

Intentional Affect: An Alternative Notion of Affective Interaction with a Machine

Shazia Afzal, Cecily Morrison, and Peter Robinson
University of Cambridge, Computer Laboratory
William Gates Building, 15 JJ Thompson Avenue
Cambridge, CB3 0FD UK
(+44) 1223 763541

{Shazia.Afzal, Cecily.Morrison, Peter.Robinson}@cl.cam.ac.uk

ABSTRACT

Affective Computing envisages truly effective human-machine interactions as being affect-sensitive. The field is both motivated and influenced by an understanding of emotion in an environment, that of person to person, that differs from its eventual application, person to machine. Analysing data obtained in a potential application environment – computer-assisted learning - we highlight the limitations of such an understanding and propose an alternative stance to affect, that of intentional affective interaction.

Categories and Subject Descriptors

H.1.2. User/Machine Systems – Human Factors; H5.2. User Interfaces – Evaluation/methodology; J.1. Computer Applications - Education;

General Terms

Measurement, Design, Experimentation, Human Factors

Keywords

Affective computing, emotion recognition, computer-assisted learning

1. INTRODUCTION

Research in Affective Computing [13] builds on the premise that adapting applications based on the emotional state of users leads to compelling and effective interaction with machines. This has often been interpreted to produce scenarios of use like the following: if a computer tutorial senses frustration, then it can adapt the content that the user receives to mollify that negative emotion, much like a human teacher would do. Such scenarios however, have an implicit assumption that people ‘interact’ with machines in the same way that they do with humans – that is, they suppose that users follow the same protocols of emotional behaviour. They expect that: (1) nonverbal behaviour associated

with emotional state will be similar to that observed in human-human interaction; and (2) users will accept the same type of adaptive intelligence from a machine as from a person. In this paper we explore the viability of this assumption and the resulting implications within the context of computer-assisted learning environments.

Although not as an explicit theoretical stance, the assumption that humans interact with machines as they do with humans is inherent in the methods and practices of Affective Computing. This can be observed in the way affect is conceptualised and subsequently modelled, as well as in how representative data is prepared for the development and training of potential affective computing technologies. For example, most of the computational techniques for recognising emotions are developed using databases that are oriented to prototypical representations of a few basic emotional expressions used with humans rather than collected from interaction with machines. Another example is the assumption that people appreciate having their environments changed by a machine, as they would from a well-meaning person.

This paper does not contain a full-fledged experiment or comparative study, but recounts our research journey as a thought experiment into the role of emotion when humans interact with machines – a theoretical contribution to the field rather than an experimental one. After providing some background (Section 2), we present our data collection exercise (Section 3) that motivated further exploration of emotion expression in a human-machine setting. We then present video analysis of the same data, pulling out inconsistencies of people’s actions with common assumptions (Section 4). We gradually weave an analysis that leads to the proposal of *intentional affect*, that is, the user’s deliberate use of expression to communicate with the machine as opposed to the machine inferring the unconscious emotional expressions of the user. The ramifications of thinking about using intentional affect to contribute to computer tutorial systems and the resulting research opportunities conclude the discussion.

2. BACKGROUND

Effective tutoring by humans is an interactive and guided process in which learner engagement is constantly monitored to provide remedial feedback and to maximise the motivation to learn [12]. Technologists are therefore keen to emulate the effectiveness of expert human tutors in the design and functioning of learning technologies. In naturalistic settings the availability of several channels of communication facilitates the constant monitoring necessary for such an interactive and flexible learning experience.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

It is known for instance that teachers use a range of communication channels to assess a student's learning state, including "the student's facial expressions, body language, intonation, and other paralinguistic cues" [9]. Advances in Affective Computing have opened the possibility of capturing the social dynamics of expert human mentoring using computer vision techniques and statistical inference to allow affect recognition from nonverbal behaviour like facial expressions, body gesture, head pose, voice and physiology.

The data drawn upon in this paper was originally aggregated to train such an automated facial affect recogniser. The objective was to collect naturalistic data in the target scenario - an increasingly emphasised stance in the field of Affective Computing to ensure that such systems generalise to real-world scenarios [2, 3]. However, the difficulties in designing the exercise urged the researcher in charge of the project to collaborate with an ethnographer to conduct detailed video analysis of the data. The intention was to reflect upon the role of emotions in human-machine interaction and develop a more qualitative understanding of emotion expression in this setting. The outcome, although unexpected, was a re-evaluation of the importance of context in human expressivity and the re-thinking of assumptions inherent in the methods and goals of Affective Computing.

3. DATA COLLECTION EXERCISE

3.1 Design

Seven participants (4 female, 3 male) were invited to our usability lab and were video recorded while doing two (sequentially varied) tasks: an interactive map-based geography tutorial and a card-matching game lasting approximately 30 minutes. The head and shoulders of the participants were captured by a video camera mounted on the computer monitor. After completing the two tasks, participants were asked to fill in three self-report measures of dispositional expressivity. The session finished with a semi-structured interview after which subjects self-annotated their videos for emotional content.

The tutorial asked participants to read about countries and their locations and was followed by a test of their learning. There was no time limit and participants took on average 20 minutes to complete this activity. The card matching activity instructed participants to choose an appropriate match out of four cue cards based on colour, shape, or number. It contained a number of planned events, including the screen blanking out, matches not being possible, and variation in feedback/scoring. The two tasks were chosen to ensure a variety of emotion expressions. The card matching game contained triggers/events in order to induce emotional responses in a more controlled manner than what was expected from the self-paced tutorial.

3.2 Measuring Expressivity

The three tests, BEQ, EES & ACT (see Table 1), were chosen to help us obtain an expressivity parameter for each of our participants. Nonverbal research indicates that there are differences in the manner and intensity by which people express their emotions. Self-report measures of nonverbal expressiveness assess such individual differences in the generation and/or expressions of emotions and a more general tendency to display affect spontaneously and across a wide range of situations [14]. Our choice was based on how the construct of emotion is conceptualized in each test, which component of emotion is

assessed, the target population, short administration time, availability, and psychometric properties like reliability and internal consistency. Although each of the three tests measures individual expressivity, they do so by looking at different attributes as outlined briefly in Table 1. It was unclear which attributes -- behavioural changes, dynamic style, or changes in overall behaviour -- would be most comparable to a measure of expressivity of human-machine interaction, so all were tried.

Table 1. Self-Report Measures of Individual Expressivity

<p>Affective Communication Test (ACT) [5]</p> <ul style="list-style-type: none"> • dynamic expressive style • ability to transmit emotion <p>Berkeley Expressivity Questionnaire (BEQ) [6]</p> <ul style="list-style-type: none"> • behavioural changes associated with expression of emotions • considers positive and negative emotions <p>Emotional Expressivity Scale (EES) [8]</p> <ul style="list-style-type: none"> • individual difference variable • consistency across situations and communication channels

In order to compare the measures of general expressivity to that in a human-machine setting, a comparable measure was needed for emotion expression in the human-machine interaction. The absence of an established comparative measure in HCI however, led us to adopt an eclectic approach to measure the overall expressivity in the interaction sequence. For each subject video, the number of non-neutral emotional expressions along with a global approximation of the number of emotions perceived was obtained. To further qualify the dynamics of expressions we used some dimensions drawn from speech annotation [11] that characterise the quality of the movement, as shown in Table 2 below. Detailed results from the comparative measures and the standard expressivity tests are not presented here for the sake of argument clarity but are reported elsewhere [1].

Table 2. Dimensions of Expressivity

<p>Overall activation: amount of activity</p> <ul style="list-style-type: none"> • Static/Passive, Neutral, Animated, Engaged <p>Spatial extent: amplitude of movements</p> <ul style="list-style-type: none"> • Contracted, Normal, Expanded <p>Temporal extent: duration of movements</p> <ul style="list-style-type: none"> • Slow/Sustained, Normal, Quick/Fast <p>Fluidity: continuity and smoothness of movement</p> <ul style="list-style-type: none"> • Smooth, Normal, Jerky <p>Power: strength and dynamics of movements</p> <ul style="list-style-type: none"> • Weak/Relaxed, Normal, Strong/Tense <p>Repetitivity: repetition of same expression/gesture many times</p> <ul style="list-style-type: none"> • Low, Normal, High

3.3 Annotation for Emotional Behaviour

Automatic prediction using machine learning relies on extensive training data which necessitates preparation of labelled representative data. This also serves as a baseline for training and testing different techniques and is therefore crucial for development and evaluation of computational models of emotion. Annotation can be achieved in two ways, by participants and by experts. The use of self-annotation is thought to give a subjective account of emotional experience. The strategy suffers from the known problem of low reliability and dependence on voluntary disclosure. Experts can provide observational assessments of emotion which overcome these problems but depend heavily on coder interpretations and are sensitive to the affect-decoding skills of the experts. We decided to try both these methods for purposes of comparison.

The self-annotation was implemented using an interval-based coding system [10] where participants were asked to label every 20 seconds of their recorded video for emotional content. Participants did this immediately after the tasks ended for the entire duration of video, approximately 30 minutes. The idea was that participants would be able to associate what they felt with what they expressed. Observation of the labelling process (via the one-way mirror of the usability lab), indicated that this was not the case. Although participants responded differently to watching their own expressions, some surprised, mimicking and laughing at themselves, and others embarrassed, rushing through the video; the reactions did not suggest that they associate a feeling with an expression but rather interpreted the expression as they might if it belonged to another in a social setting. This level of cognitive mediation confounds the way emotion perception is studied – as the initial reaction or instinctive recognition. For this reason, and the noted boredom of the participants, the self-annotation was considered to be unreliable and discarded.

In order to gain a more objective annotation, an expert coder used a standard annotation tool called ELAN¹, to observe and identify emotional behaviour. In light of the fact that the videos carried low emotional content the interval-based coding scheme was replaced with event-based coding at a macro-level. This is a method of capturing the social meaning of behaviour rather than specific cues or displays [10]. The purpose was to identify and tag emotional segments from the underlying videos. This proved to be problematic as the demarcation of emotional segments was extremely noisy. Emotional expressions often overlapped, co-occurred or blended subtly into a background expression. Even to a human expert, it was difficult to determine what constituted an emotional expression. This observation is not uncommon, and research involving naturalistic data has acknowledged this and has reported low inter-rater agreement on annotations with naturalistic data [3, 4].

3.4 Rethinking the Data

The process of designing and carrying out our data collection exercise produced more questions than data – each stage causing us to question what we meant by emotion. In order to gain a range of emotional expressions it was necessary to design a task that induced more observable emotions. This however would be contrary to the application-oriented nature of this research. When

choosing indexes of individual expressivity, the variety of possible measures for the same concept made it questionable whether these measures were context-specific and would apply to the human-machine interaction. Indeed, the expressivity measures did not correlate, positively or negatively. Lastly, the difficulty and unreliability of the emotion annotation process again prompted us to question what we were measuring. The inconsistencies we encountered between conceptual notions of emotion and the practical encounters we had collecting it prompted us to consider more closely what we meant by emotion and for what purpose we were collecting the data.

We took an exploratory approach and examined the data using video analysis. Two data themes emerged: (1) how task differences affect expression; and (2), how situational reactions affect displays of emotion. Building on these observations, we discuss the alternative of intentional interaction with machines and suggest the most productive way to pursue this concept in light of our data.

4. VIDEO ANALYSIS

4.1 Inferred Emotion Recognition

The aim of affect-sensitive technology is to interpret a user's affect state from his/her nonverbal behaviour. Applications such as a computer tutorial can adapt intelligently based on its understanding of user behaviour and without need for the user to express explicit intent. The goal is to get an insight into the emotional state of a user from observable signs like facial expressions or gesture. In the following two sub-sections, we examine the practicality of this desire by taking examples from our data.

4.1.1 Task Difference

We found that consistently across our participant group, regardless of level of expressivity, more emotional reactions were observed during the card game than during the tutorial. In the card matching game, sharp but frequent expression changes were observed as in Figure 1a. During the tutorial, shown in Figure 1b, faces became slack and an emotional expression occurred only every few minutes. The expression changes were infrequent, sustained, and slow. This is perhaps not surprising as the card game changed rapidly, giving many 'events' to which to respond while the tutorial required periods of concentration during reading.



Figure 1: Emotional Expressions: a) expressive face during the card-matching game; b) slack face during the tutorial

¹ <http://www.lat-mpi.eu/tools/tools/elan>

The card matching task set the pace and thus evoked reactive behaviour. It might be likened to a conversation between two people in which there is a constant stream of both non-verbal and verbal ‘events,’ to which to react. In contrast, the tutorial put the learner in control, involving application of the individual learning style and deeper cognitive engagement. The human-machine interaction that takes place during a tutorial then is likely to include less intense expressions of emotion and rather contain periods of low expressivity during concentration.

Machine recognition of emotion is going to prove technically difficult in situations like tutorials. As an alternative solution to this technical problem, one might induce more expressivity by adding ‘events’ through a pedagogical agent. Interestingly, some studies show that such interaction aides often prove distracting and are perceived as unnatural (c.f. [16]). This suggests that the connection between concentration and low expressivity in our data is not a matter of chance; concentration likely lowers the threshold of emotions and affects both the quantity and quality (dynamics) of expressive behaviour. It seems that type of activity with the machine has a substantial impact on the nature and expressivity of an individual. This in turn implies that *the applicability of affect-sensitive interfaces is constrained to certain types of tasks.*

4.1.2 Situational Reactions

Each participant indicated that they were confused when the screen went blank during the card game. Nonetheless, their reactions to confusion were quite different. P7, for example, laughed, while P3 appeared to be extremely concerned, almost alarmed. Although both indicated that they felt the same emotion – confused, their reaction to the situation engendered very different facial expressions. This data suggests that there is a distinction between felt emotions and situational reactions and it is rather the latter that is expressed. A related example, P3 expressed anxiety throughout the tutorial, while P7 never expressed anxiety. The former, as discovered during self-coding of the video, always feared getting something wrong, while the latter did not worry even when he did get something wrong. The two had very different attitudes towards learning.

It is not clear how an application should interpret user behaviour in those instances when it is the users’ attitude that is manifest and not necessarily emotion. The dominance of individual learning style in managing attitude rules out the simplistic notion of adapting content and pace of learning based on the learner’s affective state. *There are latent factors then that govern responses to interaction and these can have a direct effect on apparent user behaviour.* Deciding how and when an emotion recognition system should give credence to behavioural cues is indeed going to be difficult.

4.2 Intentional Affective Interaction

In the above section, we suggest that spontaneous emotion recognition from apparent user behaviour may not be practical. On the one hand, concentration and interest seem to reduce expressivity causing technical difficulties; and on the other hand, the distinction between felt emotions and situational reactions creates design issues. For a more meaningful adaptation and interaction we suggest an alternative – intentional affective interaction, wherein a user, understanding the consequences of their non-verbal behaviour, makes an active effort to be

understood. This solves the technical problem of producing emotions to recognize as well as the design problem of not knowing whether a facial expression should be interpreted or not. However, our data suggests that designing an intentional affect responsive interface is not simply a design problem.

4.2.1 Considering Social Consequences

Six out of our seven participants indicated on their questionnaires that they would interact differently if they knew the computer could respond to their affective state. From this we originally hypothesized that this ‘difference’ would be a magnification or conscious regulation of behaviour as happens when one tries consciously to communicate an emotion, such as pleasure. However, two incidents in our data indicate that this may not occur since interacting with a computer is devoid of the usual social consequences that stimulate non-verbal behaviour in everyday life.

One participant, P2, was quite flamboyant when speaking with the researchers, liberal in her use of body gestures and facial expressions. It was clear that she intended these non-verbal behaviours to draw attention and reinforce her opinion rather than express a felt emotion. It is unlikely that she would use a similar strategy with a computer. Or even if she did, interpreting her expressions would be problematic. Indeed, she was one of the least expressive individuals to the computer in our study.

Data from another participant, P1, suggests a further complication. P1 had some of the lowest expressivity test scores but was very expressive while using the computer, so much so that she surprised herself. Not only did she evidence quick and dramatic changes of emotion but engaged in other expressive behaviours such as gasping and ‘giving the finger’ to the machine. In discussion with P1, she revealed that she worked in a male-dominated technical environment where emotional displays encouraged a gender stereotyped image that she wanted to avoid. Being with a computer gave her an outlet for expressing herself. Her apparent comfort with a computer was partly due to its being a machine and thus non-judgemental. If the computer could understand her, she would be less likely to use the same expressions or ‘abusive’ gestures as an emotional outlet.

Social consequences are a significant part of the use of non-verbal expressions during interaction. Our analysis indicates that awareness of the machine’s passivity and lack of social consequences will affect the user’s expressivity. *Research into intentional affective interaction then cannot rely on data obtained in a setting without an emotionally sensitive interface as currently done.* New methods of data collection are needed.

5. DISCUSSION

In light of the inconsistencies of expressivity observed in our data -- the importance of task type on emotion expression and the distinction between felt emotion and situational response – we suggest that intentional affective interaction with a machine could be a promising solution. We note that current research methods are impractical to apply to this problem. We propose two alternative means of investigation, one theoretical and one practical. Accountability, a fundamental feature of ethno-methodology, provides a basis to conceptualize human-machine interaction, and ‘new media arts’ provides a means to explore it.

Accountability [7] propounds that within a community there is a shared understanding of how to act and interpret action.

Moreover, all behaviour is accountable and interpreted. For example, if someone passes you in the corridor, they can choose to greet you or not. They cannot choose to be invisible in order that their actions have no social consequences. One is currently not accountable in an interaction with a computer. To design an affect sensitive program, then, we would need to consider how to create a community that could act and interpret actions.

Designing such a community, even if it were just a user and her computer, could be done using 'new media arts'. Suchman [15] argues that new media artists are at the forefront of exploring the boundary between humans and machines not from observing and cognizing about humans but rather by building systems that provoke users to explore a particular type of interaction. Our future work is to build a system that encourages its user to explore the process of becoming accountable with the goal of gaining design insight into the most appropriate, or 'natural' communication for intentional affective programs.

6. SUMMARY & CONCLUSIONS

We have presented an analysis that leads us to question both the accuracy and lack of nuance in the assumption that people 'interact' with machines in the same way as they do with other humans. Apart from the methodological issues associated with the identification and measurement of affect, there are additional issues that need to be considered for viable application of affect-sensitive techniques. Our video-analysis results indicate that people express themselves less during a cognitively engaging task like a tutorial than during the faster-paced activity of a card-matching game. It seems that there is a likely conflict between emotion expression and concentration, indicating that emotion recognition for environments demanding concentration may prove to be difficult and of limited application. Furthermore, the problem in distinguishing between a felt emotion and a situational response makes it difficult to utilize the recognized expression for the purposes of an application.

We propose intentional affective interaction between humans and machines, in which the user knows that the machine is reacting to its expressions and actively utilizes them. Our data suggests that in order to pursue this design idea further, we must gain a better understanding of how *intentionality* influences interaction and how expression is related to social context - something the computer will never have.

7. ACKNOWLEDGMENTS

We would like to acknowledge the funding support of the Gates Cambridge Trust and Boeing Inc. as well as thank all the participants in our study.

8. REFERENCES

- [1] Afzal, S., & Robinson, P. 2008. Dispositional Expressivity and HCI, Emotion in HCI, HCI 2008, Liverpool
- [2] Batliner, A., Fischer, K., Huber, R., Spilker, J., Noth, E. 2003. How to find Trouble in Communication, 40 (1-2), pp. 117-143
- [3] Cowie, R., Douglas-Cowie, E. & Cox, C. 2005. Beyond emotion archetypes: Databases for emotion modelling using neural networks, *Neural Networks*, 18, 371-388
- [4] Devillers, L., Vidrascu, L., Lamel, L. 2005. Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 18, pp. 407-422
- [5] Friedman, H.S. Friedman, H.S., Prince, L.M., Riggio, R.E., & DiMatteo, M.R. 1980. Understanding and assessing nonverbal expressiveness: the Affective Communication Test. *Journal of Personality and Social Psychology*, 39(2), 333-351
- [6] Gross, J.J., & John, O. P. 1995. Facets of emotional expressivity: Three self-report factors and correlates. *Personality and Individual Differences*, 19, 555-568
- [7] Heritage, J. 1984. *Garfinkel and Ethnomethodology*. Polity Press
- [8] Kring, A., Smith, D.A., & Neale, J. M. 1994. Individual differences in dispositional expressiveness: Development and validation of the emotional expressivity scale. *Journal of Personality & Social Psychology*, 66, 934-949
- [9] Lepper, M.R., Woolverton, M., Mumme, D.L., & Gurtner J. 1993. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In Lajoie S.P. & Derry S.J. (Eds.): *Computers as Cognitive Tools*, 75-105. Lawrence Erlbaum, Hillsdale, New Jersey
- [10] Manusov, V.L. 2005. *Sourcebook of Nonverbal Measures: Going Beyond Words*. Lawrence Erlbaum Associates
- [11] Martin, J., Abrilian, S., Devillers, L., Lamolle, M., Mancini, M. & Pelachaud, C. 2005. Levels of Representation in the Annotation of Emotion for the Specification of Expressivity in ECAs, *Intelligent Virtual Agents 2005*, LNAI 3661, pp. 405-417
- [12] Merrill, D.C., Reiser, B.J., Ranney, M. & Trafton, J.G. 1992. Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *Journal of the Learning Sciences*, 2, 277-305
- [13] Picard, R.W. 1997. *Affective Computing*. The MIT Press, Cambridge
- [14] Riggio, R.E., & Riggio, H.R. 2005. Self-report measures of emotional and nonverbal expressiveness. In V. Manusov (Ed.), *The sourcebook of nonverbal measures: Going beyond words* (pp. 105-111). Mahwah, NJ: Erlbaum
- [15] Suchman, L. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press
- [16] Zakharov, K., Mitrovic, A. & Johnston, L. 2007. *Pedagogical Agents Trying on a Caring Mentor Role*, Artificial Intelligence in Education, Los Angeles