

# Designing Empathic Computers: The Effect of Multimodal Empathic Feedback Using Animated Agent

Hien Nguyen

Department of Computing Science  
University of Aberdeen  
+44 (0)1224 27 2223  
csc286@abdn.ac.uk

Judith Masthoff

Department of Computing Science  
University of Aberdeen  
+44 (0)1224 27 2299  
j.masthoff@abdn.ac.uk

## ABSTRACT

Experiencing emotional distress is the number one reason why people who are undergoing behaviour modification (e.g. quitting smoking, dieting) suffer from relapses. Providing emotional support is an effective way to help them overcome the unpleasant effects of negative affect and adhere to their regimen. Building computers with such ability has grabbed the attention of the HCI community in recent years. Early research has shown some promising results when adopting strategies of how we comfort others, but many questions on how to build such systems remain unanswered. This paper presents the results of a 2 (modality: animated vs. no visual) by 3 (intervention: non-empathy vs. empathy vs. empathy and expressivity) between-subjects study that investigates the impact of two important factors and their interaction in the design of such systems: (1) different ways of expressing empathy, and (2) the modality of delivering such content. Findings and implications for the design of empathic computer systems are discussed and directions for future research are suggested.

## Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems – *human factors, software psychology*.

## General Terms

Human Factors, Affective Computing, Design, Experimentation.

## Keywords

Human Factors, Affective Computing, Design, Experimentation.

## 1. INTRODUCTION

Building computers that can provide emotional support to their users to help them deal with experienced emotions and offer them encouragement and comfort during difficult times has important implications in many areas. Yet, we have only found a handful of studies addressing this problem [1-3], leaving several questions on how to build such systems unanswered. The key of emotional support is the ability to recognize people's feelings and express appropriate empathic feedback through language and non-verbal gestures such as affective facial expressions [4]. In this paper, we extend previous studies [1-3] to investigate the impact of these two important factors and their interaction in the design of such systems: (1) different strategies of expressing empathy verbally, and (2) the modality of delivering such content (i.e. text-only vs. speech and other non-verbal gestures). The paper begins with a discussion of the undesirable effects of negative emotions, the role

of empathy in helping people to overcome such affective states, and related research on empathic computer systems in Section 2. Section 3 describes our research questions. This is followed by the experiment design, results, discussion, and conclusion in Section 4, 5, 6 and 7 respectively.

## 2. BACKGROUND

### 2.1 The Function of Emotions

Emotions arise when we attend to, interpret, and analyse ongoing events and situations that are relevant to our personal goals. As those events attain, maintain, or thwart our goals, they evoke appropriate emotions. We experience positive emotions when our desired goals are attained or maintained and undesired goals are avoided, and negative emotions when our desired goals are threatened [5]. We, therefore, have the tendency to adopt or cling to behaviours that make us feel good but to avoid or drop behaviours that make us feel unpleasant. The emotions that we experience in everyday life events, in turn, have a significant impact on our moods, psychological well-being, physical health [6], and affect the way we make decision, process information, and interact with others in future events and situations. They can hurt as well as help. In general, negative emotions such as distress, frustration, anger, anxiety and sadness have been found to be associated with many unpleasant effects: increased likelihood to take risk, selective encoding and retrieving of negative information making us even easier to become more negatively aroused [7], decreased ability to pay attention, to think creatively, to interact harmoniously with others, even weakened immune system and cardiovascular functioning [6]. Thus, alleviating negative emotions can undoubtedly have many beneficial consequences.

### 2.2 Empathy as Emotional Support

Providing emotional support is an effective way to help people overcome any negative affect that they experience [6], and a powerful way to provide such support is through the expression of empathy [8]. Empathy is commonly referred to as the ability to detect what others feel and to experience that emotion ourselves. The value of empathy comes not from understanding others' feelings, but what we do as a result of this. Empirical evidence indicates that expressing accurate empathy can lead to positive psychological, physical and health outcomes (e.g. helping people cope more effectively with problem situations, maintain a positive self-esteem, sense of social inclusion, higher life satisfaction, enjoy better physical health, quickly recover from injuries, better adhere to treatment regimen [6, 9], as well as building trust, and strengthening the relationship between the helper and the hearer

[8]. On the other hand, inept expression of empathy can also lead to many undesirable outcomes [6]. Hence, should we choose to provide emotional support, we must do it effectively.

### 2.3 Emotional Support Using Computers

Klein et al. define two types of emotional support that computer systems can provide: passive and active support [3]. Current computer systems offer many ways of passively helping people to manage their moods and emotions such as playing music, videos, displaying humorous comic strips, allowing people to communicate electronically, or acting as game consoles. While these methods are widely used and effective, people have to actively seek for and initiate these interactions, “*they are not automatically offered by computer systems*” [3]. Some initial attempts have been made to build systems that actively offer emotional support to their users (e.g. chatbots such as ELIZA, computer and robotic pets such as Furby, Tamagocchi, or the famous Sony’s Aibo). However, very few studies have so far investigated how users respond to systems that exhibit emotional-support ability. A study done by Klein et al. [3] showed that displaying empathy via simple text dialogues led to a significantly longer playing time of a computer game in subjects who had been made frustrated by the same game earlier compared to similar text-based interfaces that only allowed users to express their feelings (i.e. venting) or ignored their feelings altogether. However, no significant difference in post-experiment self-reported mood was found. Following Klein, Brave et al. [2] also showed that the subjects’ attitudes toward an empathic agent were improved, which included greater likeability, trustworthiness, as well as greater perceived caring and felt support when compared to its non-empathic counterpart. In his study, both agents were represented by a photo of a human face and capable of expressing empathy via text messages appearing in speech bubbles and photographed facial expressions. Recently, Bickmore and Schulman [1] also extended Klein’s study to compare the effectiveness of two comforting methods. The first method extended Klein’s venting condition by allowing the subjects to use speech-based input (hence offer them an additional mechanism to express their feelings through intonation and prosody), and adding neutral feedback to the subjects’ expressions. The second method replicated Klein’s first condition. In addition, their system utilizes an embodied agent that uses both verbal and nonverbal channel to convey additional information such as emphasis, turn-taking cues, discourse shifts, and affect. The results obtained are somewhat similar to that of Klein’s study. Greater empathic accuracy was more efficacious at comforting the subjects, even at the cost of restricting their input when compared to venting. The study, however, lacked a control condition, hence it is impossible to conclude that the comforting effect was caused by its intervention.

### 3. Research Questions

Despite the promising results of these early researches, a number of questions on how to build such a system remain unanswered. In human-human interactions, empathy is effectively communicated through both verbal and nonverbal channel. When expressing our empathy, our tone of voice, facial expressions, posture, eye movements, arms and legs’ positions are at least as important as what we say. The results of the aforementioned studies indicated some success in adopting these channels, yet failed to answer: (1) how each channel individually impacts the comforting effect and the users’ perception of the system and (2) how they interact with each other.

Can displaying empathic feedback alone help the users feel at ease and improve their attitude toward the system without the presence of speech and other affective non-verbal gestures? Does adding an embodied animated agent increase such effects considering that people can recognize emotions in synthetic speech [10,11] and facial expressions of animated agents of different forms (e.g. robot, on-screen synthetic head) as accurately as emotions exhibited by other people [12,13]?

## 4. EXPERIMENTAL METHOD

### 4.1 Method

The questions raised in Section 3 will be tested in a 2 (modality: ANIMATED vs. NO VISUAL) by 3 (intervention: EMPATHY AND EXPRESSIVITY vs. EMPATHY vs. NON-EMPATHY) between-subjects study, where the subjects are randomly allocated to one of six conditions.

The subjects are led to believe that they are going to take part in a study of verbal skills and web credibility judgement, in which they are asked to complete three tasks. While the first task allegedly measures their verbal skills, the second and the third task measure their judgement of web credibility, and their memory respectively. In fact, the first task is designed to induce a mild negative mood state in the subjects (the mood manipulation task). After completing the first task, the subjects undergo one of the three interventions that are designed to either alleviate their negative mood or ignore it altogether (see 4.3 for more details). The second and the third task measure the effect of mood on the subjects’ judgements and memory, hence consequently act as measurements of their mood after intervention.

### 4.2 Subjects

Eighty-four students and staff of the university (37 males, 47 females; average age: 25.4 years, stdev: 7.9 years) completed the experiment. Subjects were recruited via fliers, advertisements posted on the University news’ webpage, and emails. All subjects received a £5 voucher to compensate for their time and effort.

### 4.3 Procedure

Subjects took part in the experiment one at a time. They were told by the experimenter that this was a study of verbal skills and web credibility judgement, and they would be interacting with an interactive computer agent named Mary who would take them through the study.

In the ANIMATED conditions, Mary is represented by a human like animated agent (with a synthesized voice) (see Figure 1). On the contrary, in the NO VISUAL conditions, Mary has neither a visual representation nor a voice. Mary first interacted with the subjects for a few minutes to introduce herself, to welcome them and to ask for some personal information (e.g. age, gender) and their current mood (mood before manipulation). In the EMPATHY AND EXPRESSIVITY and EMPATHY conditions, Mary’s script was extended with giving feedback on the subjects’ information (e.g. “*I hope you enjoy your study*”) and asking caring and polite questions such as “*Are you sitting comfortably?*” or “*Hopefully, you will get more comfortable as we go along. Before we start, could I please have some of your information?*”. This aimed to build trust and a relationship between the subjects and the agent, and to help them to familiarize with interacting with Mary. The agent then disappeared and left the subjects to complete the mood induction task (the first task).

Upon completion of the task, the agent asked all subjects to rate their mood at the time (mood after manipulation). The agent in the NON-EMPATHY conditions ignored their feelings altogether and asked some distracting questions (e.g. *“have you participated in similar tests before?”*), then instructed them to start the web credibility judgement task, while the agent in the EMPATHY AND EXPRESSIVITY and EMPATHY conditions attempted to comfort the subjects should they feel less positive or encouraged them should they have achieved their goals. The agent first expressed that they recognized the changes in the subjects’ mood, emphasized the worst negative affective state, and offered the subjects a chance to clarify if their feedback was incorrect (e.g. *“I understand it’s not all that fun”, “You seem to feel less positive than before because you did not perform as well as expected. Is that about right?”*). If their judgements were correct, the agents acknowledged the validation of the subjects’ affective state (e.g. *“You’re not the only who feels this way”*) and provided some empathic statements (See 4.6 for more details). In addition, the agent in the EMPATHY AND EXPRESSIVITY conditions also offered the subjects a chance to freely express themselves unless the subjects decided not to talk about it (e.g. *“No, I don’t want to talk about it”*). The agent thanked the subjects for sharing their feelings and before asking for permission to start the web credibility judgement task (e.g. *“Thank you for sharing your feelings with me. Can we start the next task?”*).

In the web credibility judgement task (the second task), the subjects were asked to explore two websites discussed in 4.5 and look for information about *“alcoholism and its consequences”* and judge the credibility of the information after they had finished exploring each site.

Upon completion, they were asked to rate their mood once again (mood after intervention) and to answer some distracting questions (e.g. *“What is alcoholism?”*). They were then instructed to recall any thoughts about each website and identify whether each thought was negative, positive, or neutral (the third task).

Finally, the agent offered the subjects a chance to watch a short happy video to help them feel better. This is a positive mood induction procedure, which acted as one of our mechanisms to ensure all subjects leave the experiment in a stable, neutral, or positive mood.

At the end of the session, the experimenter debriefed the subjects about the actual aims and rationale for the study.

#### 4.4 Mood Manipulation Technique

Our mood manipulation task induced a mild negative affective state by manipulating the difficulty of an analogy test so that the subjects are likely to perform worse than their expectation. The subjects started with an apparently easy trial test consisting of 6 multiple-choice questions with no time limit, which was followed by a considerably more difficult actual test consisting of 24 questions and a time limit of 4 minutes. The questions used in the test are obtained from [alliqtests.com](http://alliqtests.com)<sup>1</sup>. Between the trial test and the actual test, the subjects were asked to set a goal of how many questions they think they will get right in the actual test based on their result of the trial test (e.g. 50 – 75%). Feedback about their performance was given upon their completion of the actual test. The feedback consisted of the number of questions they had answered correctly, a negative (positive) comment about their

<sup>1</sup> See <http://www.alliqtests.com/tests/3/6/>.

performance by comparing their result to their goal (e.g. *“(Sorry!) You have (not) achieved your expected target.”*) and a sheet containing all their answers as well as the corresponding correct answers.

#### 4.5 Materials

Our animated agent, Mary, was designed using Haptেক<sup>2</sup> technology. Mary is a half-body, female embodied agent that is capable of generating facial expressions, hand, arm, and head movements (see Figure 1 below). Mary employed a synthesized American English voice developed by AT&T. We opted for a voice with American accent as we predicted that as most subjects are either of British origin or living in Britain, they would have a higher tolerance for a foreign voice should certain words be mispronounced.

The two websites used in the web credibility judgement task were <http://www.medhelp.org> and <http://www.intelihealth.com>. Both websites provide general health information, and have been verified to have low and high credibility respectively. Their credibility ratings were conducted by the Consumer Reports WebWatch, a non-profit publisher, and the Health Improvement Institute (HII)<sup>3</sup>.

#### 4.6 Design of the Agent

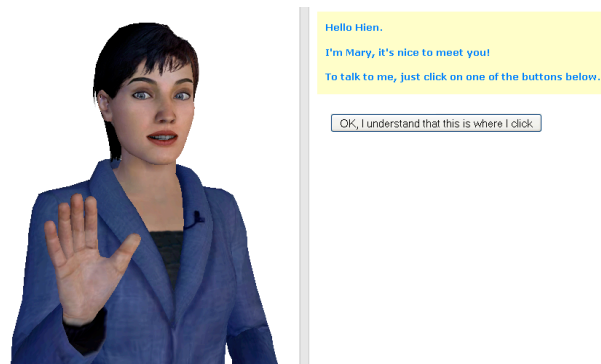


Figure 1. The experiment’s user interface.

The subjects interacted with Mary through the user interface shown in Figure 1 above. The text spoken by Mary was also displayed along with the list of possible replies that the subjects could choose to react to Mary.

Taking findings from research in nonverbal behaviours, a number of considerations were taken into account when designing the animated Mary to ensure her realistic behaviours and perceived warmth and empathy:

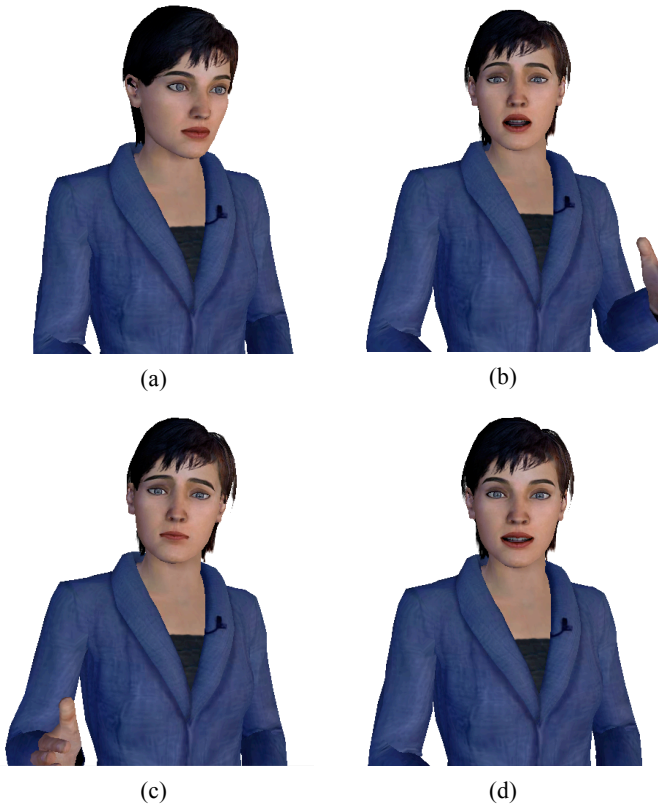
- Mary uses several facial expressions including happy, neutral, concerned, and sad to add affections to her speech.
- Mary keeps eye contact with the subjects while talking, but looks away and toward the answers when waiting for them to choose their reply. This acts as an additional turn-taking cue, and avoids the subjects to feel *“being stared at”*.
- Mary blinks every 3-4 seconds.

<sup>2</sup> See <http://haptেক.com/>.

<sup>3</sup> See <http://www.healthratings.org/>.

- Mary has a moderate open-arm position.
- Mary uses a variety of hand, arm, finger, body (slight forward), and head movements (e.g. nodding) while talking.

Figure 2 below shows some examples of Mary’s behaviours: (a) looking away while waiting for the subjects’ answer, (b) and (c) expressing empathy, and (d) giving encouragement.



**Figure 2. Mary’s various expressions.**

To actively provide emotional support to users, we adopt a number of strategies widely used in Motivational Interviewing [8], sociology, and communication theories, some of which are highly similar to strategies used in [1,3]. The list of strategies used for the EMPATHY intervention are summarized below:

- Using small talk to build trust and relationship between Mary and the subjects (e.g. *“Nothing feels like home, doesn’t it. But I hope the city can keep you happy while you’re here”*). Such technique has been proved to be effective in several studies (e.g. [9]).
- Being polite and friendly. This includes acknowledging the subjects’ opinions of the experiment and asking for permission to start a task or acquire more information when appropriate (e.g. *“I see you didn’t like this test. Hopefully, you will find the next task more interesting and enjoyable. Would you mind if we continue?”*, *“Thank you for sharing your feelings with me. Can we start the next task?”*).
- Acknowledging any mood change in the subjects by actively asking for the subjects’ current affective state at appropriate times (e.g. *I can see you didn’t perform as well as expected.*

*How do you feel right now?*) and comparing it with their previous state (e.g. *“You seem to feel less positive than before (because you did not perform as well as expected). I can see that you are feeling upset, angry, and afraid in particular.”*).

- Offering means to enable the subjects to correct the agent’s judgements in a timely fashion (*“You seem to feel... Is that about right?”*).
- Conveying a sense of sympathy and empathy to the subjects should they feel less positive (e.g. *“I am sorry to hear that things did not go better, but that is completely normal”*).
- Acknowledging the validation of the subjects’ emotional state (e.g. *“You’re not the only who feels this way. Such intelligence tests can make people apprehensive and make them actually perform worse”*).

For the EMPATHY AND EXPRESSIVITY intervention, Mary also encourages the subjects to freely express their feelings (e.g. *“Perhaps you could tell me more about why you feel this way”*). For each expression, Mary replies with a neutral feedback (e.g. *“I understand. Anything else you want to tell me”*).

## 4.7 Measures

### 4.7.1 Measuring affective state

In this study, we employ two different methods to measure the subjects’ behavioural, and subjective reactions simultaneously.

The subjects’ subjective experience was measured using the Short-Form Positive and Negative Affect Schedule questionnaire (PANAS-S) [14], which consists of 5 positive and 5 negative affective adjectives (e.g. *active, upset*). For each adjective, the subjects rate the extent to which they are feeling the corresponding state on a 5-point Likert scale where 1 = *very slightly or not at all* and 5 = *extremely*. The subjects were asked to give their self-rated mood three times: (1) before the mood induction, (2) after the mood induction, and (3) after the web credibility judgement task but before they were asked to recall information about the two websites (i.e. after intervention).

The subjects’ behavioural reactions were measured using their performance on the second and the third tasks. According to the Affect Infusion Model [15], the subjects who experience negative mood should be more critical with their judgements of the two websites’ quality in the second task, and should recall more negative thoughts about the two websites and less thoughts overall than the subjects who experience neutral or positive mood in the third task.

### 4.7.2 Measuring judgements of web credibility

The subjects’ judgements of web credibility were measured using five 7-point Likert scale items where 1 = *not at all* and 7 = *a lot*, which measure the extent to which the subjects find the information on the websites biased, comprehensive, caring, trustworthy, and expert.

### 4.7.3 Measuring attitudes toward Mary

The attitudes of the subjects toward Mary was measured using four 7-point Likert scale items, which indicate how enjoyable, caring, likeable, and trustworthy they find Mary over the course of their interactions.

### 4.7.4 Measuring Mary’s felt empathy

To measure the degree to which the subjects felt empathically understood by Mary, we adopt a scale originally developed by

Barret-Lennard ([4, p.173]). The scale consists of six statements, which describe the degree of empathy from low to high (e.g. “*She totally ignores my feelings*” to “*She appreciates what my experience feels like to me*”).

#### 4.7.5 Measuring the perception of the design of Mary

We also measured the subjects’ opinions of Mary’s language, as well as appearance, gestures, and voice where appropriate. The subjects’ opinions were measured on a 7-point Likert scale where 1 = *I don’t like at all* and 7 = *I like it a lot*.

## 5. EXPERIMENTAL RESULTS

Our results will be presented in five separate subsections, which discuss the followings: (a) our mood induction, (b) the subjects’ perception of our design of Mary, the effect of different interventions and modalities on (c) the subjects’ self-rating moods, (d) the subjects’ judgement of web credibility, and (e) the subjects’ attitudes toward Mary. As gender, nationality, English ability, and occupation of the subjects had neither any main nor interaction effect on our dependent variables, we will not consider these variables any further.

The results of four subjects were removed. Two subjects were removed as many of their measurements were at least 3 standard deviations (stdev) away from the mean relative to other subjects in their group. Further analysis of their comments during the debriefing revealed that they were not comfortable at all with this type of interaction (e.g. “*generally (strongly) dislike these animations*”, “*feel mocked when an (obviously soulless) image is made to pretend to be caring, or feeling and understanding in general*”, “*feel silly for listening to a computer*”). Two other subjects were removed as their positive mood at the start of the experiment was extremely low (both gave the lowest possible rating of 5) in comparison with the other subjects ( $m = 17.30$   $stdev = 4.08$ ). This also ensured there was no significant difference in the subjects’ mood between the conditions before induction. The allocation of the subjects in the six conditions of the experiment is shown in Table 1 below.

**Table 1. The allocation of subjects in the six conditions.**

	Empathy and Expressivity	Empathy	Non-empathy
<b>Animated</b>	13 (6m 7f)	15 (6m 9f)	12 (3m 9f)
<b>No visual</b>	14 (7m 7f)	12 (7m 5f)	14 (7m 7f)

### 5.1 Effectiveness of the mood induction

Our mood induction worked as expected. A one-tailed paired-samples T-test of the subjects’ self-ratings PANAS-S showed a significant decline in positive mood ( $m = -0.65$   $stdev = 3.98$ ,  $p=.05$ ) and a significant rise in negative mood before and after the manipulation ( $m = 1.20$   $stdev = 2.50$ ,  $p<.01$ ). In particular, our manipulation made the subjects feel less attentive and inspired, but more upset, hostile, and ashamed (see Table 2 below for more statistical detail). Indeed, only ten subjects achieved their expectation, but eight of whom admitted that they “*expected to do better*” or “*much better*”, while only four other subjects thought they “*did well enough*” despite not achieving their target. This is further confirmed by the subjects’ comments during the debriefing about how unexpectedly difficult they found the test and how

surprised and frustrated they were with their results especially as they had done so well in the trial test.

**Table 2. Mood change caused by the mood manipulation**

	Change of PANAS		One-tailed paired-t	
	mean	stdev	df	p
<b>Attentive</b>	-0.28	0.99	79	$p<.01$
<b>Inspired</b>	-0.29	1.22	79	$p<.05$
<b>Upset</b>	0.65	1.10	79	$p<.01$
<b>Hostile</b>	0.13	0.58	79	$p<.05$
<b>Ashamed</b>	0.40	1.01	79	$p<.01$

### 5.2 Perception of Mary

Overall, the subjects liked Mary and the language she employed. On a 7-point Likert scale, Mary’s language was rated 5.21 ( $stdev = 1.14$ ), while Mary’s appearance, gestures and synthesized voice were rated 5.00 ( $stdev = 1.30$ ), 4.83 ( $stdev = 1.17$ ), and 5.05 ( $stdev = 1.20$ ) respectively. Further analysis revealed no difference between conditions.

### 5.3 The effect of Mary on the subjects’ self-rated mood

We first examined how the subjects’ mood changed over the course of the experiment by looking at the change of positive and negative mood before mood induction and after intervention. The change in positive and negative mood in each condition is shown in Table 3 and 4 below.

**Table 3. Change in positive mood before mood induction and after intervention.**

	Change in positive mood		
	Empathy and Expressivity	Empathy	Non-empathy
<b>Animated</b>	0.69 (3.71)	-1.40 (2.69)	-2.83 (4.61)
<b>No visual</b>	-0.71 (2.92)	0.67 (3.06)	0.86 (4.37)

**Table 4. Change in negative mood before mood induction and after intervention.**

	Change in negative mood		
	Empathy and Expressivity	Empathy	Non-empathy
<b>Animated</b>	0.08 (1.32)	0.33 (2.77)	-0.67 (3.23)
<b>No visual</b>	-1.00 (1.57)	-1.17 (2.13)	-0.43 (1.60)

Our analysis showed an interaction effect of modality and intervention on the change in the subjects’ positive mood ( $p<.05$ ). Examining each level of intervention, there was no difference between the two modalities for both empathic interventions. For

the NON-EMPATHY intervention, the NO VISUAL agent led to a better recovery of positive mood ( $p < .05$  with Bonferroni correction). There was no difference between different interventions at each level of modality.

So neither of our empathic interventions performed better than the non-empathic counterpart. The addition of an animated agent also did not lead to any improvement. With hindsight, there are two possible explanations for these results: (1) our induced mood did not last as long as we had expected or (2) our non-empathic conditions were not as unempathic as we had planned. As we expected, there was a significant reduction in the subjects' negative mood after intervention, which was measured after the web credibility task ( $m = -1.65$   $stdev = 1.95$ ,  $p < .01$ ), but this could not be credited to the interventions alone since even the subjects in the NO VISUAL, NON-EMPATHY group reported a similar significant reduction ( $m = -1.00$   $stdev = 1.47$ ,  $p < .05$ ). Perhaps, the fact that the NON-EMPATHY agents asked for the subjects' feelings and took them through the experiment might be enough for many subjects. Indeed, further analysis of the subjects' comments suggested the subjects' appreciation for such trivial form of support (see 5.6 for more details).

#### 5.4 The effect of Mary on web credibility judgement

Overall, the subjects gave a significantly higher credibility score to the highly credible website in comparison to its lowly credible one ( $m = 5.25$   $stdev = 1.03$  vs.  $m = 4.12$   $stdev = 1.42$ ,  $p < .01$ ).

We predicted that the subjects' credibility ratings are affected by their mood at the time of judgement and biased in the direction of the prevailing mood. As the web credibility judgement task was designed not to affect the subjects' mood (this was confirmed during the debriefing) and their mood after intervention was measured right after they completed the task, their mood at the time of judgement should be the same as their mood after intervention. Consequently, the subjects' credibility ratings should reflect their mood after intervention. Indeed, their perceived credibility of the two sites was positively correlated with their positive mood after intervention ( $r = .34$ ,  $p < .01$  and  $r = .27$ ,  $p < .05$  for the highly and lowly credible website).

However, we found neither any main effect of modality or intervention nor interaction effect of modality and intervention on the credibility judgement of the two sites (see Table 5 & 6).

**Table 5. Credibility ratings of the highly credible website.**

	Empathy and Expressivity	Empathy	Non-empathy
Animated	5.08 (1.14)	5.21 (1.00)	5.00 (1.17)
No visual	5.34 (1.20)	5.52 (0.71)	5.36 (1.01)

**Table 6. Credibility ratings of the lowly credible website.**

	Empathy and Expressivity	Empathy	Non-empathy
Animated	4.22 (1.13)	3.76 (1.70)	4.30 (1.42)
No visual	3.99 (1.45)	3.90 (1.57)	4.59 (1.29)

The results also revealed some notable behaviour of the subjects with regard to the time spent on each website. The subjects who received either empathic intervention spent more time on reading and took longer to judge each website (see Table 7 for statistical detail).

**Table 7. The mean time (and stdev) spent on each website.**

	Empathy and Expressivity	Empathy	Non-empathy
<b>Time spent on the website with high credibility (seconds)</b>			
Reading	272.8* (154.2)	291.6* (196.4)	162.7 (117.0)
Judging	318.9 (150.9)	320.4 (204.5)	226.2 (143.6)
<b>Time spent on the website with low credibility (seconds)</b>			
Reading	255.7 (221.3)	239.3 (159.2)	227.6 (149.9)
Judging	289.3 (232.9)	271.8 (156.9)	270.0 (152.1)

\* Significant difference at the .05 level (Bonferroni correction) between the corresponding intervention and the NON-EMPATHY intervention.

Considering we encouraged the subjects to spend as much time as they wish to carefully read and judge each website and there was neither a minimum nor maximum time restriction, these findings show that the subjects who received either empathic intervention either could pay more attention to the task or they were more willing to continue interacting with the system and to perform what was asked of them more dutifully.

#### 5.5 The effect of Mary on the subjects' recall of information

Overall, the subjects recalled significantly more positive thoughts than negative thoughts ( $m = 5.14$ ,  $stdev = 2.60$  vs.  $m = 3.93$   $stdev = 2.38$ ,  $p < .01$ ). This is in accordance with the tendency of people likely to see others in a positive light [16]. In particular, the subjects recalled significantly more positive thoughts for the website with high credibility ( $m = 3.25$   $stdev = 1.80$  vs.  $m = 1.33$   $stdev = 1.71$ ,  $p < .01$ ) and significantly more negative thoughts for the website with low credibility ( $m = 2.60$   $stdev = 1.67$  vs.  $m = 1.89$   $stdev = 1.77$ ,  $p < .05$ ).

We expected the subjects with better mood to recall more thoughts overall, and in particular more positive and less negative thoughts. However, our results show only an interaction effect of modality and intervention on the total number of thoughts recalled ( $p < .05$ ) such that for the EMPATHY intervention, the subjects who received the NO VISUAL condition recalled significant more thoughts than subjects who received the ANIMATED condition ( $m = 12.17$   $stdev = 4.17$  vs.  $m = 8.73$   $stdev = 3.65$ ,  $p < .05$ ). This is inline with the results of their self-rated mood reported in 5.3.

#### 5.6 The subjects' attitude toward Mary

Attitude toward Mary is calculated as the average of Mary's perceived enjoyment to interact with, trustworthiness, caring, and likeability. Overall, the more positive attitude the subjects formed toward Mary, the more empathic Mary was perceived ( $r = .40$ ,  $p < .01$ ), and the more positive mood change occurred before mood induction and after intervention ( $r = .33$ ,  $p < .01$ ) as well as before and after intervention ( $r = .28$ ,  $p < .05$ ).

Our analysis showed a significant effect of intervention on the subjects' attitude toward Mary ( $p < .05$ ). Post-hoc analysis using Tukey's HSD revealed that the subjects who received either empathic intervention found Mary more enjoyable, caring, trustworthy, and likeable than those who received no intervention (see Table 8 for more statistical detail).

**Table 8. Attitude toward Mary.**

	Empathy and expressivity	Empathy	Non-empathy
<b>Enjoyable</b>	5.63* (0.84)	5.37 (1.31)	4.69 (1.54)
<b>Caring</b>	4.85 (1.49)	5.33 (1.30)	4.46 (1.56)
<b>Trustworthiness</b>	4.52 (1.67)	5.04 (1.61)	4.27 (1.87)
<b>Likeable</b>	5.56 (1.05)	5.48 (1.53)	4.65 (1.60)
<b>Overall attitudes</b>	5.14 (1.02)	5.31 (1.21)	4.52 (1.44)

\* Significant difference at the .05 level (Bonferroni correction) between the corresponding intervention and the NON-EMPATHY intervention.

There was also a significant interaction effect of modality and intervention on Mary's perceived trustworthiness ( $p < .05$ ), and Mary's felt empathic understanding ( $p < .01$ ).

For the EMPATHY AND EXPRESSIVITY intervention, the ANIMATED agent was found significantly more empathic ( $p < .01$ ), trustworthy ( $p < .05$ ), as well as more enjoyable, caring, and likeable than the NO VISUAL modality, and consequently led to a significantly more positive attitude toward the agent ( $p < .05$ ) (See Table 9,10 for more detailed statistics). For the other two interventions, no significant difference was found between the two modalities.

For the ANIMATED agent, both the EMPATHY and EMPATHY AND EXPRESSIVITY interventions were perceived to be more enjoyable, caring, trustworthy, and likeable than the NON-EMPATHY intervention (see Table 9 below).

**Table 9. Attitude toward the ANIMATED agent.**

	ANIMATED Agent		
	Empathy and expressivity	Empathy	Non-empathy
<b>Enjoyable</b>	5.92 (0.86)	5.40 (1.45)	4.75 (1.66)
<b>Caring</b>	5.31* (1.03)	5.27* (1.53)	3.92 (1.17)
<b>Trustworthiness</b>	5.23 (1.48)	5.20 (1.82)	3.67 (1.88)
<b>Likeable</b>	5.92 (0.95)	5.67 (1.54)	4.58 (1.73)
<b>Overall attitudes</b>	5.60* (0.90)	5.38 (1.39)	4.23 (1.43)

\* Significant difference at the .05 level (Bonferroni correction) between the corresponding intervention and the NON-EMPATHY intervention.

For the STATIC agent, there was no difference between interventions with respect to the subjects' attitude toward Mary (see Table 10). But her perceived empathy was significant lower in the EMPATHY AND EXPRESSIVITY intervention than that

in the EMPATHY ( $p < .01$ ) or NON-EMPATHY ( $p < .05$ ) interventions (see Table 11).

**Table 10. Attitude toward the NO VISUAL agent.**

	NO VISUAL Agent		
	Empathy and expressivity	Empathy	Non-empathy
<b>Enjoyable</b>	5.36 (0.75)	5.33 (1.16)	4.64 (1.50)
<b>Caring</b>	4.43 (1.74)	5.42 (1.00)	4.93 (1.73)
<b>Trustworthiness</b>	3.86 (1.61)	4.83 (1.34)	4.79 (1.76)
<b>Likeable</b>	5.21 (1.05)	5.25 (1.55)	4.71 (1.54)
<b>Overall attitudes</b>	4.71 (0.96)	5.21 (0.97)	4.77 (1.46)

Analysing the subjects' comments further showed that our NON-EMPATHY conditions were not perceived as bad as we had planned. The subjects' attitude toward such conditions was still strengthened by the fact that the agents took them through the experiment and employed an informal language, consequently made everything easy to understand and follow (e.g. "*She explained everything in a way I could easily understand*", "*Easy, informal tone – makes it clear what has to be done*").

Clearly, our data showed that giving empathic feedback led to a more positive attitude toward the system, including greater likeability, trustworthiness, perceived caring, and enjoyment to interact with regardless of the modality of delivering such content, although the difference was somewhat more profound when the system is represented by a human-like agent. Interestingly, the ANIMATED agents are better than the NO VISUAL agent when they behave more human-like and vice-versa. We speculated that people build up a higher expectation of the resemblance of system's behaviours to our owns when the system is represented by a human-like agent. Indeed, the ANIMATED, NON-EMPATHY agent provided the worst positive mood recovery as well as the lowest attitude toward the agent.

**Table 11. Mary's felt empathy.**

	Empathy and Expressivity	Empathy	Non-empathy
<b>Animated</b>	4.69 (1.49)	3.93 (1.67)	4.00 (1.65)
<b>No visual</b>	2.64 (1.34)	4.67 (1.23)	4.29 (1.59)

## 6. DISCUSSION

In this paper, we systematically investigated two important factors and their interaction in the design of empathic computers: (1) different ways of expressing empathy, and (2) the modality of delivering such content. Our results lead to a number of important implications:

### (1) The positive attitude of the users toward active support

Our findings (5.6) suggested that people appreciate even the smallest form of help. Systems that can actively support the users

to achieve their goals by just simply taking them through the process can lead to a more positive attitude toward the systems.

### **(2) The positive outcome of an empathic system regardless of its representation**

Our findings (5.6) extend the result obtained by Brave [2] where an agent with empathic emotion was found to lead to more positive ratings of the agent when both agents were represented by a photo of a human face. Our results showed that expressing empathy improves the liking of the system regardless of its representation. However, if the system exhibits empathic understanding, it is best to use an embodied conversational agent.

Consequently, an empathic system could improve the users' attention and willingness to engage with a task (5.4). Considering the strong correlation between the subjects' positive mood change and their attitude toward the system (5.6), an empathic system is also likely to be able to help the users feel at ease.

### **(3) A system represented by a human-like representation is expected to be empathic.**

When a system has no visual representation, our results (5.6) suggested that people do not seem to care or expect whether such a system can understand and care for their feelings. However, when a system is represented by a human-like agent, its lack of empathy can lead to a negative user experience, and worsen the users' attitude toward the system. This is perhaps because we have learned not to expect such ability from a standard computer system, however the novelty of interacting with a human-like agent causes us to orient more socially toward the system.

### **(4) Adding a mechanism to allow the users to freely express themselves does not improve the effectiveness or liking of an empathic system.**

Both studies done by Klein and Bickmore [1,3] found that allowing the users to freely express their feelings either through the traditional keyboard-based input or the more novel speech-based input did not provide any positive effect on the users' behaviour or affective state. Similarly, our results showed that adding such mechanism along with neutral feedback does not lead to any improvement when compared to giving empathic feedback alone. It is, therefore, more beneficial to restrict the users' inputs to obtain more accurate empathic feedback.

## **7. CONCLUSION**

In summary, our study showed the potential of computer systems that can actively provide emotional support to their users. In the design of such systems, being able to give accurate empathic feedback is the most important. Such ability can improve the users' attitude toward the systems. Additionally, adding an embodied agent can help to improve the systems' liking and perhaps their effects on the users. However, embodied agents should be avoided if the system is not designed to be empathic or if the system does not want to build up a high expectation of its ability to deal with affect.

## **8. ACKNOWLEDGMENTS**

This research is partly supported by the Nuffield Foundation Grant No. NAL/00258/G. We would also like to thank Filip Drozd for his suggestions on Mary's scripts, and Dr Louise Philips for her suggestions on the experiment's method.

## **9. REFERENCES**

- [1] Bickmore, T., Schulman, D. 2007. Practical approaches to comforting users with relational agents. Conference on Human Factors in Computing Systems (CHI), San Jose, CA.
- [2] Brave, S., Nass, C., Hutchinson, K. 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, 62,161-178.
- [3] Klein, J., Moon, Y., Picard, R. 2002. This computer responds to user frustration: Theory, design, results, and implications. *Interacting with Computers*, 14, 119-140.
- [4] Vincent, S. 2005. *Being Empathic: A Companion for Counsellors and Therapists*. Radcliffe Publishing.
- [5] Stein, N. L., Hernandez, M. W. 2007. Assessing understanding and appraisals during emotional experience: The development and use of the narcoder. *Handbook of Emotion Elicitation and Assessment*. Oxford University Press.
- [6] Greene, J., Bureson, B.R. 2003. *Handbook of Communication and Social Interaction Skills*. Lawrence Erlbaum Associates, Inc., Publishers.
- [7] Wilson, E., MacLeod, C., Campbell, L. 2007. The information-processing approach to emotion research. *Handbook of Emotion Elicitation and Assessment*. Oxford University Press.
- [8] Miller, W. R., Rollnick, S. 1991. *Motivational Interviewing*. The Guildford Press, New York.
- [9] Bickmore, T., Picard, R. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer Human Interaction*, 12, 2, 293-327.
- [10] Cahn, J.E. 1990. The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, 8, 1-19.
- [11] Nass, C., Foehr, U., Brave, S., Somoza, M. 2001. The effects of emotion of voice in synthesized and recorded speech. In *Proceedings of the AAAI Symposium: Emotional and Intelligent II: The Tangled Knot of Social Cognition*, North Falmouth, MA.
- [12] Schiano, D.J., Ehrlich, S.M., Rahardja, K., Sheridan, K. 2000. Face to interface: Facial affect in (hu)man and machine. *Proceedings of CHI'00: Conference on Human Factors in Computing Systems*, 193-200.
- [13] Massaro, D.W., Cohen, M.M., Beskow, J., Cole, R. 2000. Developing and evaluating conversational agents. In: Chassell, J., Sullivan, J., Prevost, S., Churchill, E. (Eds.), *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 287-318.
- [14] Thompson, E. R. 2007. Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38, 2, 227-242.
- [15] Forgas, J. P. 1995. Mood and judgment: the affect infusion model (AIM). *Psychol Bull.*
- [16] Forgas, J. P. 1985. *Interpersonal behavior: The Psychology of Social Interaction*. Oxford: Pergamon Press.