The Aesthetics of Reading

Kevin Larson (Microsoft) & Rosalind Picard (MIT)

Dr. Kevin Larson
Microsoft Advanced Reading Technologies
1 Microsoft Way
Redmond, WA
Phone: (425) 703-5204
Fax: (425) 936-7329

Dr. Rosalind W. Picard
Director of Affective Computing Research
Co-Director Things That Think Consortium
MIT Media Laboratory
20 Ames Street
Cambridge, MA 02139
Phone: (617) 253-0611
Fax: (617) 253-5275

Abstract

In this paper we demonstrate a new methodology that can be used to measure aesthetic differences by examining the cognitive effects produced by elevated mood. Specifically in this paper we examine the benefits of good typography and find that good typography induces a good mood. When participants were asked to read text with either good or poor typography in two studies, the participants who received the good typography performed better on relative subjective duration and on certain cognitive tasks.
The Aesthetics of Reading

Kevin Larson (Microsoft) & Rosalind Picard (MIT)

Abstract

It has been previously demonstrated that when people are in a positive mood that they will perform better on cognitive tasks that involve creativity (Isen, 1987). It is fairly easy to manipulate mood, the presentation of a small gift or watching five minutes of a humorous video is enough to generate measurable differences on certain cognitive tasks. In this paper we demonstrate that it is possible to apply this finding to the measurement of affect in software systems or other product evaluations. When participants were asked to read text with either good or bad typography in two different studies, the participants who received the good typography afterwards performed better on Isen’s cognitive tasks as well as on subjective duration assessment.

Introduction

Microsoft’s Advanced Reading Technology team is working to improve the quality of on-screen text. This involves improvements to rendering technologies (e.g. ClearType), fonts, and text layout. There is also a large basic research effort, much of which is being conducted with academic collaborators.

Some improvements to on-screen text have had measurable performance benefits. Gugerty, Tyrrell, Aten, & Edmonds (2004) have found reading speed and comprehension advantages for the ClearType rendering engine over the basic black & white rendering engine. In a thatistoscopic lexical decision task they found that participants were statistically reliably more accurate at recognizing words rendered with ClearType, with a magnitude difference of roughly 17%. In a sentence comprehension study they found statistically reliable differences in favor of ClearType for both reading speed and comprehension.

But other on-screen reading improvements don’t demonstrate the large performance differences that we see with ClearType. Barbara Chaparro at Wichita State University (unpublished) investigated the performance difference between documents with good page layout and poor page layout and found no speed or comprehension differences between these two conditions. Users reported with Likert scale questionnaires that they greatly preferred the good page layout documents. Good page layout includes typographically correct headers, paragraph indentation, good figure placement, and block quoting.
Figure 1: Pages from page layout documents. The right-hand page has good image placement, good headers, and well marked paragraphs. There was no reading speed or comprehension difference, but the good page layout was greatly preferred.

Chaparro also investigated a variety of typographic improvements grouped together under a technology called OpenType. OpenType provides application support for features such as ligatures, kerning, small caps, old style numerals, and sub/superscript. In this study no reading speed or comprehension differences were found, nor did users report preference differences with the Likert scale preference questionnaires. It appears that these differences, which are very obvious to any typographer, are too subtle for most readers to notice explicitly.

Figure 2: The right-hand sample paragraph demonstrates the OpenType ligatures, kerning, small caps, old style numerals, and sub/superscript features. There were no reading speed, comprehension, or preference differences between these two conditions.
The art of typography

Typographers are attuned to subtle features when they design and set type. One such feature that is quite noticeable to the readers’ perceptual system is symmetry. It is a surprisingly difficult challenge to make and render symmetric type. All the strokes across a font need to be of equal weight - if one vertical stem is heavier than the next then the relative darkness will appear as a dark spot on the page. The white space within characters needs to be balanced - if the space under one arch of an m is not equal to the other arch or to the arch of the n, this will also appear as a dark spot on the page. And the white space between characters needs to be equal to complete the desired symmetry.

Figure 3: The first paragraph has uneven stroke weight and uneven spacing both within characters and between characters. The second paragraph only has poor spacing between characters. The third paragraph has symmetric characters and spacing.

Designing type with even spacing between characters appears to be a straightforward task, but this is also a design challenge. If all characters had vertical stems on the left and right side then it would be a simple task of measuring the distance between one letter and the next. Unfortunately round and triangular characters make this a much more difficult task. When the simple horizontal distance rule is applied between characters of different shapes, the result is uneven spacing. Instead typographers look at the apparent space between character pairs. This results in two round characters being placed closer together than two vertical characters.
Figure 4: The samples on the left are spaced with an equal amount of horizontal distance between each letter. The samples on the right are adjusted by a typographer to have an apparent equal amount of space between each letter.

Type design and setting is an artistic enterprise with the goal of creating a beautiful page of text. While it was surprising that there were no performance differences in the page setting study, it was astonishing that there were no preference differences in the OpenType study. The features that are part of OpenType are considered essential to typographers and are generally believed to improve the aesthetics of the page. The participants generally agreed that the OpenType features were beneficial after they were pointed out. It would be beneficial to have a methodology that can detect typographic improvements without first drawing attention to the differences.

The impact of positive emotion

Our goal with this project is to develop a measure that is sensitive to improvements in aesthetics. By extending two earlier methodologies we hope to find one that is successful in detecting differences. The first methodology is based on the adage time flies when you’re having fun. Participants’ perception of time is manipulated by the enjoyment of their activity. The second methodology is based on the finding that participants perform better on certain cognitive tasks when they are in a good mood.

Weybrew extended Zeigarnik’s work on task interruption by demonstrating that task interruptions cause participants to overestimate task duration (Weybrew, 1984). While interruptions cause task durations to be overestimated, non-interrupted tasks tended to be underestimated. Weybrew also found that more engaging tasks tend to be underestimated. Recent work has turned this finding into a useful usability measure called relative subjective duration (Czerwinski, Horvitz, Cutrell, 2001). Relative subjective duration (RSD) measures participant’s perception of how long they have been performing a task. Difficult tasks tend to be overestimated in duration while easy tasks are underestimated in duration.
Our hope is that RSD not only detects task difficulty, but also aesthetic differences. In our studies we will access RSD by interrupting the participants after they have been reading for a certain period of time. We expect that duration of reading tasks with good typography and aesthetic qualities will be underestimated by participants while duration of tasks with poor typography and aesthetic qualities will be overestimated by participants.

The second methodology is based on research conducted by Alice Isen and her colleagues that has shown that participants who are put in a good mood before performing certain cognitive tasks will perform better than participants who are not (Isen, Daubman, & Nowicki, 1987). Participants can be placed in a good mood by receiving a small gift such as a candy bar or by watching five minutes of a humorous video. After bring induced into a good mood, participants perform better on creative cognitive tasks such as the candle task (Duncker, 1945) and remote associates task (Mednick, 1962) then a group of participants that were not induced into a good mood.

If a candy bar or humorous video can induce a good mood, can good typography induce a similar kind of good mood? We expect that after reading documents with good typography that participants will perform better on the tasks that Isen used to measure creative cognition.

We conducted two studies where half of the participants read with good typography and half with poor typography. Our hypothesis is that the participants in the good typography condition will underestimate time with RSD and perform better on cognitive tasks then the participants in the poor typography condition. This would suggest that good typography does elevate mood.

**Study 1**

In our first study, we asked participants to read text with high quality typography or with poor typography and took three kinds of measurements: Relative subjective duration, Likert scale preference questions, and performance in the candle task. Each participant was given a Tablet computer with special software that let them read a full issue of the New Yorker magazine.

**Methods**

- **Subjects**
  Twenty participants were selected from the Microsoft database of computer users in the Puget Sound area who are willing to participate in usability studies in exchange for software gratuities. Half the participants were female; all were aged 20-40 years old and had 20/20 or better corrected vision. All participants classified themselves as occasional readers of the New Yorker magazine. Ten participants were randomly assigned each to the good typography or poor typography condition. One participant in the poor typography condition was familiar with the candle task, and her data from that task was discarded.
Materials

Prototype software created by the Microsoft ePeriodicals team was used in this study. ePeriodicals are electronic versions of print magazines designed for the Tablet PC. Each page is designed to fit perfectly on a full screen in portrait orientation without any scrolling. Page turning (page up and page down) happens with the Tablet PC hardware buttons. It is also possible to navigate directly to the table of contents or any article directly with the simple pen user interface. In this study we used the content from the January 5, 2004 edition of the New Yorker magazine. This includes the text, images, and advertisements used in that edition of the print magazine. The page layout differed from that of the print magazine to accommodate the size of the Tablet PC screen.

We created two versions of the New Yorker ePeriodical. The good typography version was the best ePeriodical we could make at the time using the New Yorker font with ClearType and good hyphenation and justification. The bad typography version used the bitmap version of the Courier font and had an extra 2 points of space added between every word. While it looks terrible, users had no trouble reading the text – and the content was exactly the same in the two conditions.

Figure 5: The New Yorker with good typography
Half the participants received the good typography ePeriodical and half received the poor typography version. The participants were not told until after the study was complete that there was a good and bad version. Each participant was given a brief tutorial on the ePeriodical user interface which included how to use the page up and page down hardware buttons, how to navigate to the table of contents, and how to navigate to any article. Participants could choose to read anything they wanted from this ePeriodical.

Participants were told that they would be reading for 20 minutes. They were interrupted during the reading session to collect the RSD. For the first four participants (two in each condition) we interrupted the participants after fifteen minutes. Each of these four participants estimated that they had been reading for fifteen minutes. For the remaining participants we changed the RSD interruption interval to sixteen minutes. After collecting the RSD data, the participants continued to read until twenty minutes passed.

After the reading session, the participants were given a simple preference questionnaire. It included six 7-point Likert scale questions which asked participants to indicate their agreement with statements such as “The text was easy to read”, “The page layout looked like a magazine”, and “I felt like I had control”.

Finally, participants were given the candle task, a creative cognitive task that Isen has found to detect mood differences. In the candle task participants are given a box full of tacks, a candle, a
match, and a corkboard affixed to a wall; their task is to attach the candle to the corkboard in such a way that the wax won’t drip all over the place when lit. They were given ten minutes to solve the task. The task is considered correctly solved if the tacks are emptied from the box, the box tacked to the corkboard, and the candle placed inside the box. All other solutions are considered incorrect. One of the participants in the poor typography condition said that she knew the solution to the candle task because she was familiar with the task from a psychology course in college. This participant's data was not included in the candle task results.

**Results**

We took three measures: Relative subjective duration, a Likert scale preference questionnaire, and the candle task. We found reliable differences with RSD and the candle task, but none with the preference questionnaire.

With RSD we found that participants in the poor typography condition underestimated their reading time by 24 seconds on average, while participants in the good typography condition underestimated their reading time by 3 minutes and 18 seconds on average. This is a reliable difference, $t(18)=2.17, p<.05$. This data includes the lack of difference with the first four participants who had an earlier interruption point. Because participants in both conditions underestimated the amount of time they had been reading, this indicates that the reading task was pleasant and engaging. The greater underestimation in the good typography condition indicates that good quality typography is responsible for greater engagement during the reading task.

With the candle task we found that 4 of 10 participants successfully correctly solved the task in the good typography condition while 0 of 9 participants correctly solved the task in the poor typography condition. This is a reliable difference, $\chi^2(1) = 2.47, p < .05$. This indicates that participants in the good typography condition were in a better mood before starting the candle task then were the participants in the poor typography condition.

**Study 2**

Excited by the results of the first study, we decided to run a second test very similar to the first to confirm our findings. We made two major changes to the study. First, because in the first study we found larger RSD difference with a longer duration interruption point, we decided to increase the duration even further to 17 minutes. Second, we changed the final cognitive task from the candle task to the remote associates task, another cognitive task that Isen has shown to be impacted by positive mood.

In the remote associates task participants are given three words such as *water*, *skate*, and *cream* and asked to generate a word that will create a common compound with each of the three words. In this example the correct answer is *ice*. The three items were shown on a computer screen, and the participants pressed a keyboard button as soon as they knew the answer. The items were
shown for up to 15 seconds. After the participant pressed the button or 15 seconds had elapsed, the participants were asked to respond with the compound associate. We collected both reaction time and accuracy data from this test. Isen found that participants placed in a positive mood successfully completed a high percentage of these trials and at a faster rate then participants not placed in a positive mood.

Most other details of the study are identical to the first study. We again used the January 5, 2004 edition of the New Yorker for the ePeriodical content, and created the good and poor typography versions with the same differences. The same experimenter gave the participants the same instructions about how to use an ePeriodical. Twenty new participants were recruited for the between subjects design with the same set of selection criterion as the first study.

**Results**

In the second study we took three measures: Relative subjective duration, a Likert scale preference questionnaire, and the remote associates task. We found reliable differences with RSD and the preference questionnaire, but not with the remote associates task.

With RSD we found that participants in the poor typography condition underestimated their reading time by 2 minutes and 21 seconds on average, while participants in the good typography condition underestimated their reading time by 5 minutes and 12 seconds on average. This is a reliable difference, *t*(18)=2.19, *p*<.05. The findings are similar to Study 1, again participants underestimated reading duration in both the good and poor typography conditions, but had a larger underestimation in the good typography condition. The overestimations are greater in this study than in study 1 for both the good and poor typography. The main difference is that we extended the time duration to 17 minutes which resulted in a greater range of time estimations. Reading in both conditions still appears to be an engaging task, but reading with good typography is more engaging.

We used the same 7-point Likert scale questions as in the first study, but this time some of the preference scores were reliably in favor of the good typography:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Bad Typography</th>
<th>Good Typography</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>The text was easy to read.</td>
<td>3.6</td>
<td>5.5</td>
<td><em>t</em>(18)=2.63, <em>p</em>&lt;.05</td>
</tr>
<tr>
<td>The page layout looked like a magazine.</td>
<td>5.0</td>
<td>6.7</td>
<td><em>t</em>(18)=4.02, <em>p</em>&lt;.01</td>
</tr>
<tr>
<td>I felt like I had control.</td>
<td>4.8</td>
<td>6.2</td>
<td><em>t</em>(18)=2.55, <em>p</em>&lt;.05</td>
</tr>
<tr>
<td>It was easy to get where I wanted to go.</td>
<td>4.4</td>
<td>5.8</td>
<td><em>t</em>(18)=2.15, <em>p</em>&lt;.05</td>
</tr>
</tbody>
</table>
We expected to find preference differences in favor of the good typography ePeriodical in both studies, and do not have an explanation why reliable differences were not found in the first study. There were clear differences in text quality and magazine appearance that should have been noticeable to the reader, even though it is a between subjects design where any given reader only sees either the good or the poor typography ePeriodical. There were no actual navigation differences between the two versions, and participant ratings reflect general perceived improvement for the good typography ePeriodical.

With the remote associates test we found that the good typography participants succeeded at 52% of the tasks at an average speed of 6395ms to the poor typography’s 48% at an average speed of 6715ms. Neither the accuracy nor the speed difference is statistically reliable.

**Conclusion**

From these studies we have both discovered methodologies for measuring affect differences and that high quality typography can improve mood. RSD previously had been used to measure differences in task difficulty. In these studies RSD has been expanded to detect differences where the aesthetics of the text differed while the task (reading) is unchanged. Creative cognitive tasks have been previously shown performance improvements when participants were induced into a positive mood. These studies have shown the potential for using those same tasks to measure mood differences. However the results were mixed with the candle task showing the expected benefit while no difference was found with the remote associates task.

We have also demonstrated that high quality typography appears to induce a positive mood, similar to earlier mood inducers such as a small gift or watching a humorous video. This is an exciting finding because there are important differences between good and poor typography that appear to have little effect on common performance measures such as reading speed and comprehension. To help move the field of typography forward we need methods that can successfully measure aesthetic differences.

This is still a young project and there is much to do. One potential problem with these studies is that the affect value of the reading content was not controlled in this study since every participant could read the articles of their choosing. One article in the ePeriodical told of the loss of the author’s childhood dog. This was clearly a sad story and could have had an undue impact if more participants in one condition choose to read this story. In future work we will examine placing tighter controls on the content that the participants read in order to eliminate this as a confounding factor. We are also planning to examine subtler typographic differences. In these studies we manipulated font quality, rendering quality, and layout quality. Next we would like to see if this methodology is sensitive enough to detect aesthetic differences such as those seen in Chaparro’s page setting and OpenType studies. We hope these methodologies are sensitive enough to measure more subtle typographic differences.
Acknowledgements

Special thanks to Tracy Premo who ran the studies, and to Radoslav Nickolov and Steven Kim for helping create the good and bad ePeriodicals.

References


Stoop to Conquer: Posture and affect interact to influence computer users’ persistence

Hyung-il Ahn¹, Alea Teeters¹, Andrew Wang¹, Cynthia Breazeal¹, and Rosalind Picard¹

¹ MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139, USA {hiahn, alea, zephyrus, cynthiab, picard}@media.mit.edu

Abstract. RoCo, a novel robotic computer, has the capability to move its monitor in subtly expressive ways that respond to and promote its user’s postural movement. Motivated by Riskind’s “Stoop to conquer” research where it was found that postures congruous to the type of outcome a person received (e.g. slumping following a failure or sitting up proudly following a success) led to significantly better performance in a subsequent cognitive task than incongruous postures (e.g. sitting up proudly following a failure or slumping following success), we performed two experiments where RoCo was used to manipulate its user’s posture. Our results show that people tend to be more persistent on a subsequent task when RoCo’s posture is congruous to their affective state than when it is incongruous. Our study is the first to show that a computer’s “pose” congruous or incongruous to a user’s affective state can influence factors such as persistence in problem solving tasks.

Keywords: User Studies, Robotic Computer (RoCo), Affective Interaction, Posture and Emotion, Human-Robot Interaction

1 Introduction

Everyone knows that how you feel can influence what you think and do. However, many people do not know that there is a growing body of findings from psychology, cognitive science, and neuroscience where more subtle affective states have been shown to systematically influence cognition [10,12,13,19]. In particular, a number of studies have explored the effect of body posture on affect and cognition [16,17,5,21]. An example is the theory in Riskind’s “stoop to conquer” research [16,17], where it was found that incongruous postures, such as slumping after a success, negatively affected subsequent performance, while congruous postures, such as slumping after a failure, helped to mitigate the effects of failing.

Motivated by Riskind’s “stoop to conquer” research, we performed two experiments examining the interaction of posture and affect on persistence, creativity, and comfort. While Riskind’s experiments were conducted over 22 years ago with pencil and paper and with human-manipulated postures, we modified the experiment, adapting it to be applicable to today’s computer users, and importantly, to users of
future new technologies that we believe will have articulated smooth movements (like RoCo).

This paper is organized as follows. First we present a brief description of the new RoCo robotic computer platform to motivate this new desktop technology that moves and can get people to shift their posture naturally while working. Next we offer a summary of relevant psychological literature with respect to body, affect, and cognition interaction effects that informs and guides our work. We then present two novel user studies adapting Riskind’s experiment to the RoCo platform. Finally we discuss the findings and conclusions, and suggest future directions.

Figure 1. RoCo: a robotic computer (left) and its graphical simulator for designing new behaviors (right)

2 RoCo: A New Robotic Computer Platform

RoCo is designed to lie within a continuum that might be loosely described as having ordinary fixed desktop computers at one end, and humanoid robots at the other end. In between is a huge unexplored space, where one can begin to animate the things in the office environment that usually do not move on their own: computers, chairs, and more. With increasing interest in promoting healthy activity, we are starting with the desktop computer, considering how it might move in ways that get the person using it to move more. RoCo was carefully designed to look like an ordinary computer, but to move in ways that are completely paradigm changing. RoCo has no face or body that attempts to evoke humanoid or animal characteristics: it has a regular monitor, keyboard, and box, which are sessile. However, it also has motors that give it smooth, expressive, articulated movement.

The physical RoCo robot has five degrees of freedom (DoFs) that manipulate its mechanical “neck” with a LCD screen as its “face” and “head.” See Figure 1. Two DoFs move the mechanical neck (base yaw and base pitch) and three DoFs (head yaw, head pitch, and head roll) move the LCD display. These five degrees allow RoCo to perform a wide variety of simple motions, including nodding, shaking its head, and leaning forward. These life-like motions are sufficient to implement a wide variety of immediacy behaviors and postural changes. For example, if you lean toward RoCo, perhaps to read something tiny on the screen, it could meet you halfway. At the same time, we are designing RoCo with sensors to monitor your facial and postural movements so that it does not move in ways that distract you. Inspired by examining how and when humans move naturally, we are currently
aiming to have RoCo hold very still while you are attentive to the screen, but to look for natural breaks to maximize your movement without distracting or annoying you. For example, if you have been slumping for a while, and then turn away your gaze, then when you return your gaze, you might find RoCo has “stretched” upward, subtly encouraging you to adjust your posture upward. However, as we show in this paper, there is more to consider than your attention and your posture.

In this paper we wish to isolate how its posture interacts with and influences the person using it. In fact, for this paper, we will pre-set RoCo’s postures so that the novelty of a computer moving does not enter into our results. We do this as the first set of experiments with this new technology, to carefully control the variables influencing the outcomes.

3 Body, Affect and Cognition Interaction

3.1 Affect and Cognition Interactions

Studies from psychology, cognitive science, and neuroscience indicate that affect and emotional experience interact with cognition in significant and useful ways. Current understanding is that emotion plays a useful role in regulating learning, creative problem solving, and decision making. For example, Isen shows that a positive mood promotes a tendency toward greater creativity and flexibility in negotiation and in problem solving, as well as more efficiency and thoroughness in decision making [10]. These effects have been found across many different groups, ages, and positive affect manipulations. Other specific influences of affect on cognition have also been found for negative affective states, e.g., Schwartz argues that being in a sad mood enables better performance on certain kinds of analytic tests [19].

Emotion not only influences cognition, but it also interacts with information in the environment in ways that can enhance or hinder your ability to perform. Cliff Nass and colleagues, while trying to decide if a voice in the automobile driver’s environment should sound subdued and calm or energetic and upbeat, ran an experiment trying both kinds of voices [13]. Importantly, they also looked at the two conditions where drivers were either upset or happy (having just viewed disturbing or funny films.) In a total of four conditions, the happy or upset drivers drove in a simulator with either an energetic voice or a subdued voice talking to them and asking them questions. On multiple measures of driving performance and cognitive performance, happy drivers did better overall than upset drivers. But there was also an important and interesting interaction, highly relevant to the work in this paper. When the voice was congruous with the driver’s state (energetic/upbeat for happy drivers, subdued/calm for upset drivers) then performance was significantly better than in the two incongruous conditions. The worst performance of all four conditions occurred when the upset drivers were paired with the energetic and upbeat voice. It is this kind of effect – where performance is improved by mood congruent interaction – that we explore in this paper. However in this paper, we induce the congruence condition in an entirely new way.
3.2 Body and Affect Interactions

According to Riskind’s [16] “appropriateness hypothesis”, slumped or upright physical postures are not just passive indicators of mental states but can reciprocally affect the mental states and behavior. The results suggest that “inappropriate” postures, such as slumping after a positive success, can undermine subsequent motivation and feelings of control, while “appropriate” postures, such as slumping after a failure, help to mitigate the effects of failing. His findings suggest that it is therefore not beneficial after a failure to sit with chin up as if proud, despite that people often tell children to do that.

In Riskind’s original experiment, all the subjects were first asked to perform a cognitive task (e.g. a tracing puzzle task). The affective manipulation (positive/negative affect) was handled by the experimenter who informed the subject of his or her “score” on the task. A high score (success) was designed to elicit a positive affect in the subject, while a poor score (failure) to elicit a negative affect. After this first task, the subjects were escorted to a different room and assisted to take one of three postures reflecting appropriateness (neutral/slumped/upright) under the false pretense of a biofeedback experiment. The subjects were required to hold this posture for 8 minutes before relaxing it and performing a subsequent cognitive task (e.g. additional puzzle tracing tasks). Riskind found that subjects in incongruous postures (stooped/slumped following success, upright following failure), felt like they had less control, showed less motivation in persistence tasks, and reported higher depression than subjects in congruous postures. His study suggested that a slumped versus upright posture orientation can guide and moderate information processing and responses to positive and negative mood-relevant stimuli.

4 Our Purpose, Hypothesis and Predictions

In this paper, we explore whether a computer’s “posture” can influence its user’s subsequent posture, and if the congruence of the user’s body state with their affective state during a task leads to improved task measures, such as persistence in problem solving. This research serves as a baseline study to investigate RoCo's ability to manipulate both the user's posture and the user's cognitive and affective state, illuminating the capabilities of this new technology. The key question, therefore, is how we design new technologies to beneficially influence the interactions between a human user’s body, affective, and cognitive states. We wish not only to provide an ergonomic experience, but also to foster healthful computer usage and improved task outcomes.

Our study expands on the appropriateness hypothesis [17], predicting that congruous posture guides an individual towards self-regulating behaviors while incongruous posture leads to self-defeating behavior. Taking advantage of the unique RoCo research platform, our experiment introduces a different posture manipulation method that allows the subject to perform dependent measure tasks on a computer while in the manipulated posture. Thus, while Riskind measured the effect of a prior posture on a subsequent cognitive task, we can now measure the effect of the posture
concurrent with the task. Our prediction is that RoCo will be an effective agent for manipulating posture and inducing the “stoop to conquer” effect.

In Riskind’s original experiment, subjects were asked to either slump or sit upright under the false pretense of a biofeedback experiment. In his study, a human experimenter was responsible for posture manipulation. While this kind of manipulation is useful for detecting the “stoop to conquer” effect, it is not practical in real applications that aim to utilize this effect in a more natural way. However, when a user works on the RoCo platform, by changing RoCo’s posture, we have been able to get RoCo to subtly lure the user into a target posture without seriously interrupting his or her workflow. Also, in our experiment using the RoCo platform, since RoCo is responsible for posture manipulation instead of a human experimenter, this change makes the manipulation significantly more subtle and unobtrusive than in the Riskind experiment.

5 Experiments and Results

5.1 EXPERIMENT 1

This experiment measures persistence on a helplessness task, creativity on a word association task, and general spatial cognition on a puzzle task as a function of congruous and incongruous postures following affect manipulation [3, 20].

Subjects. Seventy-one naive subjects were recruited from our school and the surrounding area. Subjects were given a $10 gift certificate to Amazon.com as compensation for their participation. In this study there were six control conditions each of which involved a mood manipulation (success / failure) and one of RoCo’s posture states (slumped / neutral / upright). Subjects were assigned to one of the six conditions based on the order that they signed up to participate in the study.

Preliminaries. When subjects arrived they were first greeted by the experimenter then led to a standard PC. The experimenter read the following standard set of instructions aloud to the subject: “Please be seated. In front of you is a standard computer setup with mouse, keyboard, monitor and a pen tablet for use in the tracing puzzles. You may arrange these components on the desk any way you like. Please read the instructions carefully as you go. The height of the chair is adjustable with a lever underneath the seat. I will be outside the curtains, if you have any questions or get confused, but in general, please try do as much on your own as possible.” The experimenter then left the area while the subject was shown a two minute video clip previously shown to induce neutral affect [18].

Success-Failure Manipulation. Half of the six conditions involved inducing a feeling of success, while the other half involved inducing a feeling of failure. This was accomplished as follows. Subjects were given a series of four tracing puzzles to solve. They had two minutes to solve each puzzle. To solve a puzzle, the subject must trace over the design without lifting a pen from the puzzle or retracing any lines. In this case, the puzzles were presented on a standard LCD screen and pen tracing is done with a computer pen and tablet input device. The puzzles used are the same set used by Riskind [16] in his studies as well as by Glass and Singer [7]. To create a
success condition, all four puzzles were solvable. Generally each subject was able to solve at least three out of the four. Unsolved puzzles were usually the result of not carefully reading the instructions beforehand or difficulty using the pen and tablet interface. Regardless of how the subject actually performed, a results chart was displayed and subjects were told they scored an 8 out of 10. For the failure condition, the first and last puzzles were insolvable. The sense of failure was further reinforced by displaying the same results chart as in the success condition, except in this case they were told that they scored a 3 out of 10.

**Posture Manipulation.** Following the success-failure manipulation, the subject’s chair was rolled over a few feet to RoCo, the position of which had already been preset to slumped, upright, or neutral, relative to the first PC. These positions are shown in Figure 2. Notice that they are not quite the same as can be obtained with the typical degrees of freedom on a desktop monitor, although people certainly are capable of slumping or sitting up straight in front of an ordinary desktop monitor. The poses of RoCo are somewhat exaggerated to more strongly encourage sitting up or slumping relative to the neutral position used during the interaction with the regular PC. The subject, while seated in the same calibrated-height chair, was then asked to perform another series of puzzles, this time on RoCo. The subject was video taped as a manipulation check.

**Figure 2.** RoCo’s postures: neutral (left), slumped (center), upright (right)

**Dependent Measures.** The experiment examined three dependent measures: persistence, spatial cognition, and creativity. However, in this paper we only discuss the results about persistence, because we do not have enough space to discuss all the dependent measures.

**Insolvable Tracing Task to Test Persistence:** The subject was given four mathematically insolvable tracing puzzles with a time limit of two minutes for each. This task assumes that the fewer the number of tries in the allotted time, the lower the subject’s tolerance for a frustrating task. Some of the puzzles are the same as those used in Riskind’s original study. Additional puzzles were created by transforming some solvable into insolvable. Debriefings showed that only people who knew the mathematical rule for solvability ahead of time were able to distinguish solvable from insolvable puzzles; their data was dropped.

**Debriefing.** Following the dependent measure tests, each subject was given a full debriefing. As a check on the success-failure manipulation, subjects were asked how well they thought they performed on the first part. All subjects in the failure manipulation responded with answers like “not well”, “below average”, and “ok”, suggesting that the manipulation was successful. Similarly, most subjects in the
success case responded with answers such as “well” and “above average”. Four subjects in the success condition who had trouble with the tracing puzzles in part one reported that they did not do well. Their data were omitted since the manipulation was not successful. Following the manipulation check, the details of the study were disclosed including the impossibility of some of the tracing puzzles and the fabricated test results in part one. Four subjects also reported at this time that they knew the tracing puzzles were mathematically impossible. Their data were also omitted.

**Main Results.** 1. RoCo posture’s influence on the user posture: An outside hypothesis-blind person coded the changing user posture for the video data collected from 64 subjects. Based on the states of the chin, shoulder and back of the user, the coder classified the user posture into three basic states (Slumped / Neutral / Upright) every 30 seconds. We could not code three subjects’ video data because they sat down too close to RoCo so the camera didn’t capture their posture properly. The video analysis shows that RoCo’s posture strongly influenced the user’s posture in both success and failure conditions (See Tables 1 and 2). The most frequently occurring posture state during the subsequent tasks was used for counting the user posture in these tables. Most subjects (about 70% of all subjects) tended to keep the dominant posture for over 80% of the task time. Also, about 15% of all subjects changed the posture state every 5~7 minutes. 2. Persistence on Task: As predicted, the analysis on the persistence on the insolvable puzzles data (summarized in Table 3 and shown in Figure 3) did reveal a statistically significant interaction effect, F(2, 57) = 4.1, p < 0.05. Further simple effects analysis by success-failure outcome revealed that success subjects exhibited more persistence when they used RoCo in its upright position (M = 11.97) after their success than when they used RoCo in its neutral position (M = 8.32), or in its slumped position (M = 8.15). F(2, 57) = 7, p < 0.01. However, unlike in Riskind’s study, failure subjects showed no statistical difference across postures, F(2, 57) = 0.1. We address this in the discussion and in Experiment 2. Also, there were no main effects for either the success-failure or the posture manipulations, F(2, 57) < 2, p < 0.2 and F(2, 57) < 3, p < 0.07 respectively.

<table>
<thead>
<tr>
<th>RoCo</th>
<th>Slumped</th>
<th>Neutral</th>
<th>Upright</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slumped</td>
<td>9</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Upright</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

**Table 1.** RoCo posture’s influence on the user posture in the success condition (the number of subjects is shown)

<table>
<thead>
<tr>
<th>User</th>
<th>RoCo</th>
<th>Slumped</th>
<th>Neutral</th>
<th>Upright</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slumped</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Upright</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** RoCo posture’s influence on the user posture in the failure condition (the number of subjects is shown)
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Slumped</th>
<th>Neutral</th>
<th>Upright</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>N 8.15</td>
<td>8.32</td>
<td>11.97</td>
</tr>
<tr>
<td>Failure</td>
<td>N 8.33</td>
<td>8.75</td>
<td>8.41</td>
</tr>
</tbody>
</table>

Table 3. Average number of tracing attempts

Discussion. We adapted a number of factors from Riskind’s original study to work with RoCo, which may explain why our results differ for the failure condition. In Riskind’s study, subjects were taken to a separate room and told to hold the assigned posture for approximately eight minutes under the pretense of a biofeedback experiment. They then performed the second set of tasks without controlling for posture. However, in our study, the user is free to adopt any posture as long as he or she can still read the screen. The video footage shows that users seemed to adjust their posture, particularly while sitting back and thinking about possible solutions. They tended to move more in the slumped conditions where they reported lower comfort (especially the failure-slumped condition, which can also foster a sense of malaise). While thinking, the primary posture manipulation was relaxed. Thus, our subjects who encountered the slumped condition of RoCo did not slump as consistently as Riskind’s subjects did, as his were forced to hold the slumped position for 8 minutes, without moving. In sum, one possible explanation for why we are seeing the positive-upright effect but not the “negative-stooped” effect in our study may be that subjects did not sustain the stooped posture for a sufficiently long period of time. We designed Experiment 2 to address this problem.

5.2 EXPERIMENT 2

Experiment 2 was designed to see if having the subject hold the slumped position (giving a person a more involved task to do on RoCo during the posture manipulation) after failure would produce the “stoop to conquer” effect. Here we observed the dependent measures in each of the three conditions (failure-upright, failure-neutral and failure-slumped). Moreover, differently from Experiment 1, we included a new decision-making (gambling) task that had reading instructions and
content written in a small font on the LCD monitor in order to encourage the subject to stay in a position focused on the monitor. This task did appear (from videos) to keep subjects in the desired posture for a longer time than Experiment 1 because they had to scrutinize details on a screen for 8 minutes.

**Subjects.** Thirty-seven subjects were recruited from our school and surrounding area, each between 18 and 40 years old. Subjects were randomly assigned to one of the three conditions (failure-upright/failure-neutral/failure-slumped).

**Procedure.** The same procedure as in Experiment 1 was performed, except that they performed the decision-making task before other dependent measure tasks. Since our primary interest in Experiment 2 was the “stoop to conquer” effect on the persistence measure in the failure conditions, all subjects were assigned to the failure manipulation. Subjects used RoCo for the 8 minute decision-making task, then performed dependent measure tasks.

**Debriefing.** Following the dependent measure tests, the 37 subjects were given a full debriefing as in Experiment 1. We found that seven of the subjects did not feel bad after the failure manipulation: four subjects did not feel bad in spite of the low score, and three subjects believed that the low score given for the manipulation was not true (failed mood manipulation). Thus, their results were excluded from all our analysis, and we see 30 subjects in Tables 4 and 5. In Table 6, we found we had to omit additional 12 subjects: after the video analysis, we found that two subjects’ posture did not match with RoCo’s conditioned posture (failed posture manipulation). Also, one subject reported that she had much trouble in using the pen tablet for the tracing puzzles, and nine subjects knew the rule for whether a tracing puzzle was solvable or not. While these problems did not interfere in the mood manipulation (as verified in the debriefing), they would make comparisons of persistence unfair, because they skipped puzzles, so they are omitted from Table 6. Since the sample size is small, we risk false acceptance or rejection of the null hypothesis. Thus, below, we report all the averages and standard deviations as well as results of statistical tests.

**Main Results.** 1. **RoCo posture’s influence on the user posture:** RoCo’s posture strongly influenced the user’s posture. Also, compared with the failure condition of Experiment 1 (Table 2), the 8 minute cognitive task helped the user keep a constant posture longer (See Tables 4 and 5). 2. **Persistence on Task:** One-way ANOVA analysis was applied to the persistence measure from the insolvable puzzles data (summarized in Table 6 and shown in Figure 4). The result shows a statistically significant posture effect on the persistent measure, $F(2, 15) = 3.70, p < 0.05$. Subjects showed higher persistence when they used RoCo in its slumped position ($M = 9.75, SD = 2.50$) after their failure than when they used RoCo in its neutral position ($M = 7.36, SD = 1.58$), or in its upright position ($M = 6.85, SD = 1.60$). Thus, the better persistence of the matched combinations supports the appropriateness hypothesis.

**Discussion.** Experiment 1 showed that people tended to be more persistent on a subsequent task when they used RoCo in its upright position after success than when they used RoCo in its neutral or slumped position. However, Experiment 1 did not show a significant “stoop to conquer” effect on the same persistence measure in the negative mood conditions. We hypothesized that these results were mainly caused by the fact that subjects did not keep the target posture for a significant period of time while doing the subsequent task on RoCo. Thus, Experiment 2 was designed to
encourage subjects to hold the target posture longer before doing subsequent tasks. This new experiment allowed us to observe that people in the negative mood were more persistent using RoCo’s slumped posture than using its neutral or upright postures, thus achieving the “Stoop to Conquer” effect. Therefore, Experiment 1 and 2 show that a computer’s “pose” congruous or incongruous to a user’s affective state can influence performance factors such as task persistence. Also, we find that holding the target posture before doing other dependent measure tasks may be critical in utilizing the “stoop to conquer” effect. This makes sense if the slumping helps a person to process their negative state, and thus go on to better performance.

![Table 4](image)

**Table 4.** RoCo posture’s influence on the user posture for the initial 8 minutes (the number of subjects is shown)

<table>
<thead>
<tr>
<th>User</th>
<th>Slumped</th>
<th>Neutral</th>
<th>Upright</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slumped</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Upright</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

![Table 5](image)

**Table 5.** RoCo posture’s influence on the user posture for the total task time (the number of subjects is shown)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Slumped</th>
<th>Neutral</th>
<th>Upright</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>9.75</td>
<td>7.36</td>
<td>6.85</td>
</tr>
<tr>
<td>SD</td>
<td>2.50</td>
<td>1.58</td>
<td>1.60</td>
</tr>
</tbody>
</table>

![Table 6](image)

**Table 6.** Average number of tracing attempts

![Figure 4](image)

**Figure 4.** (Failure condition) Average number of tracing attempts
Since we first wanted to observe the “stoop to conquer” effect on the RoCo platform without involving any other effects between RoCo and the human subject, the experiments in this paper did not engage RoCo’s dynamic movements of responding appropriately to the user’s affective state. When RoCo uses these dynamic behaviors, there might be additional emotion contagion effects between RoCo and the user. Thus, we sampled these behavioral positions by exposing users to three different fixed points: slumped, neutral, upright. There was probably still a novelty effect of using a monitor with wires running up it, but this effect was constant across all conditions.

RoCo is an entirely new kind of system, which can greet its user socially and move during natural interaction much like people move. Suppose the user greets RoCo cheerily, then sits and slumps. Our findings confirm the theory that an upright posture (congruent with cheery mood) could help this user be more productive. RoCo can begin to move upward, subtly, without being distracting. More likely, RoCo will observe that you are already slumped, and then choose movements to respond best to your wishes, which may include increased productivity. The three positions tested so far (and the 8min timing) provide the first proof of concept that posture-mood interaction matters in HCI. This now opens the door to investigating what timings and positions are most effective.

6 Conclusion

We use RoCo in a novel user study to explore whether a computer’s “posture” can influence its user’s subsequent posture, and if the interaction of the user’s body state with their affective state during a task leads to improved task measures, such as persistence in problem solving. These findings lend support to the theory of embodied cognition where invoking a cognitive concept invokes an associated bodily (and/or affective) state, and vice versa. When the states are congruent, there is less conflict, and more resources to devote to task performance. This paper is the first to show that mood-posture interactions influence performance for a person sitting in a chair using a computer monitor.

Research in computer-human interaction has long ignored human feelings, conducting experiments that (effectively) assume users are in a neutral mood. Our findings suggest that it might be important to bias users into multiple moods, and then measure outcomes. The field of economics has found that this makes a big difference in matching theory to real behavior (e.g., [12], where the endowment effect is reversed dependent on mood.) We suggest that the field of human-computer interaction may similarly find that measuring affective state is important, and can lead to measurably different outcomes. Persistence and perhaps many other cognitive variables are likely to be influenced by body and affective states.

Acknowledgments. This work was supported by NSF SGER award IIS-0533703.

References


Evaluating affective interactions: Alternatives to asking what users feel

Rosalind W. Picard  
MIT Media Laboratory  
20 Ames Street, 020a  
Cambridge, MA 02139  
+1 (617) 253-0689  
picard@media.mit.edu

Shaundra Bryant Daily  
MIT Media Laboratory  
20 Ames Street, 120g  
Cambridge, MA 02139  
+1 (617) 253-0369  
sbdaily@media.mit.edu

ABSTRACT
In this paper, we advocate the use of behavior-based methods for use in evaluating affective interactions. We consider behavior-based measures to include both measures of bodily movements or physiological signals and task-based performance measures.

INTRODUCTION
Recent years have seen a large increase in research directed towards adding an affective component to human computer interaction. The ability to measure user affect has become important for intelligent interfaces that aim to either establish believable interactions or alter internal behavior based on the user’s affect. Evaluating and interpreting this measure presents a challenge because of many ambiguities related to affect definition, communication, and interpretation.

Classical methods for evaluating affect tend to focus on questionnaires: asking you what you feel now, or interviews, perhaps after the experiment, with a video of your performance in front of you, asking you instant by instant to recall what you felt at each moment during the earlier task. While such “self report” methods are valuable, and we continue to use them in our work, this paper will highlight some alternatives to self-report of feelings. The discussion below is divided into two categories: body measures (e.g. changes in muscle activity), and task measures (e.g. better ability to solve a creative problem).

BODY MEASURES OF AFFECT
The last decade has brought great strides in giving computers affective perceptual abilities, with new sensors for physiology and for behavior, such as body-worn accelerometers, rubber and fabric electrodes, miniature cameras and microphones, and garment or accessory-type devices, along with new algorithms for recognizing patterns in the sensed signals such as recognition of facial expressions from video or of stress patterns from thermal imagery of the face and other physiological measures. Body measures are not presented here as a replacement for other measures, but rather as additional information that may help combat some of the difficulties encountered with questionnaires and other more subjective methods. Possibly the biggest advantage is that body measurements can be taken in parallel with the interaction rather than interrupting the user or asking him after the task.

An exhaustive list of body-based measures is beyond the scope of this paper, however, Table 1 cites a sample of existing methods (leaving out lots of examples of publications in each of these categories, and also leaving out categories, e.g. EEG and ECG-based measures, and more). Clearly there are lots of possible body measures that may capture aspects of an affective state, including the combination of multiple modalities, which can reduce the uncertainty associated with using a single measure (Mednick et al. 1964; DeSilva et al. 1997; Huang et al. 1998; Picard et al. 2001; Kapoor et al. 2004)

One benefit of these “body” measures is that they can provide additional insight into the user’s emotional state without directly relying on his cognitive judgment of his emotional state. Additionally, some of them can be used without the user’s knowledge, perhaps with the goal of limiting the amount of misinformation that may arise from his feeling of being monitored. (This can also be seen as a drawback if one is concerned about privacy and about the use of sensing without a person’s knowledge).
Table 1. Body-based measures of affect (partial set of examples)

<table>
<thead>
<tr>
<th>Modality</th>
<th>Sensor</th>
<th>Is it socially communicated?</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facial Activity</td>
<td>Video (Tian et al. 200; Barlett et al. 1999; Donato et al. 1999; Cowie et al. 2001)</td>
<td>Yes</td>
<td>Facial expressions can differ significantly from genuinely felt feelings</td>
</tr>
<tr>
<td></td>
<td>IR Video (Kapoor et al. 2003)</td>
<td></td>
<td>Highlights pupils &amp; specularityties</td>
</tr>
<tr>
<td></td>
<td>Thermal Video (Pavlidis et al. 2002)</td>
<td></td>
<td>Usually works better than ordinary video when head moves (better eye detection)</td>
</tr>
<tr>
<td>Posture Activity</td>
<td>Force sensitive resistors (Smith 2000; Mota &amp;Picard 2003; Tan et al. 2003)</td>
<td>Yes, but not as pressure</td>
<td>Good results discriminating level of interest in students in computer learning interactions</td>
</tr>
<tr>
<td>Hand Tension &amp; Activity</td>
<td>Force sensitive resistors or Sentograph (Clynes 1986; Reynolds 2001; Qi &amp;Picard 2002; Dennerlein et al. 2003)</td>
<td>Varies; depends on gesture</td>
<td>Can be sensed from handling of mouse, steering wheel, etc., and pressure has been shown to be higher during a frustrating task</td>
</tr>
<tr>
<td>Gestural Activity</td>
<td>Electromyogram electrodes (Marrin Nakra &amp;Picard 1998; Dubost &amp;Tanaka 2002)</td>
<td>Visibility varies</td>
<td>Shown for expression sensing in conducting music; other gestures largely unexplored w.r.t. expression</td>
</tr>
<tr>
<td>Vocal Expression</td>
<td>Microphone (Banse &amp;Scherer 1996; Cowie et al. 2001; Ang et al. 2002; Breazeal &amp;Aryananda 2002; Fernandez 2004)</td>
<td>Yes</td>
<td>Most methods are great for discriminating arousal but not for valence; limited to times when user is speaking</td>
</tr>
<tr>
<td>Language and choice of words</td>
<td>Text analysis tools (Goertzel et al. 2000; Elliott 2002; Liu et al. 2003)</td>
<td>Yes</td>
<td>Can be used with interfaces requiring textual input; promising for valence; trivial to sense scripted dialogue moves</td>
</tr>
<tr>
<td>Electrodermal Activity</td>
<td>Electrodes (can also be clothing snaps, metallic fabric, etc.) (Picard &amp; Scheirer 2001)</td>
<td>No; except perhaps sweaty palm</td>
<td>Good at detecting changes in arousal but doesn’t distinguish positive/negative, and can also be triggered by non-affective changes</td>
</tr>
</tbody>
</table>

**TASK MEASURES OF AFFECT**

A variety of findings have shown ways that affective states tend to influence various behaviors on subsequent tasks. To the extent that such findings are robust, they can be used to indirectly assess aspects of affect that may have been elicited during an interaction. For example, Isen and colleagues (1987) have demonstrated that positive affect can influence the way cognitive material is organized and have shown that this enables broader forms of thinking, and consideration of less typical solutions, which is useful in creative problem solving (Isen 1987). Using a variety of different techniques such as gifts, comical movies, or refreshments, a positive affect state was induced in the study participants. The slightly positive emotional state benefited subjects’ performance on tests such as Duncker’s (1945) candle task, the Mednicks’ Remote Associates Test and medical decision-making with hypothetical patients (Mednick et al. 1964; Isen 1987; Isen 1991). Subjects also better integrated the material presented to them and exhibited an ability to better organize their protocols as compared to a control group (Isen 1991). Also, like Isen, Schwarz has found that a negative affective state corresponds to a higher level of spontaneous causal reasoning, which fosters bottom-up, data driven processing. Therefore, when involved in an analytical task, it may actually help to be in a sad mood (Schwarz 2002).

Recently, this kind of “task measure” was examined to see if exposing users to reading tasks using two different kinds of fonts impacted it. The hypothesis was that people reading a passage written using good typography would perform better on the candle task and on the remote associates test than readers reading the same content presented with poor typography. While the first such study of this kind was small (N=20), the findings were supportive of this hypothesis (Larson &Picard 2005).

The typography study also examined another indirect task assessment method that we think is of increasing interest for assessing affect. This measure involves asking somebody “how long do you think you spent on that task?” and, its ratio to the actual time spent is known as subjective duration assessment (Czerwinski et al. 2001). Using this measure (of how long they thought they were working on it) it has been shown that difficult tasks tend to be overestimated in duration while easy tasks are underestimated in duration. We hypothesize that this measure might also be related to frustration, which predicts it would also be influenced by task difficulty and by other factors such as time pressure and irritating aspects of the task. In two separate typography studies, this measure was found to be significant (p< 0.05), each study with N=20: In both studies, subjects using the good typography
underestimated their reading times by a significantly larger amount than did subjects using bad typography. In one of the studies, this difference held even though subjects’ self-reports of the quality of the typography did not differ significantly.

We have been interested in the generality of such time-based measures for indirectly assessing affect. Recently, Picard and Liu proposed a new variation, “relative subjective count (RSC),” based on asking people who were interrupted many times during the day by the technology being investigated, “How many times does it seem like you were interrupted by this technology?” This perceived number was divided by the actual number of interruptions to obtain the RSC. Comparing two nearly identical systems, which differed mainly in their expressions of empathy, they found that people had a significantly lower RSC when the technology was empathetic. This measure also agreed with self-reported views of how stressful the technology was. We suggest that the RSC might provide a new indirect way of assessing affect related to the stress or irritation associated with an interaction (Liu 2004).

While these kinds of assessment measures are new and require much more investigation before they are fully understood, they potentially offer a nice alternative for exploring certain states such as stress and frustration related to an interaction without having to ask directly about any negative aspects of the user’s experience.

In another area of interest, Lerner, et al. have shown that affect has important influences on economic decision-making. Positive affect was shown to reverse the endowment effect (the tendency for selling prices to exceed buying or ‘choice’ prices for the same object), while negative affect eliminated the endowment effect (Lerner et al. 2004). The affect, in this case, was evoked by using the viewing of movies followed by a writing task in which the subjects attempted to write about how they would describe how they were feeling to a friend.

While we think that behavioral task measures such as these may prove powerful for indirectly assessing when a desired positive state has likely been achieved in a group of individuals. While they are not as direct as measuring an individual’s emotional bodily reaction, and so far the results are only on populations, and not on individuals, these task-based measures can be accomplished without any special sensors or sophisticated analysis. With a large enough group of individuals, the statistical significance can potentially provide a strong assessment.

CONCLUSIONS
We have highlighted several means of assessing affect beyond directly asking somebody what they are feeling. One can imagine an interface interaction in which the user’s facial and electrodermal activity are monitored for valence and arousal information, and the interaction is followed by an assessment task such as the Duncker’s Candle Test, where better success is expected with more positive affect. Thus, information about the user’s affect can be gleaned from the physiological sources as well as the task performance. This data could additionally be compared against self-reported measures. There is a lot of room for new methods to be discovered; the ones we have presented here are just a few of the possibilities.

REFERENCES


BACKGROUND OF THE AUTHORS

Rosalind W. Picard is founder and director of the Affective Computing Research Group at the MIT Media Laboratory. The author of over a hundred peer-reviewed scientific articles in multidimensional signal modeling, computer vision, pattern recognition, machine learning, and human-computer interaction, Picard is known internationally for pioneering research in affective computing and, prior to that, for pioneering research in content-based image and video retrieval. Her award-winning book, Affective Computing. (MIT Press, 1997) lays the groundwork for giving machines the skills of emotional intelligence.

Shaundra B. Daily is a graduate student at the MIT Media Laboratory, working in both the Affective Computing and Future of Learning Groups. She holds a Bachelor’s in Electrical Engineering with honors from the Florida State University, and a Master’s degree, in Electrical Engineering and Computer Science, from Florida Agricultural and Mechanical University. Her main interests include the design and evaluation of interfaces designed to support affective development through writing, constructing, and other forms of expression as well as technologically supported community and economic development.