rests with persons desiring to reproduce or use the Material.

Please direct questions to archives@drexel.edu

Information Technology Incorporating Emotion in Dialogues

Caleb Fowler, Rosina Weber

The iSchool at Drexel

College of Information Science & Technology, Drexel University

Philadelphia, PA 19104

Abstract

Human computer communication is sometimes difficult due to lack of emotion and diversification. At times when many research projects target the development of technology to support people without adult level cognitive abilities, repetitive, and emotionless communication may challenge any chances of success with the new technology. In this paper, we examine a case-based method that relies on the listener's emotional context to recommend a communication strategy that is both diverse and embedded with relevant motivational aspects.

Keywords

Enabling Technology, Human-Computer Interaction, Emotional Interfaces, Dialogue Systems

1 INTRODUCTION

Natural and comfortable conversations between humans and computers have long been a dream of many computer users. One only needs to think of HAL in "2001, A Space Odyssey" or the computer in Star Trek to see this objective reflected in icons of our popular culture. One theme of these fantasy systems is the ability of the computer to discern the emotional state of the human and to react appropriately. Normal adults may benefit from this interface design, but other user communities may find it especially useful. Children, in particular, may greatly benefit from this type of computer interface. We examine a method that recommends a communication strategy that uses the listener's emotional context as an input. We foresee deployment of this within dialogue systems designed to interact with children. We anticipate that this will benefit children by making it easier for them to interact with and utilize the computer system.

Implementing systems for children present a number of problems. In particular, we encountered two issues. First, children view the world differently from adults, partially due to differences in physical and mental development [2]. Methods that work with adults may not perform well with children [13]. Secondly, current research supports the idea of a wide range of development in normal children [19]. This implies that wide variations in behaviour are to be expected in normal children. These two issues mean that in order to design a computer method to relate to children, we needed to use a valid framework, focused on children.

Consequently, we chose the child development framework by Piaget [17]. Specifically, Piaget [17] speculated that children progress through a number of different developmental stages, each with different behavioural characteristics:

1. Sensimotor (0-2 years): gross motor function only.
2. Preoperational (2-7 years): limited attention span, motor control, and memory capacity.
3. Concrete Operational (7-11 years): close to adult cognition.
4. Formal Operational (12+ years): adult based methods suitable.

In this paper, we limit our study to children in the preoperational stage because:

- 1 Young children are used to receiving explicit task instruction,

- 2 Parents are better at assessing a child's mood than a researcher assessing a subject (a stranger),
- 3 Our assessment method is designed to assess infants and children.

Children 2 – 7 years old have great difficulty with traditional computer interface devices. Limited short term memory and short attention spans make computer interaction difficult. However, studies have shown effectiveness in increasing attention span as well as improving short term memory recall [13] by including emotional components in dialogue systems.

Furthermore, limited fine motor control in children makes mouse and keyboard use difficult and frustrating [2]. One way to mitigate this is with an effective speech recognition system. Together, a dialogue system that incorporates emotional elements into its planning algorithms would seem to offer a powerful interface for young users.

We propose to use case-based reasoning (CBR) as the underlying methodology in the recommendation of a strategy to incorporate emotion in communication. CBR is a reasoning methodology [21] that relies on previous experiences in performing a reasoning task. The system stores these previous experiences within a case base, and retrieves cases for reuse based on how similar a new problem is to the previously stored experiences. A reuse step will compare the retrieved solution to the current problem and determine how suitable the retrieved solution is for solving the new problem. When necessary, the solution is adapted to fit the new problem [1].

CBR was chosen because:

- 1 CBR is easy to implement.
- 2 It has been demonstrated as an effective methodology to recommend suitable solutions by perceiving subtle variations in the problem context.
- 3 CBR does not require a descriptive, mathematical model in order to perform.
- 4 Exemplar cases are available.

Our work is detailed as follows: Section 2 covers related works and theoretical underpinnings. In Section 3, we present our model for dynamically altering dialogue based upon emotions. Section 4 introduces our pilot study. Section 5 discusses validation aspects of our pilot study. Finally, Section 6 presents our conclusions and future work.

2 RELATED WORKS

Our goal is to change computer dialogue by utilizing emotional information from the human partner. This challenge stands at the intersection of two major fields: Human-Computer Interaction (HCI) and Natural Language Processing (NLP). We also discuss our utilization of theories to detect emotion and methods of communicating.

Historically, in the HCI community, emotional states were not considered necessary, or even desirable elements to consider in interface design [13]. Lately, there has been an awakening in research in the psychology of emotion [8] which has, in turn, fueled a tremendous increase in the literature regarding the use of emotions in computer interface design. This is in large part due to technological improvements that have made the deployment of new methods practical by which computers may express emotions (such as multimodal interfaces [3]) as well as a trend towards ubiquitous, inexpensive, networked sensors that collect information simplifying human-computer interaction [7,18].

It has been shown that affective support can increase system acceptance, possibly by producing a feeling of increased control over the device and lower cognitive load in utilizing it [14]. However, dialogue with computers has been disappointing, despite over fifty years of steady technical progress. Studies like the one conducted by Honeywell [9] illustrate the lack of a readily accessible and easy to use methodology to utilize emotion and create diversity in communications between humans and computers.

Our work is influenced by emotional and motivational models from psychology. Emotions are desirable in computer dialogue for a number of reasons. First, they capture attention [4] from the user. Second, emotionally stimulating events are generally remembered better than unemotional events [20]. Third, even mildly positive affective states profoundly effect the flexibility and efficiency of thinking and problem solving [10,11]. Finally, moods tend to bias judgments, indicating that a user in a positive mood will judge the interface and the system more positively [5].

MAX [12] is a method for detecting emotional states, based upon ideas first proposed by Darwin. He speculated that certain emotions appear to have the same facial expressions in many different cultures [6]. MAX's original formulation was intended to be applied to infants by pediatricians in order to better assess their tiny patients' emotional state. That information may prove valuable in a diagnostic setting and can provide insight to their physical condition as well. MAX focuses on a limited set of prototypical emotional expressions.

The nine prototypical emotions are:

- 1 Interest: a state of attention to something.
- 2 Surprise: wonder or astonishment to something unexpected.
- 3 Enjoyment: the state of enjoying.
- 4 Anger: a strong feeling of displeasure or hostility.
- 5 Sadness: sorrow or unhappiness.
- 6 Fear: agitation or anxiety caused by the presence of imminent danger.
- 7 Disgust: nausea or loathing.
- 8 Contempt: regarding something as inferior or worthless.
- 9 Shame: feelings of worthlessness.

These are maximally discriminative of specific human emotional expressions, which make them relatively easy to detect and useful in applied situations

We were also inspired by communication strategies based upon the Theory of the Seven Intrinsic Motivators [15], which postulates that human behavior is motivated by seven orthogonal concepts:

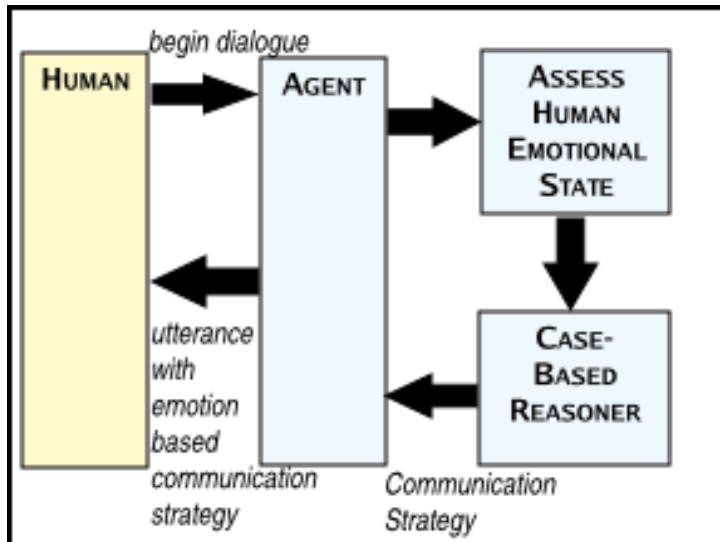
- 1 Fantasy: an opportunity to bring into reality imagined mental events.
- 2 Curiosity: a desire to experience novelty.
- 3 Challenge: a test against a personal goal.
- 4 Control: reward or punishment administered by others.
- 5 Competition: a test against another.
- 6 Cooperation: individuals working together to achieve a goal.
- 7 Recognition: acknowledged attention from respected others.

These strategies provided ready guidance in constructing unique conversational elements. When instructing a child to perform a task, the request may be phrased in such a way as to elicit one of these motivational concepts.

3 DYNAMIC DIALOGUE WITH EMOTIONS

Our approach is shown in Figure 1. We have a Human interacting with an Agent (the computer). Human engages in a dialogue by making an utterance to Agent. Agent processes the dialogue and assesses the emotional state of Human. Then, a CBR system recommends a communication strategy. The communication strategy is used to incorporate emotion to the next portion of the dialogue.

Figure 1. Dynamic Dialogue with Emotion



Agent then issues a response to Human, utilizing the communication strategy recommended by the CBR system. We do not envision the CBR system specifically presenting dialogue to the agent; rather, the agent, equipped with natural language methods, would incorporate a recommended emotion within the dialogue generation process. This emotion would then be incorporated into the computer's utterance.

We developed a prototype to begin exploring the elements of this strategy. Our first exploration involved utilization of the CBR system to select the communication strategy. This paper is limited to the incorporation of emotion based on the selection of the strategy, whereas the assessment of the emotional state of the human is future work.

4 PROTOTYPE INCORPORATING EMOTION IN HUMAN-COMPUTER DIALOGUE

As a first step in exploring the approach outlined in the previous section, we developed a prototype that selects a communication strategy and incorporates it into the dialogue based on the user's emotional state. We built a CBR prototype using episodes of parents addressing their children of ages from 3 to 7.

Cases are stored experiences, episodes that are representative of the process we are capturing. Here, our episodes are suggested statements that should be spoken in response to specific emotional states. These statements are collected, and along with other elements detailed below, are stored in a memory structure called a case base.

A CBR system will analyze a query and search its case base for cases that are most "similar" the query. There are a number of methods used to determine "similarity". However, this is not keyword matching or any other type of Boolean assessment. Rather, the results are sets of cases—similar to, but not exact.

4.1 Case Representation

Cases are represented by a set of features. In our prototype, we selected features that allowed the introduction of an activity and corresponding emotional state so that the system was able to return a suggested strategy. These features were defined as the elements of the problem, specifically: what to say, child's emotional state, child's age, dialogue location, and if the child is being stubborn.

Likewise, the systems solutions were also comprised of features. Specifically: the strategy, the *how to say*, i.e. the actual form of the recommended communication, and the strategy template.

What to say. The contents of the communication, i.e. what to say, include a verb and an object. The initial ones in our prototype are brush, teeth; leave, bath; eat, broccoli; change, clothes; stop, running.

Child's Emotional State. The emotional states of the child originate from MAX [12]. The states we are using in the prototype are mad, sad, happy, fearful, and engaged.

MAX utilizes nine emotional states; we eliminated four: surprise, disgust, contempt, and shame. Surprise and disgust were removed because they are fleeting emotional states. Contempt and shame were dropped because it may be difficult to obtain accurate cases for these emotional states—care givers may not want to discuss situations involving these emotions. Engaged was treated as the default emotional state. Emotional states are present when parents address the child. They do not refer to an intended or unintended outcome of our interaction (although they certainly could be).

Child's Age. Numeric feature indicating age from 3 to 7.

Dialogue Location. A binary feature indicating if the dialogue location takes place in the home or outside of it.

Stubborn. A binary feature indicating whether the child is being stubborn, which suggests the parent has already tried the communication before without success.

Strategy. The strategies originate from the seven intrinsic motivators discussed by [16] as underlying stimulus for the child to perform the intended task. These are fantasy, curiosity, challenge, control, competition, cooperation, and recognition.

How to Say. When you pair the strategy with the contents of the communication (what to say), the result is the *how to say*. We collected anecdotal examples of the resulting combination of strategies and contents. For example, brush, teeth; fantasy: “Let’s use this toothbrush to kill this evil plaque”. Change, clothes; cooperation: “Kid, let me help you change your clothes”. Brush, teeth; competition: “Let’s see whose teeth will look brighter”.

Strategy Template. The template is an additional component that has two purposes. One is to facilitate adaptation. Because the contents are expressed in terms of a verb and an object, they are in a form that can be considered canonical. The template determines how the emotion component incorporated by the strategy can be reused. For an example, see subsection Case Adaptation, below. The other purpose of the strategy template feature is diversity. Each case, when collected, will have one template. However, from additional collection and adaptation, a case may have multiple templates. The template in the example below was called “control countdown”. For example, one strategy is control and the canonical template was countdown. In our initial collection of 30 cases, we collected several other, different templates for control, namely, “do it now!”, “please do it”, “countdown”, “provide reason”, “divert”, “reward”, and “punishment”. These additional templates were added to our case base based on the interaction with the users.

4.2 Selecting a strategy to incorporate emotions

Case retrieval is a crucial step in CBR development, and is responsible for selecting a strategy of communication that incorporates emotion. It takes the new problem submitted to the system and assesses its similarity to candidate cases in the case base. The goal is to retrieve with higher similarity scores the cases that are more similar and thus potentially more useful to solve the new problem.

K-nearest neighbour was the case retrieval method used in our prototype. Each individual feature is compared between two cases at a time with individual similarity functions that vary by feature. For example, the similarity function for the emotional state produces a binary value, i.e., 0 or 1, there are no partial similarities. Age, on the other hand, is compared with a similarity function that outputs a value between 0 and 1. When comparing ages further away in the given scale, the function will output a value that is closer to 0, representing they are distant from each other; or closer to 1 otherwise.

The last element in similarity assessment is the representation of the relative importance of each feature. This is really challenging because it is a trial and error task unless we

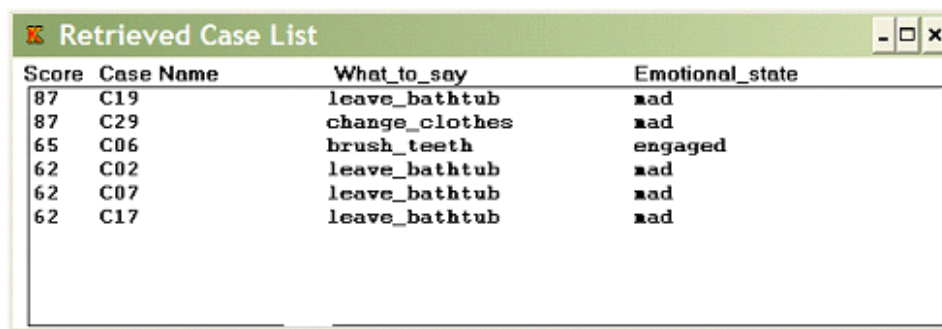
have enough exemplars to learn these weights inductively. The weights used in the prototype to represent the relative importance of features are laid out in Table 1.

Table 1. Relative importance of features

Features	Relative Importance
What to say	0.1
Emotional state	0.35
Age	0.05
Is child being stubborn?	0.3
Is child at home?	0.2

The result is a list of retrieved cases as shown in Figure 2. The higher the similarity scores, the more “similar” the cases are.

Figure 2. Example Retrieved Caption List



Score	Case Name	What_to_say	Emotional_state
87	C19	leave_bathtub	mad
87	C29	change_clothes	mad
65	C06	brush_teeth	engaged
62	C02	leave_bathtub	mad
62	C07	leave_bathtub	mad
62	C17	leave_bathtub	mad

4.3 Incorporating emotion into the dialogue

Adaptation in CBR is responsible for modifying the solution or solutions from retrieved cases to solve the new problem. In our prototype, two features are part of adaptation. From the retrieved case, the strategy template is used. From the new problem, the contents (i.e., what to say) are used.

For example, for a new problem whose contents are (eat, broccoli) while the retrieved case has (leave, bathtub) as the intention, and the strategy is “control countdown”. When applying the adaptation method, the *how to say* in the new case becomes, “Ok, you have two minutes to leave the bathtub”. The adaptation strategy would replace the verb and object to produce as *how to say* this sentence, “Ok, you have two minutes to eat your broccoli”.

5 VALIDATION

Although the development of our method is in its first stages and therefore not ready to be empirically tested, we now discuss our two main metrics with respect to our prototype. The two metrics are accuracy and diversity. Accuracy refers to the adequacy of an utterance produced by the system in the *how to say*. We consider a statement incorporated with emotion to be accurate if the majority of the user audience of a given system agrees that it is adequate. Note that this is mainly to determine whether the system suggested saying something that is improper. The scope of evaluating adequacy is determined by its audience of users because a statement incorporated with emotion can be perfectly adequate for a population of healthy adult users and very inadequate for a population of children or ill or disabled individuals.

The other metric is diversity. The system shall produce statements incorporated with emotion that are sufficiently diverse so that its users will not consider it predictive. This second metric aims at validating a system that is not repetitive and therefore its users will not be compelled to turn the system off right before they know it will say something and how it will utter it.

Data. The data we are currently using in this prototype has 30 cases. Please note that this is not a probabilistic experiment and the dataset is arbitrarily collected. Nevertheless, we

wanted to investigate the characteristics of the task of validating adequacy and diversity for this dataset.

Evaluating Accuracy. For the metric accuracy, we want to evaluate the adequacy of the retrieved strategies in an unbiased way. Because parent-child interactions typically fall into predictable patterns, we were able to formulate rules that define boundaries of reasonable and acceptable interactions. The role of these rules is in assessing whether our prototype violates any of these rules, thus recommending inadequate strategies. That is, if the similarity measure we engineered for the prototype produces results that do not violate adequacy in parent-child interactions.

We used four rules which were comprised of an emotional state, prohibited strategy pairs. The first rule is that if the emotional state of the child is *mad*, then do not select the fantasy strategy. The second rule is that if the child is in emotional state *sad*, then do not select challenge or competition. The third rule is that if the child is in emotional state *fearful*, then do not select curiosity or challenge. The fourth rule is that if the child is in emotional state *happy*, then do not select curiosity.

Table 2 shows five iterations of using the prototype. The new problem (target case) has values Yes for Stubborn and Yes for child being at home. In order to assess the adequacy of the emotional states versus the rules, that is, the compliance with the rules, we list the three most similar cases retrieved (i.e., the one with highest similarity score, the case with the second highest similarity score, and the case with the third highest similarity score).

Table 2. States when child is stubborn and at home

New Problem	Highest Sim. Score	Second Sim. Score	Third Sim. Score
mad	control	challenge	curiosity
sad	recognition	control	curiosity
engaged	curiosity	fantasy	control
happy	competition	curiosity	curiosity
fearful	cooperation	curiosity	fantasy

No one of the retrieved cases violated any of the rules. Therefore, for this limited number of iterations, the system recommended strategies can be considered as adequate.

Evaluating Diversity. Table 2 can also provide an indication of diversity. For only one of the iterations, two of the retrieved cases repeated a strategy. However, in an implementation with more cases, it is possible that the three most similar are not that diverse.

In order to guarantee diversity, we added the Strategy Template feature in the cases. This feature reflects an additional layer of abstraction available to the solutions. For the control strategy, for example, we collected several specializations, such as control countdown, control punishment. Moreover, one same strategy can be implemented multiple ways. For example, for the cooperation strategy, one of the templates is, "Let me help you [verb] your [object]." Whereas another is, "Go ahead and [verb] your [clothes] while I [verb] mine at the same time.

Therefore, we expect to enforce diversity in two ways. One is by retrieving multiple strategies that comply with a perception of adequacy within a given domain. Once the system determines that more than one strategy is applicable, it can verify the last strategies utilized in previous days or interactions with the same user and select one that is different. Another way is by randomly choosing the template. The system will become repetitive if every time it selects the next strategy different from the last one. Therefore, once the system is able to find an adequate strategy, a random function will select one of a set of templates.

5.1 Discussion

One item that caught our interest is the proportion in which a person is aware of his or her own emotional state and of the emotional state of the listener. That is, how many times are we aware of our own emotional state when we engage in a dialogue with someone? How often have we noticed our own emotional state only after the fact, when we examine

the effects our words had on the conversation? This led us to an interesting line of thought – will communication with computers be different than communication with humans because computers will focus solely on their human partner's emotional state? Computers perform without the influence of their own emotions; consequently, they may be more effective than humans because they may focus on the needs of the human rather than themselves.

Our underlying assumption is that there may be a link between emotional state as measured by MAX and an effective communication style embodied by the theory of the seven intrinsic motivators. Anecdotal evidence suggests this link exists, however the authors are unaware of any conclusive evidence of such a relationship. Furthermore, the efficacy of MAX and the intrinsic motivators have not been demonstrated in this environment.

6 CONCLUSIONS AND FUTURE WORK

This paper discussed the initial investigation on using a case-based method to select strategies to incorporate emotions in communication. Using a limited number of examples, we consider the system is able to recommend strategies that are both adequate and diverse, and therefore worthy of further research.

Further exploration needs to be undertaken exploring the link between task compliance and emotional state. One key assumption of this research is that some communication strategies are more effective in generating compliance in certain emotional states. Anecdotal evidence abounds supporting this common-sense hypothesis, however the links between these two elements require additional exploration.

Among future work is the implementation of the approach for a population of individuals and conduct a rigorous validation on the accuracy and diversity of the results of the method. Later, we will study different ways to integrate the system into dialogue planning systems and into natural language processing systems that will produce the *what to say*. The expected benefit of natural language is a richer adaptation to guarantee the adapted *how to say* from the strategy templates are correct and natural.

Another direction to complement our research is the incorporation of yet another reasoning task into the system to recognize the user's emotional state from a given dialogue. Such emotional diagnosis may also identify additional needs of the user, related to the current conversation. This additional task could trigger another function if implemented in a system such as the one described in [9].

Some open questions are: 1) Are there any dialogues that are particularly well suited or unsuited for this method? 2) Would there be any advantage in incorporating a human user model in this? 3) What overall effect does user emotions have in HCI? 4) Does incorporating user emotions or motivational strategies increase the likelihood of user compliance with the computer's suggestion?

REFERENCES

- 1 Aamodt, A. and E. Plaza (1994). "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches." *AI Communications, IOS Press* 7(1): 39-59.
- 2 Bruckman, A. and A. Bandlow (2003). Human-Computer Interaction For Kids. *The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications*. J. A. Jacko and A. Sears. Mahwah, N.J., Lawrence Erlbaum Associates: 428-438.
- 3 Cassell, J., J. Sullivan, et al., Eds. (2000). *Embodies conversational agents*. Cambridge, MA, MIT Press.
- 4 Clore, G. C. and K. Gasper (2000). Feeling is believing: Some affective influences on belief. *Emotions and beliefs: how feelings influence thoughts*. N. H. Frijda, A. S. R. Manstead and S. Bem. Cambridge, U.K., Cambridge University Press; Editions de la Maison des sciences de l'homme: vi, 249 p.
- 5 Clore, G. C., R. S. Wyer, Jr., et al. (2001). Affective feelings as feedback: Some cognitive consequences. *Theories of mood and cognition: A user's handbook*. L. L. Martain and G. C. Clore. Mahwah, NJ, Lawrence Erlbaum Associates: 63-84.
- 6 Darwin, C. (1896). *The expressions of the emotions in man and animals*. New York, Appleton.
- 7 Fraden, J. (1996). *Handbook of modern sensors*. Heidelberg, Germany, Springer-Verlag.
- 8 Gross, J. J. (1999). Emotion and emotion regulation. *Handbook of personality: Theory and research*. L. A. Pervin and O. P. John. New York, Guilford: 525-552.

- 9 Haigh, K. Z., L. M. Kiff, et al. (2004). The Independent LifeStyle Assistant™ (I.L.S.A.): AI Lessons Learned. The Sixteenth Innovative Applications of Artificial Intelligence Conference (IAAI-04), San Jose, CA.
- 10 Hirt, E. R., R. J. Melton, et al. (1996). "Processing goals, task interest, and the mood-performance relationship: A mediational analysis." Journal of Personality and Social Psychology **71**: 245-261.
- 11 Isen, A. M. (2000). Positive affect and decision making. Handbook of Emotions. M. Lewis and J. M. Haviland-Jones. New York, The Guilford Press: 417-435.
- 12 Izard, C. (1983). The maximally discriminative facial movement coding system. Newark, DE, Instructional Resources Center.
- 13 Jacko, J. A. and A. Sears (2003). The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications. Mahwah, N.J., Lawrence Erlbaum Associates.
- 14 Lisetti, C. and C. LeRouge (2004). Affective Computing in tele-home Health. 37th Hawaii International Conference on System Sciences (HICSS-37 2004), Big Island, HI, USA, IEEE Computer Society.
- 15 Malone, T. and M. Lepper (1987). Making learning fun: A taxonomy of intrinsic motivation for learning. Aptitude, learning and instruction; v.3. R. E. Snow and M. J. Farr. Hillsdale, NJ, Erlbaum: xiii, 370p.
- 16 Malone, T. and M. Lepper (1987). Making learning fun: A taxonomy of intrinsic motivation for learning. Aptitude, learning and instruction. R. E. Snow and M. J. Farr. Hillsdale, NJ, Lawrence Earlbaum.
- 17 Piaget, J. (1970). Science of education and the psychology of the child. New York, Orion Press.
- 18 Saffo, P. (1997). Sensors: The next wave of infotech innovation. Institute for the Future: 1997 Ten-Year Forecast. Menlo Park, CA, Institute for the Future: 115-122.
- 19 Schneider, K. G. (1996). "Children and information visualization technologies." Interactions **3**(5): 68-73.
- 20 Thorson, E. and M. Friestad (1985). The effects of emotion on episodic memory for television commercials. Advances in consumer psychology. P. Cafferata and A. Tybout. Lexington, MA, Lexington: 131-136.
- 21 Watson, I. (1999). "Case-based reasoning is a Methodology not a Technology." Knowledge-Based Systems **12**: 303-308.