

Computer  
Response to  
User Frustration

Also referenced as: **MIT Media  
Laboratory Vision and Modeling Group  
Technical Report  
TR#480**

Note: This technical report has been updated as of 11/22/00 to  
accommodate a minor change to Appendix A.

Jonathan T. Klein

Bachelor of Science in Computer Graphics and Human-Computer Interaction  
University of Michigan, Ann Arbor  
June 1996

Submitted to the  
Program in Media Arts and Sciences  
The MIT School of Architecture and Planning

In partial fulfillment of the requirements for the degree of  
Master of Science in Media Arts and Sciences, Massachusetts Institute of  
Technology

February 1999

© Massachusetts Institute  
of Technology, 1998.  
All rights reserved.

---

Author Jonathan T. Klein  
Program in Media Arts & Sciences  
September 9, 1998

---

Certified by Rosalind W. Picard  
Associate Professor of Media Technology  
MIT Media Laboratory

---

Accepted by Stephen A. Benton  
Chairperson  
Departmental Committee on Graduate  
Students  
Program in Media Arts and Sciences



# Computer Response to User Frustration

Jonathan T. Klein

Submitted to the  
Program in Media Arts and Sciences  
on September 10, 1998, in partial fulfillment of the  
requirements for the degree of Master of Science in Media Arts and Sciences

## Abstract

Use of computer technology often has unpleasant side effects, some of which are strong, negative emotional states that arise in humans during interaction with computers. Frustration, confusion, anger, anxiety and similar emotional states can affect not only the interaction itself, but also productivity, learning, social relationships, and overall well-being. This thesis presents the idea of designing human-computer interaction systems to actively support human users in their ability to regulate, manage, and recover from their own negative emotional states, particularly frustration. This document describes traditional theoretical strategies for emotion regulation, the design of a human-computer interaction agent built by the author to actively help relieve frustration, and an evaluation that shows the effectiveness of the agent.

A study designed to test this agent was conducted: A system was built that elicits frustration in human subjects. The interaction agent then initiated several social, emotional-content feedback strategies with some of the subjects, in an effort to help relieve their emotional state. These strategies were designed to provide many of the same cues that skilled, human listeners employ when helping relieve strong, negative emotions in others. Two control groups were exposed to the same frustrating stimuli, one of which was given no emotional support at all; the other enabled subjects to report problems and "vent" at the computer. Subsequent behavior was then observed, and self-report data was collected.

Behavioral results showed the agent was significantly more effective than the two controls in helping relieve frustration levels in subjects. These results demonstrate that strategic, social, emotional-content interaction with a computer by users who are experiencing frustration can help alleviate this negative state. They also provide evidence that humans may benefit emotionally in the short term from computers that respond in socially appropriate ways toward their emotions. The implications of this work suggest a wholly new role for computers in human life. Sociological ramifications of this new role are also discussed.

Thesis Supervisor: Rosalind W. Picard  
Title: Associate Professor of Media Technology



Computer  
Response to  
User Frustration

Jonathan T. Klein

Thesis readers

---

Sherry Turkle  
Professor of Sociology  
Program in Science, Technology, and Society  
MIT

---

Clifford Nass  
Associate Professor  
Department of Communication  
Stanford University

---

Youngme Moon  
Assistant Professor of Business Administration  
Harvard Business School  
Harvard University



## ACKNOWLEDGMENTS

I would like very much to thank Rosalind Picard for her tireless efforts, keen insight and seemingly limitless patience with me in this effort specifically, and in my Media Lab tenure in general. Apart from being the most wonderful academic advisor I could imagine, she is also a great person, an honorable person, and a privilege to know in her own right.

I would also like to especially thank Youngme Moon, whose vision, brilliance, expertise and timely help in the study's conceptualization, protocol design and statistical methods helped to make this thesis project what it is. Youngme, I simply could not have done this study without you, and I will always be in your debt.

Sincere gratitude is also extended to Cliff Nass, who also provided key, timely insights and guidance to this study and overall project, and to Sherry Turkle, who helped me start to piece together what this project's implications might actually mean in people's lives. To her, my other readers, and Roz: Thanks for giving me a shot, for the privilege of your company and your insight on this project, and for believing in me.

Special thanks go to Jocelyn Riseberg Scheirer, whose advice, guidance, and example helped shape this thesis in meaningful ways, whose prior work laid some of the groundwork for this study, and whose friendship and collaboration made it all much easier to do.

A debt of gratitude goes to Professor Keith Oatley, who generously read my proposal, and came down from Canada to sit down with me, and give me valuable advice and insight into the project. I would also like to thank Terry Winograd, for reading the proposal and giving me great comments, and for his wonderful warmth. Thanks also go to Dr. Phyllis Sanchez, clinical neuropsychologist extraordinaire, who gave me critical, grounding feedback early on from a living, breathing clinician—and to Louis Tassinary for putting me in touch with her.

And, of course, thanks goes to the incomparable Matthew Norwood, for his great, invaluable help (not to mention good nature) in the early, critical stages of the project, and for building the first version of the experiment software.



Thanks to Lorin Wilde for helping to shape the proposal, and to Matt Gorbet and Margaret Bradley, for their help in assembling critical materials. Thanks are also due to professors Jerome Kagan, Francine Benes, and Kurt Fischer at Harvard, who listened and gave feedback. Thanks also to Eric Trimble for providing the spurious transceiver.

Thanks must invariably find their way to Jennifer Healey, who with Jocelyn served as an independent coder in my data analysis. Apart from that bit of timely help, you are a real friend, and you make me laugh. Thanks also to Raul Fernandez, for his friendship, willingness to code my data and/or roll sushi on command, and for putting up with me as an officemate while I wrote my thesis.

I also would like to thank Scott Brave for his insight, friendship, and company as my officemate my first year at the Lab. His values and perspective are reflected in the focus and scope of this thesis. Thanks also to Arthur Nasson and Shannon Guthrie for their humor, support and relief during this process, as well as for repeatedly feeding my wife in my absence.

Thanks are also due to Kate Mongiat, Martín Hadis, Shari and David Metcalf, Sunil Vemuri, Kelly Conway, Doug Willen, Kathi Blocher, Laurie Ward, Teresa Marrin, Tanzeem Choudhery, Tom Minka, Martin Szummer, B. J. Fogg, Linda Peterson (of course), and all of Vismod for their warmth and support during this process.

And, of course, thanks to Elliot Soloway, Judy Olson, Chris Quintana, Susanne Jul, Jeff Spitulnik, Michael Bailey and all the other wonderful folks at Michigan who helped get me to this point. And thanks to Tom Atwood, whose key support, wit and wisdom I value and appreciate.

Many, many thanks and gratitude are due to Ellen's mother and father, Mrs. Marilyn Petri and the late Henry Petri, Sr., for all their love, support and timely help during this process. Thanks also to all of my own loving, supportive parents. You are many, but you are all very dear. Special thanks go to my mom, Sandra Addor, for igniting my interest in psychology and emotions in the first place. And, of course, thanks to Donald Kaplan (a.k.a. Grampa), who has been very supportive, throughout the process (and he even waded through the lengthy thesis proposal).

Most of all, though, I wish to express my deepest gratitude, love and thanks to my wonderful wife Ellen. Without her immense support, tireless (and sleepless) care for our daughter, and inexhaustible supply of fresh perspective (on demand, no less), I could and would never have done any of this.

And thanks to little Hannah, who always seems to have a smile for her daddy.

*For my biggest benefactors  
(mothers and shakers all)*

*Ellen,  
Sandra,  
Marilyn,  
and especially Lynne,  
who didn't have to*

*For doing so much, so quietly, for so long*



## TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b>		<b>7</b>
<b>CHAPTER 1</b>		<b>15</b>
Introduction		
<b>CHAPTER 2</b>		<b>19</b>
Background		
Emotion Theory	19	
<i>Functions of Emotions</i>	19	
<i>The Nature of Emotions</i>	20	
Frustration Theory	21	
Emotion Regulation	22	
<i>Active, Social Approaches to Emotion Regulation</i>	23	
<i>One Active, Social Technique: Active Listening</i>	23	
Computer Systems that Frustrate	25	
<b>CHAPTER 3</b>		<b>27</b>
Related work		
Computer Systems that Help Relieve Frustration	27	
Computer Systems that Support Emotional Communication		28
<i>Eliza</i>	28	
<i>Computer Psychotherapy Programs</i>	28	
<i>Computer Pets, and Related Research</i>	28	
Active Computer Support for Emotion Regulation		30
<i>Complaint Sites: Active, Computer-Mediated, Human Support</i>	30	
<i>Active Machine Support of User Frustration?</i>	30	
Media and Social Interactions		31

<b>CHAPTER 4</b>		<b>33</b>
	<b>Theory and objectives</b>	
	Applying Active Listening Theory to Interface Design	34
	Applying Other Strategies for Emotion Regulation Support to Interface Design	36
	Criteria and Objectives for the Study	37
	<i>A Model for the Protocol Design</i>	38
	<i>The Model Adapted</i>	38
	Evaluation Criteria, Predictions, & Protocol Overview	39
<b>CHAPTER 5</b>		<b>43</b>
	<b>Method</b>	
	Participants	44
	<i>A Note on Subject Attrition</i>	44
	Apparatus	46
	<i>The Physical Set-up</i>	46
	Software	47
	<i>Technical Design of the Game</i>	47
	Description of the Game	49
	Design of Interactive Questionnaires & CASPER Agent	51
	<i>Controlling for the "Conversational Quality"</i>	51
	CONTROL Condition	52
	VENTING Condition	52
	CASPER Condition	55
	CASPER Repair Dialogs	56
	Procedure	57
	<i>Recruiting Subjects</i>	57
	<i>Off-site Subject Contact</i>	57
	<i>Motivating the Subject</i>	58
	<i>On-Site: Running the Subject</i>	58
	Game 1	60
	The DELAY Condition	60
	The NO-DELAY Condition	62
	<i>The Interactive Questionnaires (Response Conditions)</i>	62
	<i>All Subjects Play the Same Game 2</i>	62
	<i>Exit Questionnaire &amp; Debriefing</i>	63
	Measures	64
	Data Treatment	65
	<i>Video Record Not Considered</i>	65
	Data Cleaning	65
<b>CHAPTER 6</b>		<b>67</b>
	<b>Results</b>	
	Frustration - Manipulation Check	67
	Game-Playing Behavior	67
	Satisfaction Level After Game 2	70
	Game-Playing Behavior, Controlling for Gender	70

	Game-Playing Behavior, Controlling for Emotional Arousability	70
	Game-Playing Behavior, Controlling for Gaming Experience Level	70
<b>CHAPTER 7</b>		<b>71</b>
	<b>Discussion</b>	
	Discussion of Primary Results	71
	<i>Manipulation Check Successful</i>	71
	<i>Significant Behavioral Results Found</i>	72
	Unexpected Results	73
	<i>A Rebound Effect for Game 2</i>	73
	<i>A Floor Effect Found for Frustration Level</i>	74
	<i>Low Scores (and Variance) for VENT Conditions</i>	75
	<i>Little Statistically Meaningful Self-Report Data</i>	75
	<i>One Interpretation: The More Frustrated One gets, the Longer One Plays</i>	76
	<i>Other Alternate Interpretations of These Results</i>	77
	<i>Possible Targets</i>	79
<b>CHAPTER 8</b>		<b>81</b>
	<b>Theoretical and Practical Implications</b>	
	Theoretical Implications	81
	<i>Implications of Principles for Designers</i>	82
	Practical Implications	83
	<i>An Illustrative Scenario From Real-Life Corporate America</i>	84
	Further-Reaching Design Implications	85
<b>CHAPTER 9</b>		<b>87</b>
	<b>Sociological and Philosophical Implications</b>	
	Soothing, Salving, Modeling: For the Best?	88
	<i>Main Effects</i>	88
	<i>Modeling and Meeting Human Needs</i>	88
	Heartstrings and Pursestrings: For the Worst?	89
	<i>Emotion Management Assistance vs. Emotion Manipulation</i>	89
	<i>Trusting the Image in the Mirror...</i>	90
	... and <i>Trusting the Reality the Mirror Reflects</i>	90
	<i>Computers as Friends: The Next Generation?</i>	90
	<i>Sapping Needed Wind From Sails</i>	91
	<i>Absent Empathy</i>	92
	<i>Synthetic Sympathy:</i>	92
	Situating This Work in Context	92
	<i>What This Approach is Not: Computer Psychotherapy</i>	93
<b>CHAPTER 10</b>		<b>95</b>
	<b>Future Directions</b>	
	<i>Determining Effective Components of the Present Study.</i>	95

	<i>Double-Blind Validation of this Experiment</i>	96	
	<i>Going Public</i>	96	
	<i>Repeat Performances?</i>	96	
	<i>Determining Deeper Effects of this Genre</i>	96	
<b>CHAPTER 11</b>			<b>97</b>
	<b>Conclusions</b>		
<b>APPENDIX A:</b>			<b>99</b>
	<b>Off-Site Subject Experience Materials</b>		
	The Soliciting Flier	99	
	Informational Email	100	
	Confirmation and Questionnaire	101	
	Questionnaire Receipt Confirmation	103	
	Email Reminder the Night Before	104	
<b>APPENDIX B:</b>			<b>105</b>
	<b>On-site Subject Experience Materials</b>		
	Briefing script	105	
	Consent form	108	
	On-screen instructions, Game 1	110	
	Directions for answering questionnaire (all conditions)		112
	Directions for Game 2	113	
	Exit questionnaire	114	
	Debriefing statement	117	
<b>APPENDIX C:</b>			<b>119</b>
	<b>Data Analyses and related documents</b>		
	Frequency distribution of subject age ranges		119
	Manipulation check	121	
	ANOVA: Main behavioral measure	123	
	Gender analysis	126	
	Trait Arousability analysis	128	
	2nd Trait Arousability Analysis	130	
	Game-Playing Experience Analysis	132	
	Frequency Dichotomy analysis	133	
	Game 2 performance analysis	136	
	Analysis of time played: CONTROL vs. VENT		138
	Analysis of time played: CASPER vs. VENT		139
	Analysis of time played: CASPER vs. CONTROL		140
	Self-report on frustration level, after questionnaire		142
	Analysis of self-report for questionnaire effectiveness		143
	Analysis of satisfaction levels, Game 2	144	
	Coding directions for independent judges, part 1		145
	Coding directions for independent judges, part 2		146
	Responses to On-Line, Open-Ended Question		147
<b>REFERENCES</b>			<b>151</b>
<b>BIOGRAPHICAL NOTE</b>			<b>156</b>

## CHAPTER 1

# Introduction

What can and should your computer do when you get frustrated while you use it?

The solutions commonly heard by this author are to try to determine and fix the problem that is making you frustrated, or better yet pre-emptively try to prevent the problem from happening in the first place. Indeed, the latter solution has been the traditional, if tacit, view of the Human-Computer Interaction (HCI) community for years (e.g. Norman 88). The problem with these approaches, beyond serious feasibility problems, is simple:

They are almost always insufficient.

Indeed, all of these approaches ignore the user in an important way: They tend to leave the user still feeling frustrated, and still feeling negatively toward the system.

Let's examine the feasibility issues first. As for the pre-emptive approach, computer software and hardware designers can indeed do much to make their systems easier for people to use, doubtless mitigating many potential sources of frustration in the process. It's also evident that computers are growing in numbers and shrinking into embedded systems, quietly moving into more and more venues of modern life. However, human beings have complex, busy lives, only portions of which may involve working directly with computer systems. Indeed, what frustrates us may or may not have anything to do with a computer system.

The ability to pre-empt frustration in users becomes an impossible task when the source of frustration could be the computer system as easily as it could be a fight with a spouse the night before, or a traffic jam on the way in to work. Further, even



sources of frustration that arise in interaction with these self-same computer systems (incompatibility between newer operating system versions and existing software, for example) are increasingly emergent, from the fast-growing complexity of modern technology alone.

To build a system that is able to accurately determine the *cause* of the user's frustration is an equally daunting task. First, it would require nearly constant surveillance of the user, and *at least* strong artificial intelligence (AI)—knowledge, intelligence, memory, insight, and reasoning ability equivalent to that possessed by an intelligent human being. The term “at least” is emphasized, since humans are themselves notoriously bad at determining the object of their own emotions, let alone the emotions of others (Zajonc 1998).

Humans are even worse at being able to *fix* the problems that can bring on frustration, since so many potential elicitors in modern life (information overflow, other people, and the speed of the Web, for starters) are beyond one's immediate, personal control. Perhaps not as daunting a task as trying to prevent emotional elicitors before they occur, this approach is nonetheless just as impractical, requiring incredible amounts of intelligence, not to mention power. If our own mothers had trouble finding and fixing the cause of our frustrations when we were two years old, what makes us think that computers can do better?

Beyond its lack of feasibility, another major flaw in the “find and fix” approach to addressing frustration is the fact that it actually does nothing to address a critical aspect of the situation: the emotions themselves. Once a person is experiencing frustration, in place of one problem there springs at least two: The original elicitor (or elicitors), as well as the state of frustration itself.

With elevated levels of adrenaline and other neurochemicals coursing through their body, the person feeling frustration suddenly has diminished abilities for attention (Kitayama 94), memory retention (Kahneman 73), learning (Lewis 89), thinking creatively (Isen 97), and polite social interaction (Goleman 95), among other things—as well as a penchant for getting more frustrated in the immediate future.

Frustration, then, can impair a person's ability to find either the eliciting problem, or solutions to it. Also, a frustrating episode in interaction with a computer system can leave the user feeling negatively disposed toward the system itself. If humans have a penchant for treating computers as if they were other people, as has been recently demonstrated (Reeves 96), such negative experiences can alter perceptions of trust, cooperation and good faith on the part of the user. Worse, such experiences can injure what, to users, appears to be one of their most important working relationships.

If these approaches show little hope of effectively addressing user frustration, what, again, can and should the computer do?

Perhaps a more fruitful (and maybe more effective) way of answering the original question is to rephrase it: What, if anything, can the computer do to address the emotional state you are experiencing, rather than addressing the eliciting problem?

This thesis presents evidence that computer systems can do a great deal to support users when they experience frustration, and much more than they have hitherto done. Further, evidence suggests that such systems can leave the user

feeling much more positively disposed toward the computer system—the *same* system that, previously, had been the perceived source of the user’s frustration in the first place.

In the course of this thesis project, the author has designed and built a relatively simple interaction agent in software that has been demonstrated to actively and significantly help relieve frustration in human subjects. This software agent not only validates the theory used to create it, but represents a dramatic leap in the ability of computer systems to directly and actively support a key aspect of human life: Their emotions. Specifically, the system described herein provides active support for the user’s own ability to manage and regulate his or her emotional states, particularly frustration, to the user’s personal advantage, and with a number of side benefits.

But how does such an agent work? What would this agent be like to interact with? Consider the following interaction between a user and a computer, occurring in an office in the not-too-distant future:

**User:** Argh!! I can’t get this thing to print! Oh God. I’m not going to make it.

**Computer:** Having problems?

**User:** Yes! I’m late for a deadline, and I can’t get this stupid file to print. Grrr...

**Computer:** Geez, you’re really frustrated right now, huh?

**User:** You got that right. If I don’t get this file to print, I may not get my paper into the conference on time.

**COMPUTER:** Well, I’m just a dumb computer, but I’m sorry you’re in such a spot. It sounds really difficult.

**USER:** Thanks. Gee, this printer’s obviously hosed. Hmm, maybe I can print this thing to another printer that’s not being used...

Here’s another approach to the same situation:

**USER:** Argh!! I can’t get this thing to print! Oh God. I’m not going to make it.

**COMPUTER:** Having problems?

**USER:** Yes! I’m late for a deadline, and I can’t get this stupid file to print. Grrr...

**COMPUTER:** Hmm. You sound frustrated. You know, Shelly and Bob both experienced frustration while using this printer yesterday.

**USER:** Really. (leans out the door). Hey is Shelly or Bob still around?

**BOB:** Yeah, I’m here, what’s up?

**USER:** I’m having trouble getting output from Printer 4.

**BOB:** Yeah, I think the page sensor is flaky. Send it again as max color output. That fixed it for us. We’ll get it repaired on Monday.

**USER:** Great! That worked. Thanks! (turns to monitor) And thank *you*.

The two scenarios above have several things in common. They are examples of strategies for human-computer interactions during times when the user is experiencing emotional duress: The user may be feeling some combination of frustration, anger, anxiety, and/or other emotions, some or all of which may be directed at the computer system itself. In both cases, however, the computer's responses actively acknowledge and address the user's emotional state. Neither strategy uses advanced artificial intelligence (AI) in the service of its goals, and neither strategy seeks to try to determine or address the *cause* of the user's emotional state. Both strategies aim to support the user in managing his or her feelings, using principles known in a number of communities, from social psychology to crisis management, from parenting theory to consumer affairs.

The strategies illustrated above may not prove to be the most effective ones for the situation, and some elements of these approaches are obviously questionable: The computer referring to itself as "I", for example, or the computer offering that it's "sorry" for the user. Such approaches may be found to mislead the user as to the computer's capabilities, or present the computer as "caring" when no such state exists in the machine. For that matter, the ability of the computer to discern frustration is still a research question, and the computer's report of the emotional experiences of others (as in the second scenario) raises privacy issues. Some of these approaches may also simply prove ineffective in helping the user to manage her own feelings, and recover from strong, negative emotional states. Further, we may find that different personalities may prefer different kinds of responses from the machine (Reeves 96), and that the strategies illustrated above may be inappropriate for the user in question.

Still, the very idea that both systems are able to address and respond to the user's emotional state represents an important departure in HCI, both in research and in practice. Beyond some work at the periphery of the field, current theory in HCI reflects a lack of direct consideration for the user's emotional state, despite occasional lip-service to the issue. When reading current HCI literature, for example, it seems doubly strange that this important aspect of the user's experience continues to be largely ignored by this community.

Underlying all HCI research and design is the belief that the people using a computer system should come first. Their needs, capabilities and preferences for performing various activities should inform the ways in which systems are designed and implemented. People should *not* have to change radically to 'fit in with the system', the system should be designed to match their requirements.

*From Human-Computer Interaction, by Preece, et al, 1994, p. 15*

Computers are beginning to be recognized as being capable of relating to their users on more than just an information-processing level. Clearly, indeed, humans are much more than information processors. Humans are affective beings, motivated to action by a complex system of emotions, drives, needs, and environmental conditioning (Myers 89). The extent to which computers are able to respond to this vast, albeit ill-charted, aspect of humanity may well be the extent to which computers are truly able to "match the requirements" of their users.

A major focus of inquiry in the Affective Computing Research Group at the MIT Media Laboratory is addressing this mismatch in human requirements, and exploring how computer systems may be designed to support the fact that humans feel, as well as think and work. A particular challenge in this area is the problem of designing systems that meaningfully respond to user frustration (Picard 97).

This thesis explores solutions to this problem, and focuses on one promising solution: The design and construction of a software agent prototype that is designed to help relieve frustration and related, negative affective states in users, using human social communication as a model. It then describes a study that was conducted to test this agent. The results of this study show strong behavioral evidence that this device was successful in achieving its goals. A discussion of a number of side benefits for this new genre of HCI ensues, as does a frank discussion of some of the risks involved in developing and using such a device.

The approach used in this thesis deals almost exclusively with the emotional event—the user’s perceptions of her experience of the event, and the affective state that ensues in the user—rather than the eliciting event itself. In other words, this work explores ways to address the *state* of the user, not the event or events that elicit that state. By contrast, charging a computer system with figuring out *why* a user is frustrated or otherwise upset (let alone figuring out how to solve such problems, either pre-emptively or in real time), is akin to solving the problem of domain-independent machine intelligence: It remains an open research question, and is certainly beyond the scope of this thesis.

However, the work described herein demonstrates that active, computer support for the user in managing his or her own emotional state is an effective and worthwhile means of improving that state, thereby increasing productivity and satisfaction with the computer system. To describe how this approach works, however, a brief review of the literature on emotion theory, frustration theory, and emotion regulation is in order.

## CHAPTER 2

# Background

## Emotion Theory

Emotions have historically presented a vexing problem for researchers in psychology, since nearly everything about them, save their expression, are internal processes, hidden from view. Recently, however, neuroscientists such as Joseph LeDoux (LeDoux 94) and Antonio Damasio (Damasio 94) have made research inroads into mapping neuroanatomical structures and processes that mediate emotional functioning in the human brain. And, despite fundamental open questions remaining about emotions and their interrelationship with memory, attention, cognitive processes and brain function, much has been learned, especially in recent years, about how and why emotions work.

### *Functions of Emotions*

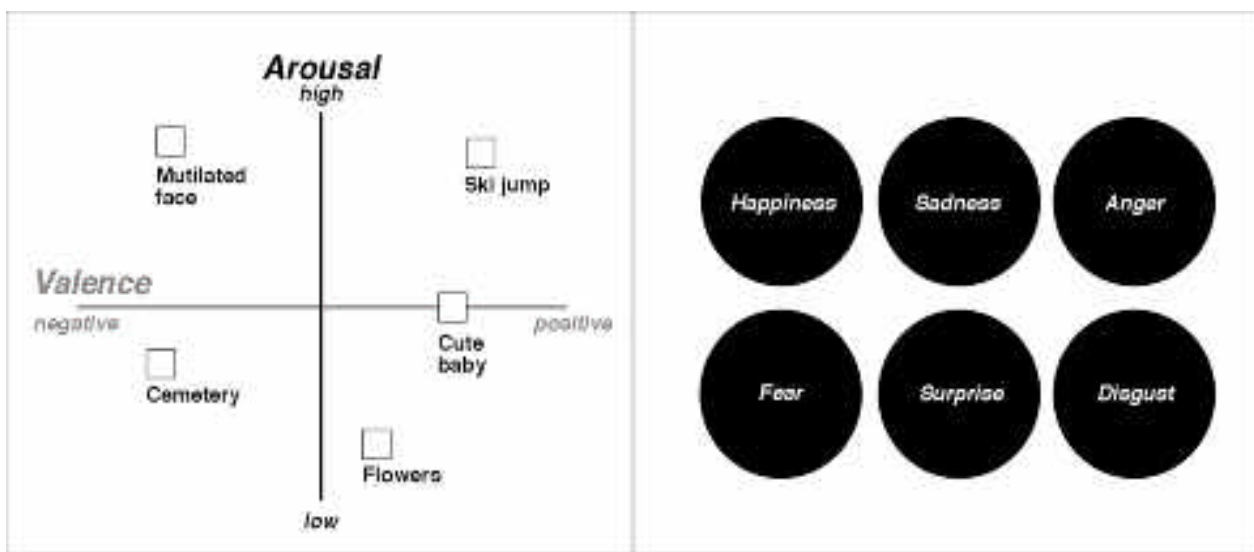
Literature on emotion theory (Zajonc 98, Fridja 86, Izard 90) identifies a number of possible functions for emotions. Although there are many more proposed functions for emotions described in the literature, three such functions relevant to this project are listed below:

- Emotions are descriptive. Emotions act as barometers to internal state, external predicament and progress in meeting needs and goals. For example, fear can act to encourage a person to retreat from a perilous situation; likewise, a sense of happiness and peace can mean an individual has no pressing, unmet needs or goals. In this sense, emotions, together with drives such as hunger and fatigue, act to let the individual know what is going on both within and around her.
- Emotions motivate. Humans seem to possess emotions at least in part to help them meet needs and goals (Tomkins 62), and keep them from harm (Zajonc 98). Emotions apparently motivate us to action via several neural response paths that work at different levels of the brain and at different speeds, which seem to correlate with different strategies for dealing with our environment (LeDoux 94). Sudden, sharp pain in an extremity yields an automatic reflex reaction. Certain stimuli such as sudden, looming objects yield a quick response of surprise, an “early-warning system”, which in infants prompts them to freeze (Kagan 94). For more complex situations, however, emotions often act as motivational “fuel” for an individual’s ability to meet needs and accomplish goals. Anger and frustration can motivate a person to confront impediments to her needs and goals she feels she can overcome. Fear can act to help keep the individual safe from potential dangers. Fear and other emotions can serve as motivators to positive action as well.
- Emotions communicate. Emotional expression is often described as social and communicative by nature (Averill 90). Humans are, after all, fundamentally

social beings: Infants rely completely on others to meet their needs at birth and throughout early childhood, and continue to rely on others to help meet their needs to varying degrees throughout life. It seems clear that a primary function of emotion is to communicate state information to others, in order to enable them to assist in meeting the needs of the individual. In the course of development, humans also need to learn a great deal from others about how to negotiate with their environment, and even how to negotiate with themselves. *Emotion regulation*, the ability to manage one's own emotional states (see Section 2.3, below), for example, is widely believed to be learned from others—primarily through the *attachment relationship*, the bond between child and primary caregiver (e.g. Brennan 95). Emotion regulation is also thought to be learned through modeling, seeing how others express and manage their own emotions. For many reasons, then, communication of state information is a vital aspect of emotion.

### The Nature of Emotions

The precise nature of emotions remains the subject of dispute among emotion theorists. Some (e.g., Lang 90) describe emotions as *continuous*. In this scheme, emotions are conceptualized as being placed in a multi-dimensional space of emotional attributes. A popular conception of continuous emotions is a 2-dimensional plane (see Figure 1.1A), in which one vector of this plane is emotional valence—whether the emotion is positive or negative, and to what degree—and the other vector is emotional arousal—how stimulated or excited a person is. Other theorists (e.g. Ekman 92) argue for a model of emotions that are *discrete* (see Figure 1.1B): Five, six or more identifiable emotional states such as *happiness*, *sadness*, *anger*, *fear*, *surprise*, *disgust* and *no emotion* are defined; one's emotional state is generally described as primarily one of these states.



Figures 1.1 A and 1.1B: Continuous vs. discrete emotions. Figure A (left) is a 2-dimensional diagram of continuous emotions, in

which valence—whether the emotion is positive or negative, and to what degree—is the horizontal axis, and emotional arousal—how stimulated or excited a person is—is the vertical axis. Pictures of various stimuli are placed on the 2D plane according to common reactions people have to them (Picard 97). Figure B (right) is a simple model of discrete, basic emotions (e.g. Ekman 92), that are described and treated as distinct from one another.

### Frustration Theory

Most of what we think of as emotions seems to share properties of both continuous and discrete conceptions. Indeed, for our purposes, *frustration* may be viewed as either a discrete state, related to anger, or as a strong (highly aroused), negative emotional state on the continuous scale.

Research on frustration was begun by behavior theorists in the 1930s; hence, much of the early work in this field involves experiments with animal behavior. Lawson, after Rosenzweig (Lawson 65), defines frustration as “the occurrence of an obstacle that prevented the satisfaction of a need.” ‘Need’ in this case can be interpreted to mean either a need or a goal. Quoting from Riseberg, et al. (Riseberg 98), frustration may be thought of as

an increase in arousal subsequent to a frustrating event. This kind of frustration is referred to as unconditioned or primary frustration, in which there is a hypothetical unconditioned reaction to the frustrating event. This acts to create an effect on the response with which it coincides, the immediate consequence of which is a short-term increment in generalized, energizing drive or arousal (Amsel 92). Primary frustration, in this view, has an affective or emotional component.

In behaviorist theory, frustration has also been associated with aggression, suggesting an action/reaction behavioral loop. Although animal behavior observed by behaviorists can hardly be said to reflect human behavior, the connection between frustration and anger (and related states such as hatred and aggressiveness) is clear. Indeed, Oatley showed that frustration is the most common precursor to, and possibly elicitor of, anger (Oatley 94).

If left alone, arousal level tends to decay slowly, such that removal from the eliciting environment into a serene one may, in time, bring one’s arousal level down; the individual consequently experiences calming down. Frustration also tends to be cumulative: When one is already experiencing frustration arousal, it can take very little in the way of further frustrating stimuli to exacerbate arousal levels through the roof. Further, depending on the person’s current state, stimuli may appear more or less frustrating: If a person is positively aroused, for example, stimuli that would normally frustrate her might not bother her at all. Similarly, if a person is *already* frustrated, even stimuli that do not, ordinarily, elicit frustration may contribute to increased frustration arousal (Amsel 90).

In practical life, as social and emotional beings, humans often find themselves in situations in which they’ve “gotten their Irish up”, yet the solution to a problem is not surmountable, or deemed not worth surmounting. A person experiencing

frustration may find herself feeling exasperated, yet knowing that her frustration will not help accomplish her goal.

In this common situation, out of a person's grappling with one problem (the situation), there grows a second: A strong, negatively-valenced emotional reaction. At this point, the person may be said to have two challenges to confront: the stimulating problem *and* the emotional state. Often, the emotional reaction can prove even more of an obstacle to the person than the initial problem itself. However, humans possess skills and strategies for *emotion regulation*, which can mediate frustration arousal levels to varying degrees.

### Emotion Regulation

Humans routinely employ *emotion regulation* throughout their lives (Gross 95). Emotion regulation is at once an aptitude and a skill for modulating and managing one's emotional state: From making sure that lows don't get too low or highs too high, to keeping anger in check, this critical skill allows humans to generate emotional responses that are appropriate to changing situations (such as "keeping one's cool" in the face of upsetting stimuli, and soothing oneself when such stimuli have passed). What's more, emotion regulation provides humans with self-discipline and motivation to achieve their goals (Mayer 95). Emotion regulation, also described as emotion self-management, has been identified as a primary component of *emotional intelligence*, a set of essential emotional skills, the development of which have been argued to correlate more with a person's success in life than IQ (Goleman 95, citing Vaillant 77 and Felsman 87).

Humans are able to manage their own emotional states to varying degrees of success, depending on the situation, the person's temperament (Kagan 94), degree of emotion-management skill, and specific approach, among other things. "Managing our emotions is something of a full-time job: much of what we do—especially in our free time—is an attempt to manage mood." (Goleman 95, p. 57) Failure to manage one's emotions can have profound effects: These can range from decreased productivity (due to anxiety, for example), to an inability to pay attention and learn, to injury of cooperative human relationships (for example due to expressing unmodulated anger), to increased personal stress and severe depression (Gross 95), even to addiction (Cooper 95) and other health problems (see Chapter 11 of Goleman 95 for an excellent overview).

People use a variety of methods to help manage their emotions that include internal strategies (positive thinking and optimism, for example), as well as external ones: interacting with media and/or other people, engaging in activities, and using foods and other substances such as alcohol, tobacco, and other drugs. One such strategy is simply expressing, or venting, one's feelings, often at the original stimulus—be it a person or object, such as a computer. There is split opinion over psychological benefits of unconstrained venting (or *catharsis*, as it's called in the psychology literature) when angry, however, as Goleman outlines (Goleman 95, pp. 64-65). Some emotion researchers argue that catharsis tends to have a calming effect, while others suggest that venting can actually serve to exacerbate ones' arousal, making one *more* angry.



I define two varieties of support for emotion management: *passive* supports, and *active* ones. *Passive* supports are those that are used by people to manipulate moods, without addressing or discussing the emotions themselves. These include media, activities, food and other substances. Often, interactions with other people fall into this category too: Engaging in activities such as sports with others is a strategy for emotion management: playing the game and interacting with others serves to help a person moderate her emotions, yet the emotions themselves may not be discussed. However, particularly in cases of strong, negative emotional states such as sadness, anxiety, or frustration, people often turn to other people—often some kind of support structure of family and/or friends, if available—to help them manage their emotional states. In this case, other people serve as *active* emotional regulation supports. Active support occurs when people are able to address and discuss their emotions and their elicitors directly, as a means of managing them, instead of simply trying to manipulate them indirectly. Talking to a parent or caring friend about what is upsetting, and how that makes the person feel, is an example of engaging active support.

#### ***Active, Social Approaches to Emotion Regulation***

Active emotional support using other people serves a multitude of positive purposes, for example as an impetus for social bonding and building friendships, or honing emotional self-management skills. A net effect of such social interaction in times of emotional tumult is that the upset person is often able to overcome his frustration, calm down, feel “better”, and get on with life. Further, the ability to use others to help manage one’s own emotions is also regarded as an important component of interpersonal communication. Consider the following cases:

- A commercial airline pilot describing the reasons for a projected delay to passengers, as well as his heartfelt appreciation for, and apology to, those who will be affected. This approach is often described as “managing expectations”, but managing expectations can be a way of managing emotions;
- The development of emotion regulation skills in the crying child who is able to be soothed and calmed by an attachment figure (Bowlby 69);
- Children, adolescents and adults who play team sports, and look to their coach to instill confidence in them through pep talks, “psyching them up”;
- People of all ages who, when greatly emotionally aroused (from sadness, fear or frustration, for example), seek interactions with others to comfort them, sympathize with them, and “make them feel better.”

#### ***One Active, Social Technique: Active Listening***

*Active listening* (e.g. Myers 89, Nugent 95) may be described as providing sincere, non-judgmental feedback to an emotionally upset individual, with a particular focus on providing feedback of the emotional content itself. Active listening has its roots in Social Psychology—specifically, in Rogerian, “Person-Centered” psychotherapy (e.g. Raskin, N. J., and Rogers, C., Ch. 5 in Corsini 95), although psychotherapists from many other schools have adopted it for their use. One need not be a

psychotherapist to practice this skill, however. Research in parenting theory (e.g. Gordon 70) argues persuasively that children who are experiencing strong, negative emotions often experience quick relief and recovery in response to active listening. Active listening has also been found to be effective when used with adults and children alike (Myers 89). It is used in such diverse areas as crisis counseling (such as MIT's *Nightline* crisis management hotline), education (Dusenbury 97), consumer affairs (e.g. Jenks 93) and other corporate venues (Kubota 97).

Indeed, active listening is an effective strategy for relieving strong, negative emotional states that, when done correctly and under appropriate circumstances, can diffuse and relieve such states quickly and effectively. An example of active listening, and its potential effectiveness, comes from the literature on parenting theory (Gordon 70, p. 78):

- JOHNNY:** Oh, I burned my finger! Mommy, I burned my finger. Ow, it hurts, it hurts [crying now]! My finger got burned. Ow, ow!
- MOTHER:** Oooo, it really hurts. It hurts terribly.
- JOHNNY:** Yes, look how bad I burned it.
- MOTHER:** It feels like it's burned real bad. It hurts *so* much.
- JOHNNY** (stops crying) : Put something on it right away.
- MOTHER:** Okay. I'll get some ice to make it cooler, then we can put some ointment on it.

Active listening may be described as consisting of a number of attributes that contribute to its efficacy in a wide variety of situations (Myers 89, Roid 94). These components may be described as enabling the individual encountering emotional distress to:

- feel “heard,” that another person now has knowledge of the individual’s predicament, *without judgment*;
- feel that his predicament and feelings are understood;
- feel “attended to”;
- feel empathized with;
- feel, in the presence of sincere feedback, that strong emotions are what everyone experiences, and that such emotions are “recoverable from”;
- feel that the individual’s emotional state is normal and acceptable for the situation.

Perhaps the key, “active” ingredient of active listening, however, seems to be paraphrasing, emotional-content feedback: The sense that the emotional state has been effectively communicated (Nugent 95). Indeed, if it is true that an important reason for communicating emotion to others is to convey information about the current state of the individual, then feedback confirming accurate reception of this state information should provide some measure of relief.

The effects of active listening, or even a sympathetic ear or reassuring gesture, can quickly help turn a negative emotion around (Cassidy 94). Yet humans seem to be spending decreasing amounts of time in human-human interactions. For example, apparently three times as many people live alone today as did fifty years ago (Myers 93). In place of human contact, many people seem to be spending increased amounts of time in interaction with computers—at school, at work, and at home. And the quality of interaction with computer systems, as well as the quality of interaction with others *through* computer systems, as many of us know, is often less than smooth.

### Computer Systems that Frustrate

A tacit aim of user-centered design is almost always to try to minimize the amount of frustration a user experiences when using a system (Norman 86). Yet, despite the best efforts of the user-interface design community, as well as huge gains in speed by the computer hardware industry, user frustration may be growing rather than shrinking. Indeed, more and more people are using computers, and using them for longer hours. While this trend continues, computer software systems are growing more sophisticated and feature-rich in order to compete for market share, yet market pressures insist that products be shipped with shorter and shorter development cycles (e.g., Preece 94). Complex help systems and thick user manuals provide little relief. In the words of Henry Petroski, the famous chronicler of invention and innovation, “The most amazing achievement of the computer software industry is its continuing cancellation of the steady and staggering gains made by the computer hardware industry.” With some exceptions, computers seem to be getting harder, not easier, to learn and use.

It's not just local computer systems that contribute to user frustration, either. With increased usage of the Internet, its fallibilities such as slow network traffic speeds as well as incomplete and hard-to-find resources on the World Wide Web (WWW) become more and more of an issue to users in an age of mounting pressure for increased productivity. All things considered, it seems as though users are destined to experience increasing frustration with their computer systems, despite the best efforts of UI designers and others.

Since efforts to keep user frustration from happening in the first place seems to be a losing battle in light of these circumstances, it seems logical for HCI researchers to do three things:

- 1. Acknowledge that users are often experiencing frustration with their computers today, and that there is little evidence that these experiences will abate in the near future.
- 2. Continue to work hard to try to minimize and/or eliminate eliciting conditions for user frustration;
- 3. Begin to explore design strategies that may enable computers to help relieve the frustrations that they help elicit. One promising strategy may be to leverage

knowledge from the social sciences about human-human interactions to inform human-computer interaction design (Reeves 96).

The project described herein represents a decisive step toward emotional human-computer interaction. This thesis argues that active, computer support of the management of human emotional states such as frustration is not only possible but practical, and sorely needed. What's more, active emotion regulation support strategies should prove to be effective in communication to the user while s/he works, and may require much less time than passive supports. The evidence discovered as a result of this thesis project strongly supports these arguments, and is presented in Chapters 6 and 7.



## CHAPTER 3

## Related work

## Computer Systems that Help Relieve Frustration

Certainly, current computer systems are capable of offering a number of tools that can aid humans in managing their emotions. Some may be described as passive supports, in that they are used by people to manipulate moods, without addressing or discussing the emotions themselves: Multimedia computer systems are capable of playing user-selected music, displaying humorous comic strips and animated movies, and engaging the user in a wide variety of interactive games. All such media have well-known, pre-computer roles in helping to manage emotions (Brennan 95, Catanzaro 94, Goleman 95), as people have used them routinely to escape, unwind, and change moods. Modern, networked computers also allow users to communicate with other people via email, Internet Relay Chat (synchronous text communication; each reads what the other types with only a slight delay), and other means, enabling users to enlist other people to help manage emotions.

All of these media can be used to support users in managing their emotional states, yet they are often time-consuming, and people have to *seek* these interactions—they aren't automatically offered by the system. Further, the ability of computer-related media to be used in an emotionally-supportive role is currently limited in practice: Corporations as a rule frown on employees playing games, reading comics, exchanging email and watching television on company time. Further, email and other text-based interaction methods offer limited emotional bandwidth, losing channels of expressiveness found in face-to-face, visual, tactile and audio interaction, often making it difficult to communicate emotion effectively to others. And, despite the theoretical availability of passive supports, computers *in and*

*of themselves* currently offer no active support for personal emotional regulation in humans. To our emotional lives, machines are literally deaf and dumb (Picard 97).

#### Computer Systems that Support Emotional Communication

While computers have yet to be built for the specific task of actively supporting a user's emotional regulation, many systems have been built that are at least able to communicate with users in ways that involve their emotions, in one form or another. There are many and varied means of emotional communication between humans and computers; below are several examples that suggest the design space.

##### **Eliza**

*Eliza* was a system built by Joseph Weizenbaum at MIT in the late 1960s (Weizenbaum 66). *Eliza* communicates to users in natural language via a textual exchange, in which users type in a sentence and *Eliza* responds. The system employed simple techniques for interaction, some of which were borrowed from Rogerian active listening (see Section 2.3.2, above). *Eliza* presented the illusion of an intelligence in the system, able to reply automatically with in many cases what seemed like meaningful, intelligent responses to a wide variety of things a user might say, including some words and phrases that are typically emotionally charged. Weizenbaum achieved this illusion of intelligence using some simple word parsing strategies, and word matching algorithms. His well-publicized stated purpose for building the system was solely to explore natural language processing, yet *Eliza* is often appreciated for the illusion it tenders: Many people have “chatted” with *Eliza* since its wildly popular introduction, some for its entertainment value, and some for its research value. Others, however, seem to use *Eliza* as something of an emotional support system (Ch. 4 of Turkle 95). Perhaps as a means for catharsis, some users talk to the system as if it were somewhere between a diary and a human sounding board, able to elicit emotional content from the user's input, and engage in dialog about the user's emotions.

##### **Computer Psychotherapy Programs**

In *Life on the Screen*, Sherry Turkle wrote about the phenomenon of computer psychotherapy programs, software with which one might converse in natural language (like *Eliza*), in the cause of bringing about positive change for the user's psychological health (Turkle 95). One such system, DEPRESSION 2.0, released in 1992, sought to bring about such change via a series of regular conversations, in natural language, with the user. The therapeutic model of the system is based on cognitive psychology: Altering negative behavior and thought patterns is a matter of providing the right information to the user about these thoughts and behavior, in effect helping to “reset” the user's “defaults” for thinking and acting. In her book, Turkle presents a diatribe against cognitive therapy approaches, and the computational tools that support them. Nevertheless, while systems like Depression 2.0 were not designed to engage exclusively in emotional-content interaction, the emotional component of interactions with a system charged with performing psychotherapy cannot be ignored. Depression 2.0 and packages like it may therefore

be said to use natural language interaction to influence the user's emotions, whether directly or indirectly. These systems are controversial, and will be discussed further in Chapter 9.

### **Computer Pets, and Related Research**

Other computer systems have also recently been built that support some form or another of emotional connection with users. The recent phenomenon of computer pets such as Catz and Dogz from P.F. Magic, Inc., and last year's Tamagocchi made by Bandai, Inc., does seem to foster some form of emotional communication. All three of these "pets" enable users to "pay attention" and give affection to them, either via stroking virtual bodies or by playing games with virtual creatures. All three provide emotional feedback, telling users that they are "pleased" or "displeased" with the interaction. As diversions, these "pets" arguably fall into the category of passive supports for emotional regulation, and perhaps a bit more. Since, for example, all three products yelp, meow or beep for attention when their algorithms have determined that they "have not gotten enough attention recently", they seem to require users to fulfill the role of emotional regulation supports for *them*. Whether or not these computer creatures help to model emotional regulation, or assistance thereof, for young users, remains a matter for future inquiry.

A new generation of interactive toys with varying degrees of emotional communication with their users have either appeared on the market, or are planned to appear in the near future. ActiMates' Barney (Strommen 98) communicates affect to its young users, primarily in the form of affection and apparent attention to the child. Mattel is releasing an interactive Winnie the Pooh with many similar capabilities, and Tiger Electronics is scheduled to release the "Furby" later this year. The Furby is apparently designed to have an unique kind of relationship with its user, but the specific details are somewhat sketchy (see Kirstner 98). It is doubtful, however, that any capability for addressing the user's affective state (and certainly negative affect) has been designed into these systems.

A number of other companies and research laboratories have recently been working on projects that involve emotional expression, emotional communication or emotion synthesis. Omron Corporation has developed a robotic "cat" that is capable of cowering or purring in response to perceived stimuli from users. Similarly, Yamaha Corporation and The MIT Robotics Laboratory have been working on a robotic dog that features an on-board emotion-synthesis system of its own, with means for emotional expression and some sensory capability. A number of projects at the MIT AI Laboratory are also focusing on aspects of human-computer (robotic) emotional communication as well as building "social skills" in robots (Scassellati 98, Breazeal 98).

Other systems have been built that attempt emotional communication with the user, by means of still-fairly-simple recognition of a user's emotion, and computer-aided expression of its "own" emotions. Clark Elliot's *Affective Reasoner* (Elliot 90) can interact with humans multimodally (via text, audio and graphics) with some emotional appraisal ability—with a built-in model of emotions, the system can "reason" about and appraise a person's emotional state—occasionally exhibiting what seems like empathic ability. Elliot's system, however, was aimed at achieving a working computational model of emotions, and creating multimodal interfaces for it,



for which smooth interaction with humans was the goal. Elliot was interested in making a system that could relate to humans on an emotional level; toward that end, the Affective Reasoner shows promise of achieving some level of emotional sensitivity in interaction with users. The system's goal was not emotional management support, however, but simply emotional interaction.

MIC and MUSE (Tosa 96) also attempt multimodal emotional interaction, albeit with a much cruder algorithm for it. Both systems attempt to simply mirror what it identifies as the user's current emotional state (one of six discrete states), the former via three-dimensional (3D) graphics and sounds, and the latter with 3D graphics and music. Similarly, Hara and his team have built a female robotic head that can both recognize the discrete emotions of fear, happiness, surprise, sadness, anger, and disgust, and express them. The main goal of Hara's work is apparently to create robots that will "empathize" with humans and make them feel "more comfortable as they read emotional changes expressed in our faces" (Strauss 97). The Affective Reasoner, MIC, MUSE and Hara's robot all may be said to involve emotion regulation strategies, and show potential as platforms for active emotion regulation support; however, none currently offer such support.

## Active Computer Support for Emotion Regulation

### ***Complaint Sites: Active, Computer-Mediated, Human Support***

Several Internet sites have been created where users angry or frustrated with products and services can go and complain in public. Sites include Internet newsgroups such as alt.unixhaters and alt.fan.billgates, and websites such as "miCROsoFT sucKS! aND theSE peOPLE agREE!" (<http://www.wamzle.com/blink/msft/msftagree.html>) Such sites offer users catharsis, the ability to publicly complain about their frustrating experiences and products. Often, other users with similar experiences will chime in and reinforce the original user's complaint. Although these sites enable users to address their frustrations specifically by engaging in cathartic, public venting, these systems may not offer the most effective strategy for emotional support, due to their asynchronous nature, and the mixed opinion on catharsis, or "venting" (see the section on Emotion Regulation in Chapter 2, above, as well as the results of this thesis study in Chapter 6, below, which includes a "venting" condition). Further, these interactions must be initiated by users, and often require considerable time and effort—to make a convincing case in writing against a product, for example.

### ***Active Machine Support of User Frustration?***

In MATLAB 3.5, an older version of a high-level mathematics software package produced by MathWorks, Inc., a user could type in the popular expletive "f---" at the software's command line, and the application would respond with a relevant quip, randomly chosen from a list of quips, such as "Your place or mine?" This functionality has since been disabled in later versions of the software, and a sober-toned note to that effect has replaced the witty response. A colleague recently told me of an undergraduate student at MIT named "Bob" who, while working in MATLAB 3.5 on a graphing problem set, became frustrated to the point of despair

over problems with MATLAB 3.5's syntax. In a fit of pique, unaware of the software's aforementioned response capability, Bob unwittingly typed in the word "f--" at MatLab's command line. When the machine responded with "Your place or mine?", Bob mood changed instantly. From utter frustration came surprise and delight. Here was a witty, unexpected "hack", a response to a word considered taboo in formal elements of this culture, a response programmed long ago by one or more of the software's developers, unpublicized (obviously) in the documentation. It did absolutely nothing to address the problem that elicited Bob's frustrated state, yet this strong, negative state suddenly dissolved. Instead, Bob became intrigued by the phenomenon, wanting to investigate further. What happened when he entered the word in again? Could other words elicit similar responses? Bob played around with the feature for a while, contacting friends to share the revelation of this hidden treasure, and then settled back down to work.

Hypotheses for the sudden turnaround in Bob's disposition abound: The experience was unexpected, humorous, ribald, playful, irreverent. He probably expected nothing, and got something—or discovered something—that in itself may be seen as an unexpected "reward". The interaction's "designers" probably gave little thought to planning this response, yet the fact remains that in the midst of experiencing a strong, negatively-valenced emotional reaction to a seemingly intractable situation, Bob made a bid to communicate this state to "another" (the software on his computer), and got a reaction. The machine's response was dumb—it was one of several simple strings pre-programmed into the system to write to the screen in response to a user typing in a certain string.

However, for Bob it was also a direct, immediate, relevant response to his frustrated emotional state—at or near a peak of frustration—from "another" to whom Bob expressed his feeling state (albeit crudely). Arguably, this simple hack produced some of the same components of feedback as are found in active listening (see Section 2.3.2, above): Immediate feedback of emotional information, acknowledgment in relevant language (albeit loosely interpreted here), as well, perhaps, as conveyance of a message like "your negative emotional state is understood and accepted, and here's a humorous remark that says 'it's okay'". One may argue that this event represents an example of purely passive emotional management support (i.e. "it was just a diversion."). The possibility also exists that some of Bob's amusement may have come a sense of connection to the *programmer*, and not the program. Yet the intriguing possibility remains that components of active support may have been at least in part responsible for Bob's sudden change of mood, as well as the coincident interest in playing with the system—and soon settling back down to work with MATLAB again.

A human-computer interaction design has yet to be demonstrated, however, that maintains as its explicit goal the real-time, active support of a user's management of his or her emotional state. This thesis is an investigation of the questions: Can machines actively and effectively assist in the management of emotional states in human users, particularly strong, negative emotional states? What would a user-interface design for such a system look like? How might such interactions affect the user, and the user's approach to the system?

## Media and Social Interactions

Recent work by Clifford Nass and Byron Reeves at Stanford University suggests that people display a natural propensity for relating to machines as if they were other people (Reeves 96). In over 35 different experiments, Reeves and Nass showed that humans relate to many forms of media, especially interactive media, with many of the same social rules and expectations they have for dealing with other people. Their studies revealed that users respond to computers and other media as if they had personalities, genders, emotions, politeness sensibilities, collaborative natures, and professional roles, among other things. The extent to which this “Media Equation” applies (i.e. how far humans appear hard-wired to relate to computers and other media socially) is not yet known. However, Reeves and Nass show that humans appear willing to accept machines in uniquely social activities such as being praised—one of their experiments involved building a spell-checker that praises good spelling (Ch. 4 of Reeves 96)—even when there is no real substance behind the role that the computer plays. Users were shown to prefer a system that makes them feel praised, with virtually no other indication that the system “cared” about the user, or knew what it means for the user to feel praise.



## CHAPTER 4

# Theory and objectives

Driving hypotheses of this thesis are as follows:

- 1) The Media Equation (described at the end of the last chapter) is a phenomenon that extends to other, more personal areas of social interaction such as assistance with emotion regulation. Further, this phenomenon may involve strong emotional states, and discussion of the emotions themselves with users while they are experiencing them;
- 2) A computer system can help relieve frustration in a user by means of active support of emotion regulation, hitherto assumed to be a uniquely social function;
- 3) A computer system employing active support techniques can make a frustrated user feel better, and do so very quickly—even when the computer is the source of the problem in the first place, and even when the machine is not capable of empathy or feeling.
- 4) Such a system can be built using existing technology, without requiring strong AI.

The approach is to look at ways of addressing what may be a human social need, namely the need to feel that one's emotional state is acknowledged and heard—that one's feelings are understood and accepted, and by association, acceptable—especially when one is upset. If the system makes it plain that it does not actually understand, but is still able to acknowledge emotional content by immediate feedback akin to a primitive form of active listening (see Chapter 2), can these responses help the user manage his or her emotional state? Might it do what skilled

human listeners can often do, which is to quickly help the user to change the emotional state, to feel more at ease, more comfortable, and less agitated? What is necessary to help make people feel better, and is this something computers can and should be in the business of doing?

To begin to answer these questions, it is first necessary to be specific about what computer systems can and should do in the service of providing this kind of support—what this author will hitherto refer to as Computer-aided, Active Support for Personal Emotion Regulation, or CASPER.

To inform the strategies that the CASPER agent uses, literature from a number of different disciplines was examined: Social psychology, developmental and clinical psychology, sociology, education, emotion theory, parenting theory, communications, research in empathy and crisis management, as well as principles from consumer affairs literature in industry. The goal of this interdisciplinary research was to discover successful, effective strategies for facilitating personal emotion regulation via social and emotional “triggers” that humans seek from others (Myers 89). The author sought to determine such triggers, as well as those that humans may be predisposed to accepting from media (Reeves 96).

If humans relate to computers socially, as if they were other people—as the Media Equation suggests—then people may be able to use computers the way they use other people: as at least partial “active listeners”, able to actively support humans in emotion management.

### Applying Active Listening Theory to Interface Design

If the above hypotheses are correct, then components of active listening can be designed into human-computer interaction, to provide active supports for emotion management. In order to build such supports, I propose the following guidelines for designing active-listening-style interaction into human-computer interfaces. When engaging in active computer support for emotion regulation, the system must:

- Actively solicit information about the user’s state, especially the user’s emotional state and feelings about the eliciting problem (not just a cognitive assessment of the problem). This query serves in part to acknowledge that the user *has* emotions, and may be having an emotional reaction to stimuli.
- This solicitation must be timely; i.e. it must occur when the user is upset enough to have need of (and benefit from) the support that the system has to offer. Active listening is known to be ineffective in situations in which a user simply wants information, or wants to communicate information (Gordon 70). At the very least, it is important that the system try to ensure that engaging in emotion-content dialog is of interest to the user, and is relevant to the user. In terms of timing, there are two promising strategies: The first is to address the emotional state then and there, when the person is experiencing the emotion for the first time. The second is to address the emotional state sometime after the fact (minutes or hours later, or perhaps longer). Since addressing the user’s arousal before it has a chance to mushroom is an obviously useful strategy, the former approach is probably the most interesting and most relevant for future systems. Still, since frustration arousal typically involves a long decay rate, and secondary

elicitors can often make one feel even more frustration, people who are frustrated can often stay aroused for long periods. Considering the fact that often when one is frustrated, a time constraint on a task may be involved, that instant might not be a good time to provide support (i.e. interrupt the task). The latter strategy is therefore viable, if perhaps less desirable in many cases.

- The initiation of the supportive dialogue may come either from the user’s own initiative (when upset, s/he may just “feel like talking about it”, and remember how helpful or beneficial the system was for them last time), or the system may proactively initiate the dialogue with the user. The system would require some means of sensing (or predicting with high probability) that the user is upset, whether from the suggestion of others or through an automated sensing system of its own (Fernandez 97).
- Make sure the user is able to express what she is feeling. It does no use to solicit emotional feedback from a user who is feeling confused and then provide him with choices that read “I am: very sad | sad | neutral | happy | very happy about [product X].” Instead, provide:
  - Appropriate labels;
  - Recognition vs. recall of names for relevant emotional states (the user may or may not have the vocabulary at hand for describing her state);
  - A method for describing not only the state name, but also *how aroused* the user is;
  - Multiple selections for emotional expression (especially for discrete emotions—e.g. a user should be able to say that she is frustrated *and* angry about a product, not simply one or the other). If using a pre-made list of emotional state descriptions, try to include an optional text field where a user may enter her own description of feelings.
- Provide feedback, especially on emotional content; communicate to the user that his emotional state has been effectively communicated, ideally by paraphrasing what the user is trying to communicate (Nugent 95). The supportive agent must convey that the *idea* has been communicated, and not just a string of words or phrases that is parroted back.
- Allow for repair if the feedback is judged wrong. This step is critical in active listening, to give the individual a chance to clarify her initial response, and make her feel as if her state is accurately conveyed to the listener. After all, especially at times of high arousal, language can easily fail to convey precise meaning, yet at such times it is often all the more important to the aroused person that she convey her situation and state accurately.
- Communicate a sense of empathy to the user. Or, to put it another way (since computers are not currently capable of empathic understanding or feeling), simply try to make the user feel *empathized with*. Since the user’s subjective feeling is the salient dynamic in a personal emotion regulation support system, the actual feelings of the supportive agent are ultimately non-sequitur to the interaction. This is true of humans as much as it is of machines. For example, a trained human psychotherapist may have well-honed Active Listening skills, enabling her to say and do “all the right things” in interaction with the client, yet

her attention may occasionally drift elsewhere during the interaction—when thinking of another client or a personal matter, for example. In this situation, the client may feel “listened to”, even though the therapist was not actually listening very much or very well. The rationale, though, is the same: the therapist feels as though the session was successful if the therapeutic effect was achieved—if the client comes away from the interaction feeling heard, listened to, understood. The implications for this kind of “absent empathy” on humans in the long term are unclear, and are discussed further in Chapter 9. However, empathy is considered to be of primary importance to the effectiveness of active listening, and, through careful design of the agent, the impression of empathy can and should be communicated from computer to human.

- Convey a sense of sympathy to the user as well. The term “sympathy” is often used interchangeably with “empathy”, but the meanings are subtly different. The two may be conceptualized in the following way: Whereas the message of empathy is “as I understand it, this is what you are going through”, sympathy’s message is more like “this is how *I feel* about what you’re going through.” (Ickes 97). Perhaps to a lesser degree than empathy in Active Listening, sympathy is nevertheless arguably an important component of this skill, since so many critical attributes of Active Listening (enabling someone else to feel as if their emotional state is understood, accepted and indeed acceptable, especially under the circumstances) are hard to achieve without it, and easy to achieve with it. Care must be taken to maintain the user’s feelings at center-stage, however; this is not always trivial to do with sympathy, since by nature sympathy is about the listener’s feelings, and not those of the person with the problem. Of course, the more attributions of feeling are applied to the computer, the more problematic the interaction becomes, especially in terms of the authenticity of the interaction. This topic will be discussed in greater depth in Chapter 9.
- Convey to the user the sense that his emotional state is appropriate for the situation, if possible. Perhaps the computer might communicate that the user is not alone in feeling the way he does in such a situation. Note that this sentiment is different from either sympathy or empathy. In humans, this impression may be conveyed by sharing a similar story, or even by an understanding glance—a raised eyebrow and a head nod while listening. It may not be feasible or believable for a computer system to feign such a human-like gesture, especially when communicating through limited bandwidth, such as text. There are ways for a computer communicate this impression, however. For example, this apparent sociological judgment is often a matter of simple consensus; it is easy, for example, for a system that deals with many complaints to do a simple statistical analysis of the complaints, and to generate a believable statement such as “You are not alone in your feelings about [Product X]. In fact, 84% of people [or simply *most people*] who report problems with [Product X] describe feeling extremely frustrated.”

#### Applying Other Strategies for Emotion Regulation Support to Interface Design



Components of active listening may prove to be the most effective methods for providing active emotion management support, but other promising strategies may prove effective as well:

- When speaking to angry customers, for example, consumer affairs representatives are counseled to employ a variety of interesting strategies (Jenks 93 illustrates ten of these). Some of these strategies are akin to active listening, such as acknowledging the customer's emotional state, being respectful, showing empathy and listening. Other strategies include agreeing with the customer, and apologizing, albeit without accepting or placing blame. The model of the service relationship underlying the customer/consumer affairs interaction may share much in common with the current tradition of human-computer interaction, particularly in agent-based interaction. Such strategies might be employed by a CASPER agent, albeit with care and sensitivity to context, and ensuring that the computer's capabilities are not being misrepresented.
- MATLAB 3.5 (Section 3.3.2, above) arguably employed humor in the service of supporting the user's emotion management. Humor, when used appropriately, can be an effective means of lifting spirits. Indeed, as suggested in Section 3.1, above, many people use humor (in movies, television shows and comics, for example) in the service of passive emotion management support. Morkes, et al (Morkes 98, as well as the first author's 1998 Stanford dissertation) found that humans responded socially to machines designed to express humor. Whether humor can be used effectively by machines in an active support role, however, remains to be seen. Since a sense of humor may vary greatly from person to person, and appropriateness of humor can range widely from situation to situation, many sources in the literature shy away from using humor (e.g. Jenks 93, Gordon 70). Still, the possibility remains that, in the right circumstances, humor might prove to be an effective support. Again, such strategies may require a "Wizard of Oz" approach to test their effectiveness in studies of this nature.

Many of these strategies should be easily implementable in software using existing methods. However, it is also possible that some of these strategies are currently not within the reach of technology (accurate, automatic recognition of the user's emotional state, for example). In this case, it may be possible to explore the efficacy of such strategies using a "Wizard of Oz" approach: A human, concealed from the subject, pretends to represent the computer, thereby simulating the desired, if perhaps futuristic, interaction capabilities. In this case, the "system" could demonstrate human-like sensitivity and empathy in interaction with subjects. The "Wizard of Oz" approach is known to be an empirically sound method for interface design and testing (Preece 94).

## Criteria and Objectives for the Study

The idea for this thesis was to build an interface agent that provided CASPER-style interaction, and then to test the agent on human users who were experiencing frustration, to see if it supported the hypotheses presented above.

In order to test the CASPER agent in a reliable, repeatable way, a protocol needed to be carefully designed, in which a random group of human subjects were put into a frustrating situation, and as a result of the situation would be feeling noticeable levels of frustration. The agent would then implement the CASPER theory in interaction with the user, in an attempt to enable the user to recover from the frustration in short order. Measures would then be taken of the effect of the agent on the subject. The protocol needed to provide a comparison between the effect of the agent on frustrated users, and the effect of other stimuli that were either typical of computer use (i.e. no support at all), or another strategy that some researchers argue should prove effective in this situation, such as emotional “venting” (Goleman 95). Further, in order to establish groundtruth for the efficacy of the CASPER agent, the effect of the agent (and the other competing stimuli) on users who were not feeling frustrated, or at least not nearly as frustrated, needed to be determined.

Experimentation that involves emotion is a very tricky business. As described in Chapter 2, an important function of emotions is social. Therefore, in an experiment that seeks to elicit an emotion (such as frustration) in a predictable, repeatable way, human interaction with the subject can represent a host of variables (the interaction of their moods, gestures, words, style of expression and overall behaviors, to name a few) that are next to impossible to control for, and can have very different, complex, emergent effects. In this experiment, then, human interaction needed to be tightly controlled. Computers, on the other hand, are superb at exact repetition of behavior, and offer an excellent means of controlling the subject’s experience, and providing the exact same stimulus repeatedly, and across many subjects. Ideally before, during and after the subject encounters the frustrating situation, there should be little or no human contact at all.

Another challenge in an experiment involving emotion is the problem of foreknowledge. At least in part because of the social nature of emotion as well as the intimate interplay between emotion and cognition (Izard 93), asking a person to feel frustrated is almost guaranteed not to elicit this emotion in any authentic way. In order for the subject to experience an authentic emotional reaction, then, one needs alternative means. One standard strategy in the social science community is to do a *blind study*—to leave the subject in the dark as to the nature of the study, and create a *cover story* to tell the subject, to make her believe that the experimenters are actually conducting the study to test for something else.

#### ***A Model for the Protocol Design***

Even with limited human contact and a good cover story, eliciting an emotional reaction in a human being in a predictable way is no small feat, with only a computer and the cover story to use. However, using Lawson’s theory (as described in the section on Frustration Theory in Chapter 2), Riseberg et al (Riseberg 98) created an experimental protocol for a pilot study to frustrate human subjects using only a computer and a mouse, and observe their physiological responses with non-invasive biosensors

The scenario was thus: subjects were recruited to come and play what they were told was a game in which their visual acuity was being tested. The game appeared to be a simple visual puzzle game, that subjects were told to complete as quickly and as accurately as they could. Subjects were paid US\$10 for their participation, but were told that the best, fastest scorer would win US\$100. With this incentive, subjects began playing the game—which, unbeknownst to the subject, had seemingly random delays built into it, which made it appear as though the mouse were malfunctioning. If the subject commented about it to the administrator, who was quietly reading on the other side of the same room, the administrator would say simply, “Oh, it sticks sometimes. Just keep going.” So, following Lawson’s recipe, subjects were given a goal, incentive to achieve the goal, and then were thwarted from achieving the goal. Although no baseline (i.e. non-frustration) condition was used in the study to compare the states, self-report measures in this experiment indicated that a majority of subjects felt frustrated by the delays.

### *The Model Adapted*

The above scenario served as the model for the elicitation of frustration in the present study—with some important differences.

Since the present investigation sought to raise frustration levels (only to try to lower them right afterward), as well as to control for frustration level on the whole, it was not deemed practical to have the administrator in the room with the subject during the experiment.

Further, since the subject would not be physically tethered to the apparatus (unlike in the prior study, where biosensors strapped to the body were connected by wires to free-standing equipment), there was a chance that the subject would get up and go find the administrator during the experiment, if the “problem” encountered by the subject was deemed to be great enough. This action would invariably skew the results. An apparent hardware problem cover story such as was designed for Riseberg’s study posed too great a risk, since the subject might assume that such a “local” problem could be easily remedied.

It was therefore decided that a less “local”-appearing problem needed to be implemented. As mentioned earlier, the World Wide Web (WWW, or “the Web”) has become a popular, credible source of frustration in recent years, due to often unpredictable network traffic, and often slow loading times. Indeed, a long-standing joke about the WWW is that it stands for “World Wide Wait”.

A Web-based scenario (or at least the appearance of one) therefore became the method of choice for a frustration-elicitation system with a believably remote reason for failure, but a cover story needed to be carefully crafted to support the protocol, and its apparent circumstances. It was felt that a game protocol patterned loosely after the Riseberg study was a wise idea, given our initial success with it. There would be freezing delays in this game too, occurring seemingly at random, and for seemingly random intervals. These delays, however, would be attributed to the Web.

In order to make a believable cover story, a new Web-based game that researchers at the MIT Media Lab would want to have subjects come and play would need to fit certain criteria. First, there would need to be a good reason why researchers here would need people to come in and play it. The “visual acuity test”

story in the Riseberg study would not fit here, since there would be no purpose for the Web element. What was opted for was a cover story that revolved around the idea of play-testing a prototype of a new kind of game, using a new kind of Internet protocol. The game would need to look and act the part—look slick in parts, raw and “prototype-like” in others, and feature ruses—elements of the game’s design that would look like they might be interesting enough to want to be the “real reasons” the Lab was conducting such a test.

#### Evaluation Criteria, Predictions, & Protocol Overview

The measurement criteria for determining the effectiveness of the CASPER agent seemed simple enough: Once they had completed the experience, simply ask subjects how they felt.

Unfortunately, life is not that simple. For starters, there’s the social desirability problem. Humans tend to shy away from describing themselves (and, therefore, their internal states) to others in ways that may convey weakness, vulnerability, or other attributes that they consider might make them lose face, or make them look less socially desirable.

Humans also tend to be notoriously bad (or at best, uneven) at accurately assessing how they feel (Goleman 95). Further, they are known to be bad at determining the source of, or reason for, their feelings. For these and other reasons (see Zajonc 98 for an excellent review of these issues), self-report, particularly of emotional variables, has long been known to be unreliable in social science research.

Behavior, however, is a different story. One important aspect of human emotion is that important elements of it have been shown to be non-cognitive (LeDoux 94, Damasio 94)—humans can do things that are emotion-driven, without stopping to think about them (such as jumping out of the way of an object that is rushing toward you). Similarly, emotions can affect action in ways that the person often cannot explain; they simply act, perhaps finding reasons for their behavior after the fact. So what people say they do, or the reasons people *think* they are doing them, can vary greatly. Behavior, by contrast, is well known in the psychology community to be a much more reliable indicator of state, for the very same reasons.

So it was decided that the study would rely mainly on behavioral measures, with self-report data gathered as well. The protocol to collect this data would be as follows: subjects would be given a goal and an incentive via email messages, a script the administrator would read to them, and on-screen directions. Play a game on the computer. Play it once, as hard as possible, for a fixed amount of time (5 minutes), and try to win one of two US\$100 prizes for the top score. Answer some questions on screen (different subjects would be given different interactive questionnaires; the CASPER agent would interact with the user masquerading as one of these), then play the game again. However, the second game would not be timed, and would not count towards any prizes. Players would be asked to play the second game for at least three minutes, but could quit any time after that. They could also keep playing, if they so chose, for up to 20 minutes. They would then fill out a paper questionnaire asking about how they felt about each game, why they played game 2 for as long as they did, and—most importantly—how they felt, in 5 different emotion

categories, both then and at various points in the experiment (see the Exit Questionnaire, Appendix B.6).

The behavioral measures that were the focus of the study were the time subjects spent played the second game, and how many points they scored in that game. The metric was simple: The more frustrated you are, the less likely you will want to sit and play a game that, on the whole, is a bit tedious and not very exciting to play. (Note: the tediousness of the game was confirmed in subjects' self-report descriptions of the game, discussed in *Description of the Game* in Chapter 5; some of these responses may also be found in Appendix C.17, "Responses to On-Line, Open-Ended Question".)

Also, according to frustration theory outlined above, subjects feeling more frustrated should have an impaired ability to pay attention, think creatively, or remember the layout of the maze as well as subjects who did not experience frustration. Therefore, subjects who are feeling high levels of frustration should play Game 2 for less time, and score fewer points, than players who are feeling better, more relaxed, and less agitated. Hence, subjects who were feeling less frustrated after interacting with the agent (or one of the other questionnaires) would play Game 2 for longer, score more points, and rate their experience as better, on average, than those who were feeling more frustrated. Accordingly, subjects who were feeling less frustrated should, ideally, rate their frustration levels as lower, and their satisfaction and fun levels as higher, than those who were feeling more frustrated.

The literature on "venting" suggests that it is unreliable at aiding humans in recovering from strong frustration. It assists some in calming down and feeling better, as it allows them to divest themselves of the problem, to "get it off their chests". For others, or perhaps for the same people in other situations, unsupervised "venting" (also known as *catharsis*) can simply encourage the strong, negative feelings to rise to the surface again, even long after the fact—as if the person were reminded of the situation all over again, and began running a simulation of the event in their mind, complete with their initial (or intensified) affective responses.

For venting, then, huge variance is expected, as some walk away feeling "cleansed" and others are simply reminded of how frustrated they felt, and either maintain or build their frustration levels accordingly. The effect of venting so soon after the eliciting event is somewhat unclear, however, as this author has not heard of a study directly testing venting in this manner. Still, the expectation is for behavioral and self-report measures that reflect a number of extremes; the mean for such measures should likewise remain somewhere in the middle of the spectrum.

A further hypothesis is that there may be differences in responses to the CASPER agent by gender and emotional arousability, as well as frequency of game play. A number of researchers contend that women pay more attention to emotions, interpersonal communication and relationships than men, and consequently may have higher sensitivity to the CASPER agent; the agent may therefore have higher success rates among women than men (Tannen 90).

Similarly, *trait arousability* (Mehrabian 95) is an established measure (with its own proven collection metric, by means of a 34-question inventory) that roughly corresponds to emotional arousability—how emotional, positive or negative, that an individual tends to become in response to day-to-day events. Subjects who score high in trait arousability should be expected to be more emotionally affected by the

game—feeling less tolerance for playing Game 2 after experiencing frustrating delays in Game 1, without support for emotion regulation. *With* such support (via the CASPER condition), however, subjects with high trait arousability should be expected to feel *much* better, and therefore play Game 2 for much longer.

By contrast, seasoned game players, as well as those who have experience playing games on the Web, may be more accepting of apparent Web delays, and/or more analytical of the game's (pre-designed) limitations, and consequently less emotionally invested. Therefore, seasoned game players should be less affected by the CASPER agent than others, and may be expected to play for shorter times in Game 2.



## CHAPTER 5

## Method

The study consisted of a 2 (frustration: NO DELAY, DELAY) x 3 (feedback-type: CONTROL, VENT, CASPER), between-subjects, full-factorial experiment (see Table 5.1, below). The experimental protocol and all collateral documents were approved by the MIT Committee On the Use of Humans as Experimental Subjects (COUHES Addendum to “Affective Computing: Pilot Studies”, Protocol #2321) prior to the study. The investigation used humans as blind experimental subjects—subjects who were deliberately misled as to the intent of the study, and who were not made aware of the study’s purpose until after the experiment was completed. A multimedia computer system (hardware and software) was iteratively designed, built, and tested to perform specifically for this situation.

Table 5.1: The six conditions in the 2 x 3 experiment

	NO DELAYS IN GAME 1	DELAYS IN GAME 1
CONTROL QUESTIONNAIRE	CONDITION 1 $\underline{N} = 12$ (6F, 6M)	CONDITION 2 $\underline{N} = 12$ (6F, 6M)
“VENTING” QUESTIONNAIRE	CONDITION 3 $\underline{N} = 11$ (5F, 6M)	CONDITION 4 $\underline{N} = 12$ (6F, 6M)



CASPER AGENT/ QUESTIONNAIRE	CONDITION 5 $\underline{N} = 11$ (5F, 6M)	CONDITION 6 $\underline{N} = 12$ (6F, 6M)
--------------------------------	--	--

## Participants

Human subjects were recruited ( $\underline{N} = 114$ ; see note on subject attrition, below.  $\underline{N} = 70$  for the actual study). In order to account for gender,  $\underline{N} = 11$  or  $12$  for each of the six conditions, which included  $\underline{n} = 6$  males in each condition, and  $\underline{n} \geq 5$  females in each conditions ( $\underline{n} = 6$  females in four of the six conditions; see Table 5.1 on the preceding page for  $\underline{N}$  for each condition). Participants' ages ranged from 14 to 44, where the median age was 20-24. The two youngest subjects were noted to be 14-year-old males (for details, see Appendix C.1), who were children of MIT faculty or staff. Most subjects were MIT undergraduate students, but there were a number of graduate students and professionals as well. Participants' occupation and race were noted by the administrators only to the extent that all these categories seemed to be random, and varied. However, these data were not officially recorded in this study. Subjects' prior experience playing computer games and web-based computer games were averaged together to derive a mean frequency of relevant computer game experience. These scores followed a random distribution described by a curve (see Figure 5.3, below), and ranged from 1 or 2 experiences playing games in one's life to playing several times per week.

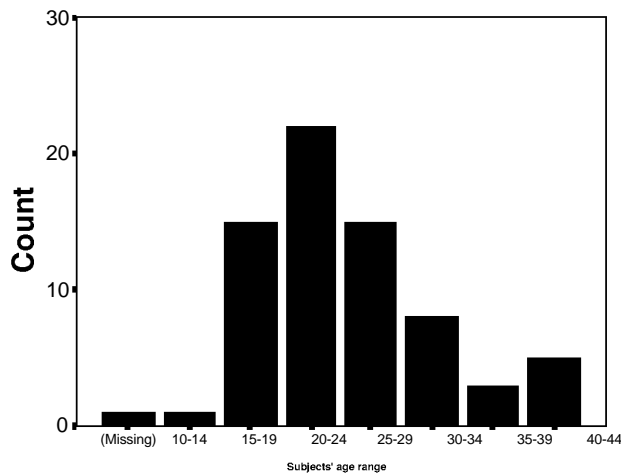


Figure 5.2: Distribution of age ranges of subjects

### ***A Note on Subject Attrition***

A subset of the recruited subjects who participated ( $\underline{n} = 34$ ) were used as preliminary user-testers in early versions of the experimental system, which required first a

revamping and later refinement of the game's playability, the various conditions of the protocol, as well as the system's overall usability. Of the remaining number ( $N = 80$ ), 2 experienced a rare bug in the final version of the software that made Game 2 fail, and 2 subjects' electronic log file data were irretrievably lost—once due to a crashing bug in the system, and once due to administrator error. Another subject was disqualified from participating (he had already participated, but neglected to tell the experimenters), and one subject asked to have his data withdrawn (his right as specified by COUHES and as stated in advance email, the consent form, and the debriefing form) for personal reasons.

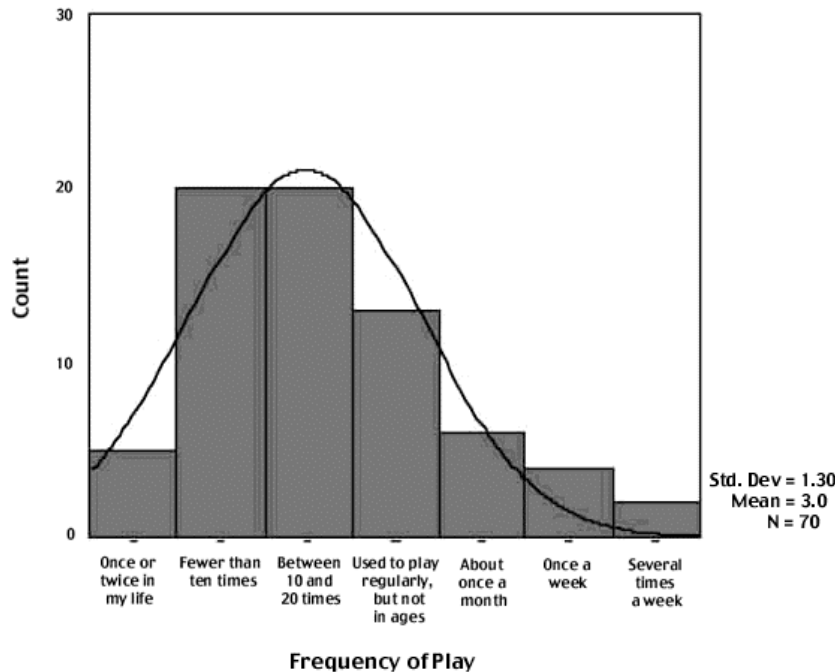


Figure 5.3. Distribution of mean relevant computer game play across subjects. Figures represent the mean resulting from combining frequency of computer game play and frequency of play of graphical, web-based computer games.

Of the remaining subjects ( $N = 74$ ), the data from four were omitted from the final study in order to homogenize sample counts across conditions. Subjects to be omitted from the study were either chosen chronologically (the last subject to participate was omitted), or for some extenuating reason such as subjects' own report of strong, external motivations for playing Game 2 for a certain amount of time, or suspect command of the English language, which was needed in order to understand and respond to the large number of written and spoken materials. Specifically:

- One subject was removed from the DELAY/CONTROL/FEMALE condition in which there was a surplus ( $n = 7$ ) because her command of English was

extremely limited, and suspect; she arrived to participate with a friend who offered to act as an interpreter, and explained that the subject was visiting from China (the offer to interpret was declined, since such different treatment of each subject was not acceptable within the protocol). And, indeed, the subject's responses to some questions demonstrated an inability to understand those questions (no other suspect subject was found in this group).

- One subject was removed from the DELAY/VENT/MALE condition in which there was a surplus ( $\underline{n} = 7$ ) because he was the last subject to be run in this condition (no other suspect subject was found in this group).
- One subject was removed from the DELAY/CASPER/FEMALE condition in which there was a surplus ( $\underline{n} = 7$ ) because she reported “hunger” for lunch as the reason she played as long as she did. She was the only female in this condition to report an external motivation for playing a certain amount of time). Note: 3 independent coders rated question #3 on the Exit Questionnaire, “Why did you play the second game for as long as you did?” for type of response, among which was an “externally-motivated reason” option (see Appendix C.16B for the text of the directions to the coders).
- One subject was removed from the DELAY/CASPER/MALE condition in which there was a surplus ( $\underline{n} = 7$ ) because he was the last subject to be run in this condition (no other suspect subject was found in this group).

## Apparatus

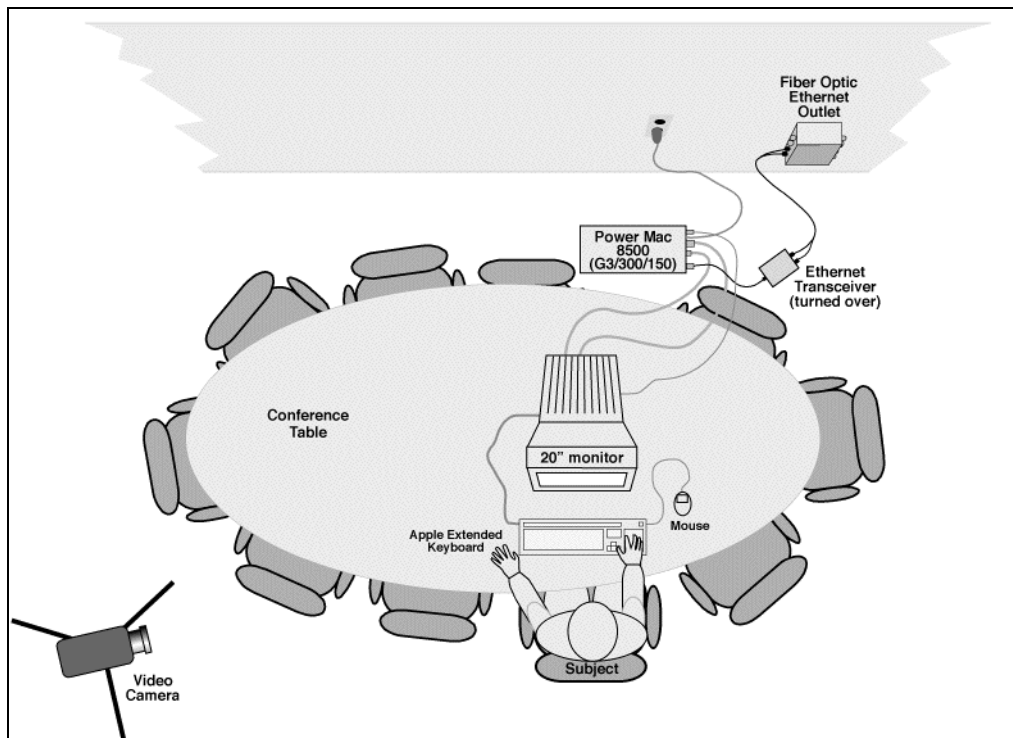


Figure 5.4: Schematic of the physical set-up for the experiment.

### ***The Physical Set-up***

The experiment took place in a well-lit conference room, equipped with an oval conference table, 10 chairs, some audio/video equipment (a large monitor and VCRs on a cart) pushed off to the side, and minimal decor (see Figure 5.4, above).

A single video camera (A small, Hi-8 Sony HandyCam) was mounted on a tripod to the subject's left, and was used to record subjects in interaction with the system. From this side angle, the camera was able to record subjects' profiles, as well as some measure of what was on the screen.

The primary hardware for the experimental system consisted of an Apple Power Macintosh 8500 tower computer, equipped with an upgraded processor board (a MaxPowr 300 MHz G3 processor with a 150 MHz backside cache), 128 MB of DRAM and 2 MB of VRAM. The computer's sole output device was an Apple Multiple Scan 20 Display (running thousands of colors at a resolution of 1152 x 870 pixels, at 75 Mhz). Input devices consisted of a standard Apple Extended (105-key) keyboard with numerical keypad on its right-hand-side (the sole means of interaction in the game), and a standard Apple Design Bus mouse. The monitor, keyboard and mouse were the only objects on the table.

Of all subjects, only two included in the study suspected that the game might not be communicating with a remote server, due to apparent tell-tale noise from hard-drive activity that only skilled ears could detect. Early pre-tests revealed this weakness in the design, however, and steps were taken to minimize its effects. The computer, originally situated next to the monitor on the table, was moved down onto the floor behind the table, with the table partially occluding its view from the subject. This was done before any of the 70 subjects included in the final study were run, to minimize the effects of the sound of the hard disk activity.

In both cases of suspicion due to hard drive noise, the degree of suspicion was judged to be mild, since both believed that the noises they were hearing might be part of the new Internet communication protocol (part of the cover story).

In support of the cover story, an Ethernet connection consisting of fiber-optic cables connecting to a CentreCOM MX80F 10BaseT transceiver box was connected from the computer to a large, visible, multi-port fiber-optic cable outlet mounted on the wall, in view of the subject. Since this outlet was not itself connected to the fiber-optic cable network, an Ethernet connection from the network to the computer could not actually be made. Consequently, the LED indicators on the transceiver box were conspicuously dark—a dead giveaway of the cover story for the technically-savvy subjects around MIT.

The solution was simply to turn the transceiver box over, so the LEDs faced the floor. Although the administrator was not present in the room with the subject during the experiment, since the subject's experience in the room was tightly orchestrated, and since the subject's actions in the room were being videotaped, it was believed that the subject would not have either time or inclination to examine the transceiver box up close. And, indeed, the cover story seemed to prove unanimously effective.

### ***Software***

The experimental software system consisted of the on-screen directions, the game system (including the simulated Web browser and accompanying delay mechanisms),

the three on-screen questionnaires (only one of which was seen by any one subject), the CASPER interaction agent and the means for recording all relevant events, times, scores, and questionnaire answers into automatically-generated logfiles. All elements of the system were constructed by the author in Macromedia Director 6.0 for the Macintosh, a platform for development of interactive multimedia. The system made heavy use of Lingo, Director's programming language. This system represented a major redesign from an early version of the system, which was built largely by Matthew Norwood, an undergraduate UROP (in computer science and cognitive science) working under the author's supervision. A number of elements of this earlier design were adopted for use in the final system.

### ***Technical Design of the Game***

The game prototype, which the author designed and built from scratch, is a simple scenario of a character seeking treasure that is lying about in an adventure-game-style maze of paths. The game features some sophisticated-appearing 3-D graphics (actually executed in 2.5-D) and animation of the sort not-yet seen in web-based games, to support the cover story.

The main character, which was modeled, animated and rendered in 3D, was designed to be viewed from a fixed-aspect-ratio and rendered orthographically (without any perspective), to support a "2.5-D" viewing paradigm. This perspective enabled her to seem as if she were moving in 3D across a flat plane, but without ever getting any nearer to the camera.

The main character, as well as all jewels and other artifacts in the game world, were modeled, rendered and animated in 3D using Macromedia Extreme 3D 2.0 for the Macintosh (see figure 5.5, below). The character was designed to be female, and the game was designed to be simple and non-violent. These features were designed to support the cover story and provide a false target for satisfying subjects' curiosity about the study and the reasons for conducting it in a research environment (the Media Lab). The lack of violence (and the game's overall theme of searching for, and collecting treasure) was also designed to make it appealing to women as well as men, to counter-balance for the effect that a typical male-oriented "shoot 'em up" game would have on female subjects.

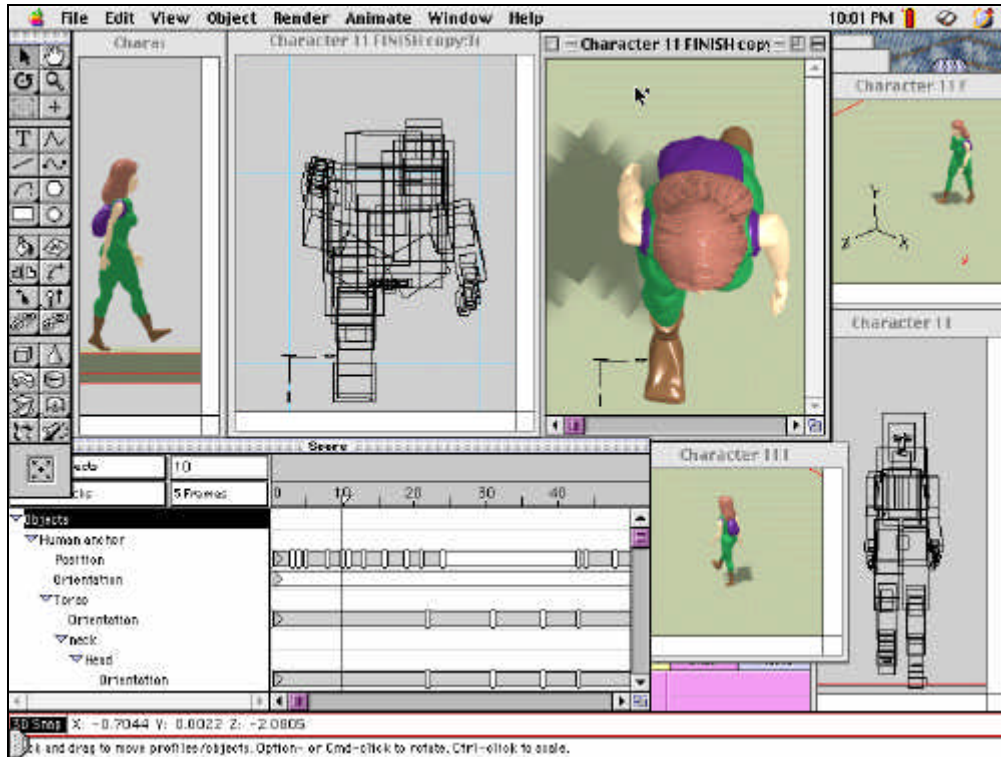


Figure 5.5: The main character, as well as all jewels and other artifacts in the game world, were modeled, rendered and animated in 3D using Macromedia Extreme 3D 2.0 for the Macintosh.

The character was constructed and animated in the 3D package to walk in one direction, turn in two directions, and bend down to pick things up in one direction. Once the modeling and animation was complete (and exact), the character was rendered from 8 different, carefully-placed camera angles, using 8 different lighting schemes that, together, made the character appear to walk, turn and bend in all directions, all from the same perspective.

The rendered images were brought into Adobe Photoshop 4.0 for the Power Macintosh, and batch manipulated to silhouette and compress the images. The images were then imported into Director. In Director, a scheme for animating and walking was developed, in which the character appeared to move in any one of 8 different directions (Every 45° of arc—up/down, left/right, and diagonally 4 ways). In the game's design, the character actually stayed stationary in the center of the screen, but would appear to move about in the world by means of having the elements in the world flow by her—as if she were on a treadmill. This scenario afforded the user complete control of the character at all times, provided the system was not in a feigned “network delay” mode. Such a scenario was an improvement over the original movement paradigm for the game, in which the character would need to react to a world that slowly scrolled by the character down the screen. Opportunities arose in that design for one's character to become trapped behind objects, which proved to be too much risk for unplanned disempowerment of the user—too similar to the delay conditions in the game.

## Description of the Game

The game needed to support the cover story down to its barest details. The briefing script read to subjects before they began play (see Appendix B.1) described the game as a “first-playable”, an alpha-version of a character-based game, in which the character was mostly fleshed out, but the world mostly wasn’t. Also due to the “early stage of the game’s development”, it lacked music and sound effects, as well as rich graphics and a subset of the things the character would “eventually be able to do in the finished game.” These features were not implemented on purpose—to ensure that subjects’ experiences with the game would not be so immersive or engaging that subjects would feel compelled to play for as long as possible during Game 2. Indeed, when asked how they liked the games, subjects reported on the Exit Questionnaire that the game seemed “somewhat boring” and “dull”.

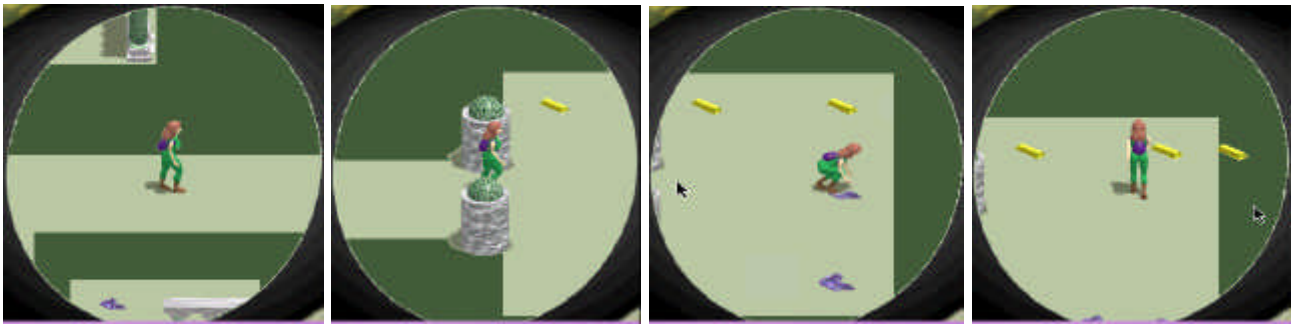


Figure 5.6: Examples of the character’s movement. Note that the character stays in the center of the screen at all times; she *looks* like she’s moving since the world is drawn as moving around her.

The game itself is relatively simple. Once the user presses the Start button (as directed in the instructions), the time counter (which reads “00:00” at the game’s outset) starts to iterate every second. The subject then uses 8 of the numeric keypad keys to move the character around in the world in any of 8 different directions (the keypad layout is provided in the directions, and ubiquitously on the game’s background graphic as well).

The character’s movement is constrained to occur only on light-colored paths, which connect to one another at right-angles, and form a matrix on top of the dark-green “grass”, where the character’s feet are unable to step. The paths form a large, complex maze, with only a small portion of the maze visible at any one time. Five different kinds of treasure may be found by the character at various points around the maze, but often with clusters of treasure, or “treasure troves” found at many of the maze’s dead ends. When the character gets near a piece of treasure, it bends down to “pick it up”, whereupon the piece of treasure disappears from the maze, and the score is updated to reflect the piece’s value (as stated on Page 2 of the directions).

There were two elements of the game that seemed a bit buggy to subjects, that actually corresponded to real bugs in the programming of the game. The first was an occasional flicker: The maze in the software was realized as a series of 23 interconnected “rooms”. Each room had paths that led into other rooms, thus forming the maze. The bug involved a very brief (<100ms) flicker inside the oval

“window on the world” at the center of the game window, each time the character passed from one room to another. It seemed to have the effect of slightly disorienting subjects, so that it was very slightly more difficult to keep track of one’s path in the maze. However, the bug actually served to support the cover story, since the game was supposed to be an “early prototype.” And, since all subjects in all conditions experienced this bug, it was not considered to be a significant factor in the experiment.

So too was a bug that involved getting stuck, periodically, while navigating the characters around hard corners. Occasionally, the character would get stuck going around a corner, with no apparent reason for the failure. The character would stop moving in that direction, and fail to continue moving in that direction. Other keys would turn the character, however, and moving the character in an intuitive direction away from the corner would solve the problem. This bug, again, was experienced by all subjects. It was a bit more troubling of a bug than the flicker, though, since it could involve making subjects feel as though the character were stuck, similar to the effect of the apparent “Web delays”. It proved enough of an issue to include a “hint” about it in the directions for Game 1. With the problem mentioned conspicuously in the directions, and since all subjects experienced the same problem—and since the character did respond to being turning, and there were no associated Web-browser delay effects, the problem was not estimated to make a significant impact on the study.

The game was designed to take at least 28-30 minutes to collect all the treasure in the entire maze. The maze itself follows a semi-linear path from section to section of the maze, with multiple, two-way connections (akin to a grid) in the middle of a number of sections. There are also several one-way paths, where the character cannot go backward after a certain point. There are no traps that cannot be escaped from.

Navigation in the maze does eventually get a bit harder, as some pathways become narrower. This type of gradual escalation of challenge is typical of computer games, to maintain interest. This measure was intended to be a little more exciting, but not by much.

The maze is not incredibly hard to navigate through, but since it cannot be visualized whole on screen, and since there is a great cost (in time) to going back down the same path that one took before, memory is an asset to performing well in the game. (Those who are more frustrated should be less able to remember where they are in the maze; consequently, they should not score as many points in the same amount of time as players who are not frustrated.)

The character takes a little time to change direction, as she actually turns around in a natural way, with natural-looking arms swings. She also takes a little time to bend down and pick up a piece of treasure—so there is a cost, in time, to both movements. Picking up the ubiquitous 1-point treasures, then, becomes something to be avoided, in favor of discovering large-point treasure troves further away. Also, the character moves slightly faster on the diagonal than she does walking at 90° angles. These phenomena are meant to be found as the means for using one’s intellect in the game. Also, in theory, those who are less frustrated should, on average, think more creatively and therefore discover more of these phenomena.



Adapting to them, they should score more points than players who are feeling more frustration.

### Design of Interactive Questionnaires & CASPER Agent

These “interactive questionnaires” all had a number of things in common, as well as key differences.

The similarities were thus: All response conditions (CONTROL/VENT/CASPER) featured an “interactive questionnaire” segment, that functioned exactly the same across conditions, and all began in the exact same way. The functionality was that of a dialog box with multiple frames which could be advanced (but not rewind), in which text was displayed, sometimes alone and sometimes in tandem with a series of labeled radio buttons, or a blank text field where subjects could enter free-form responses to questions.

All had either one (Okay”) or two (“Yes”/“No”) buttons at the bottom of the screen, used as a means of advancing to the next screen, and identifying the input on the current screen as acceptable (or, in the latter case, as acceptable/unacceptable). In any given screen with radio buttons or text field, subjects could change their minds about the input (i.e. edit their text responses using the keyboard, or select a different radio button with the mouse), until the button at the bottom of the screen was pressed.

All conditions began with the same first five frames, with identical content. The content of these screens was as follows:

“This computer is now going to ask you some questions about the game. Okay?”  
[Okay]

“Please note that your responses to the questions this computer asks are very important to the researchers, but are entirely voluntary. Please feel free to skip any question you don’t want to answer.” [Okay]

“First off, can you describe your age?” [*11 radio buttons with age ranges, from 10 years to 70+ years old*] [Okay]

“Second, what is your sex?” [*2 radio buttons: Male, Female*] [Okay]

“Okay. Now how often have you played computer games or video games before?” [*11 radio buttons with labels of different frequencies, ranging from*

*“Never” to “Many hours each day”*] [Okay]

Subsequent frames differed at this point, depending on the condition (see Table 5.7 on the following page for a list of all questions asked, with user response possibilities, grouped by condition and by common questions). All conditions were designed to take the same amount of time to complete, to control for the same

length of time between Game 1 and Game 2, during which time the subjects spends on approximately the same kind of activity. This way, any effect found for the CASPER agent could not be attributed to differences in the times that subjects interacted with each condition's response method. Time to completion was based on total number of questions/text read, similarity of question to previous questions, and similarity of response labels to those of previous questions. For the CASPER condition, estimated time to completion was based on the number of screens (with text on screen) in a typical interaction, since this number varied depending on the amount of repair dialogs (0 to 3, with an assessment screen after each repair dialog) the subject encountered. See Figure 5.8 on the page after next for a diagram of the interaction flow for the CASPER agent.

#### ***Controlling for the “Conversational Quality”***

The “chatty”, conversational feedback cues (such as “Okay”, “Hmm” and “And...”) employed here were used to adjust for the fact that the CASPER condition used dynamic (although simple) feedback—it conversed with the subject based on the subjects responses. The CONTROL questionnaire used similar cues (see Table 5.7). In this way, subjects in all conditions experienced a conversational effect, as if they were conversing informally with “another”.

#### ***CONTROL Condition***

Subject in this condition answered questions that did not enable them to report problems they encountered, or to describe how they felt about the experience. All user responses were via radio buttons; i.e., there was no blank text field available in this condition, so no means for users to use their own words to describe their experience or their reactions to it. There were four more “frequency”-style questions, three about various permutations of game play experience, and the fourth asking about frequency of play with a female main character—meant as a ruse question to support the cover story only.

The next question asked about the main character's appearance (too big? too small? visible enough? Ruse questions and responses, non-sequitur to the events estimated to be relevant to the subject). The next three questions followed suit, asking about the frequency and visibility of treasure items. The final two questions asked for 1st and 2nd top picks for what the subject would like to see implemented next in the game (puzzles, a male character instead of a female one, monsters to kill, etc.). Subjects are then thanked for their answers, and instructions for Game 2 are displayed.

#### ***VENTING Condition***

The first nine screens of this condition were identical to the first nine in the CONTROL condition (all questions up to and including the question about frequency of game play with a female main character). The rest of the questions differed from the CONTROL condition, however, in that they asked questions specifically designed to enable the subject to report the relevant problem (i.e. the network delays), as well as be able to describe exactly how they felt about it (see Table 5.7).

These questions were as follows:

“Okay. Now how well do you think you did in this game?” [0 (*The worst score*) to 10 (*The best score*)] [Okay]

“How was the speed and smoothness of the network while you played? Were there any delays?” [0 (*Smooth; no delays*) to 10 (*Awful; big delays*)] [Okay]

“Hmm. If there were any delays, do you think they affected your game?” [0 (*No, not at all*) to 10 (*Yes, greatly*)] [Okay]

“How frustrated do you think you got playing the game, all things considered?” [0 (*Absolutely not frustrated at all*) to 10 (*The most frustrated I have ever felt in my life while playing a game*)] [Okay]

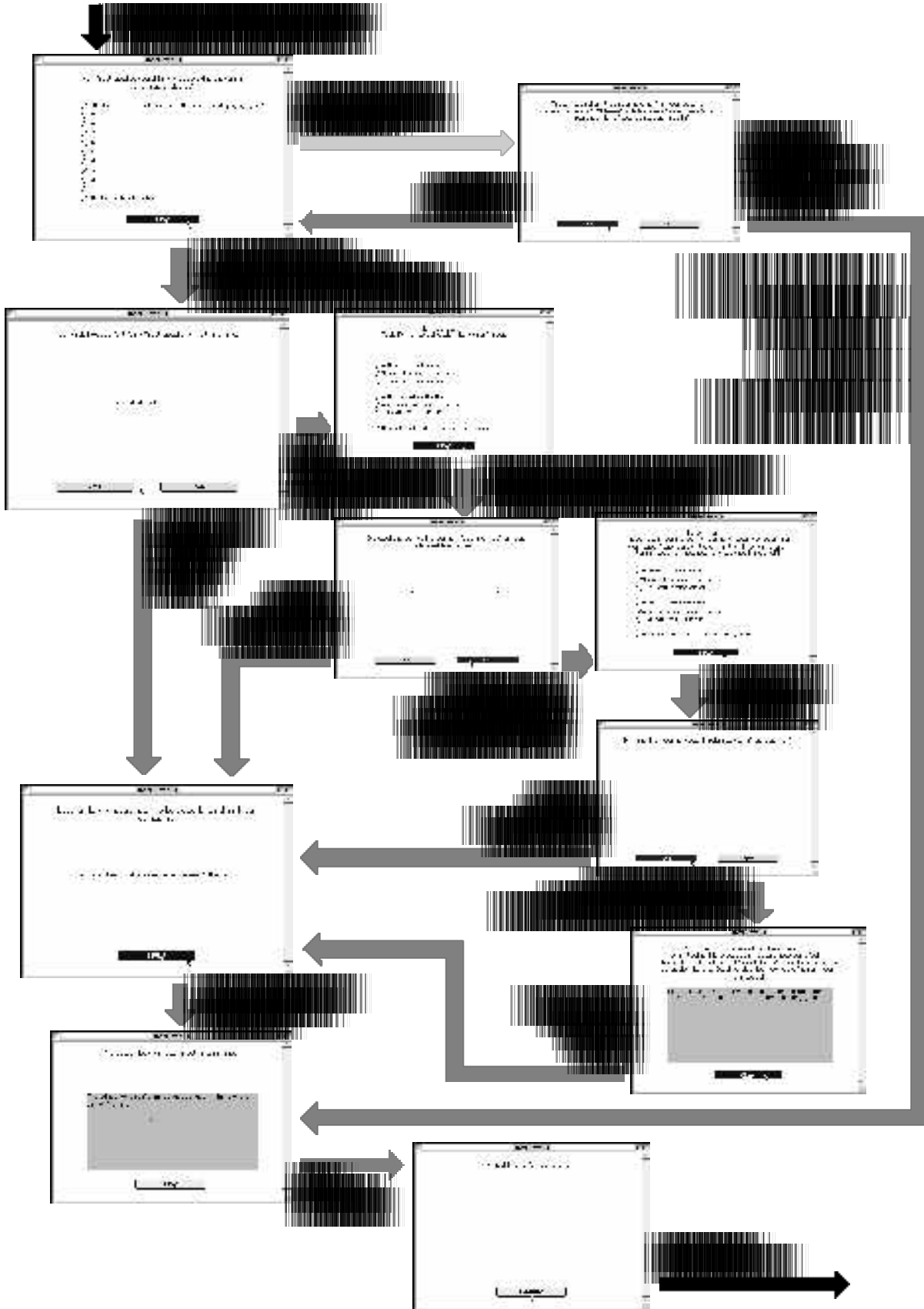
“And how much satisfaction did you experience as a result of playing?” [0 (*Absolutely not satisfied at all*) to 10 (*The most satisfied I have ever felt while playing a game*)] [Okay]

“Okay. Did playing the game make you feel anxious or tense? If so, how much?” [0 (*Absolutely not anxious or tense at all*) to 10 (*The most anxious or tense I have ever felt while playing a game*)] [Okay]

“Overall, how much fun did you have playing the game?” [0 (*Absolutely no fun at all*) to 10 (*The most fun I have ever had while playing a game*)] [Okay]

<input type="radio"/> 10-15 <input type="radio"/> 15-19 <input type="radio"/> 20-24 <input type="radio"/> 24-29 <input type="radio"/> 30-34 <input type="radio"/> 35-39 <input type="radio"/> 40-44 <input type="radio"/> 45-49 <input type="radio"/> 50-59 <input type="radio"/> 60-69 <input type="radio"/> 70+	<input type="radio"/> 10-15 <input type="radio"/> 15-19 <input type="radio"/> 20-24 <input type="radio"/> 24-29 <input type="radio"/> 30-34 <input type="radio"/> 35-39 <input type="radio"/> 40-44 <input type="radio"/> 45-49 <input type="radio"/> 50-59 <input type="radio"/> 60-69 <input type="radio"/> 70+	<input type="radio"/> 10-15 <input type="radio"/> 15-19 <input type="radio"/> 20-24 <input type="radio"/> 24-29 <input type="radio"/> 30-34 <input type="radio"/> 35-39 <input type="radio"/> 40-44 <input type="radio"/> 45-49 <input type="radio"/> 50-59 <input type="radio"/> 60-69 <input type="radio"/> 70+
<input type="radio"/> Female <input type="radio"/> Male	<input type="radio"/> Female <input type="radio"/> Male	<input type="radio"/> Female <input type="radio"/> Male
<input type="radio"/> 0 (Never) <input type="radio"/> 1 (Once or twice in my life) <input type="radio"/> 2 (Fewer than ten times) <input type="radio"/> 3 (Between 10 and 20 times) <input type="radio"/> 4 (Used to play regularly, but haven't in ages) <input type="radio"/> 5 (About once a month) <input type="radio"/> 6 (Once a week) <input type="radio"/> 7 (Several times a week) <input type="radio"/> 8 (Once a day) <input type="radio"/> 9 (At least once a day) <input type="radio"/> 10 (Many hours each day)	<input type="radio"/> 0 (Never) <input type="radio"/> 1 (Once or twice in my life) <input type="radio"/> 2 (Fewer than ten times) <input type="radio"/> 3 (Between 10 and 20 times) <input type="radio"/> 4 (Used to play regularly, but haven't in ages) <input type="radio"/> 5 (About once a month) <input type="radio"/> 6 (Once a week) <input type="radio"/> 7 (Several times a week) <input type="radio"/> 8 (Once a day) <input type="radio"/> 9 (At least once a day) <input type="radio"/> 10 (Many hours each day)	<input type="radio"/> 0 (Never) <input type="radio"/> 1 (Once or twice in my life) <input type="radio"/> 2 (Fewer than ten times) <input type="radio"/> 3 (Between 10 and 20 times) <input type="radio"/> 4 (Used to play regularly, but haven't in ages) <input type="radio"/> 5 (About once a month) <input type="radio"/> 6 (Once a week) <input type="radio"/> 7 (Several times a week) <input type="radio"/> 8 (Once a day) <input type="radio"/> 9 (At least once a day) <input type="radio"/> 10 (Many hours each day)
<input type="radio"/> 1 (Character too small) <input type="radio"/> 2 (Character too large) <input type="radio"/> 3 (Character not visible enough) <input type="radio"/> 4 (Pattern of character's outfit needs adjustment)	<input type="radio"/> 10 (The best score!) <input type="radio"/> 9 <input type="radio"/> 8 <input type="radio"/> 7 <input type="radio"/> 6 <input type="radio"/> 5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/> 0 (The most frustrated I have ever been in my life playing a game)	<input type="radio"/> 10 (The best score!) <input type="radio"/> 9 <input type="radio"/> 8 <input type="radio"/> 7 <input type="radio"/> 6 <input type="radio"/> 5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/> 0 (The most frustrated I have ever been in my life playing a game)
<input type="radio"/> 1 (Purple) <input type="radio"/> 2 (Black onyx) <input type="radio"/> 3 (Gold bar) <input type="radio"/> 4 (Red ruby)	<input type="radio"/> 10 (The most frustrated I have ever been in my life playing a game) <input type="radio"/> 9 <input type="radio"/> 8 <input type="radio"/> 7 <input type="radio"/> 6 <input type="radio"/> 5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/> 0 (Absolutely not frustrated at all)	<input type="radio"/> 10 (The most frustrated I have ever been in my life playing a game) <input type="radio"/> 9 <input type="radio"/> 8 <input type="radio"/> 7 <input type="radio"/> 6 <input type="radio"/> 5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/> 0 (Absolutely not frustrated at all)
<input type="radio"/> 10 (More non-violent game play) <input type="radio"/> 9 (A more fanciful or interesting main character) <input type="radio"/> 8 (Social interaction with other characters/players) <input type="radio"/> 7 (A more fanciful or interesting main character) <input type="radio"/> 6 (Puzzles to solve) <input type="radio"/> 5 (The ability to die and come back to life) <input type="radio"/> 4 (The ability to die and come back to life) <input type="radio"/> 3 (A male character instead of a female one) <input type="radio"/> 2 (Monsters to kill) <input type="radio"/> 1 (Monsters pursuing your character)	<input type="radio"/> 10 (The most frustrated I have ever been in my life playing a game) <input type="radio"/> 9 <input type="radio"/> 8 <input type="radio"/> 7 <input type="radio"/> 6 <input type="radio"/> 5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/> 0 (Absolutely not frustrated at all)	<input type="radio"/> 10 (The most frustrated I have ever been in my life playing a game) <input type="radio"/> 9 <input type="radio"/> 8 <input type="radio"/> 7 <input type="radio"/> 6 <input type="radio"/> 5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/> 0 (Absolutely not frustrated at all)

(END OF CONTROL QUESTIONNAIRE)



“Did your experience playing the game make you feel angry? If so, how much?”  
 [0 (*Absolutely not angry at all*) to 10 (*The most angry I have ever felt while playing a game*)] [Okay]

“Please describe your reaction to this experience.” [*Blank text field*] [Okay]

“Great! Thanks for your answers.”

Subjects were able to report on how well they thought they performed, what was estimated to be the most relevant problem they encountered that interfered with their performance, the severity of the problem, and the severity of its impact on their experience. They were then prompted to describe the intensity levels of five emotion vectors estimated to be relevant to their immediate experience. Finally, they were encouraged to report whatever was on their mind via the text field.

### **CASPER Condition**

The first seven frames in this condition were identical to screens in the first nine frames of the prior two conditions (two ruse questions were omitted to balance for time—see Table 5.7 for which questions were included). The next four frames (and questions) were identical to the questions in frames 10 - 13 of the Venting condition (“Okay. Now how well do you think you did in this game?”, to “How frustrated do you think you got playing this game, all things considered?”).

Unbeknownst to the subject, at this point the CASPER agent assumed control of the questionnaire (see Figure 5.8 for a diagram of the flow of the CASPER agent’s interaction, with reduced screenshots of the agent in action). Based on the subject’s response to the question about frustration, the CASPER agent responded with a sentence paraphrasing the subject’s description of her experience, with a focus on the emotional content. For example, If the subject reported feeling a frustration rating of 2 (from the 0-10 scale), the agent would reply, “It sounds like you felt a little frustrated playing this game. Is that about right?” [Yes] [No]

If the subject pressed the “No” button, the system responds with a *repair* frame, enabling the subject to refine or correct her assessment of how she felt, by means of relevant radio buttons, the display of which was designed to be sensible (i.e. if the subject reported feeling only a “1” on the frustration scale, and the feedback read “It sounds like you felt ever so mildly frustrated playing this game. Is that about right?” in the repair dialog there would be radio buttons that corresponded to the subject feeling “a little less frustration than that”, or “a little more frustration than that”, or “a fair amount more...” or even “a lot more...”—but not “a lot less”). Subsequent paraphrasings of the user’s state were designed to mimic natural language, to make the responses seem less “canned” (see Figure 5.8, and look *very* closely—small text). The agent maintains an integer variable that corresponds to its estimation of the subject’s frustration level. So if the subject initially reports feeling a frustration level of 5 but then elects a repair dialog to amend his answer to “a bit more frustrated than that”, the variable will amend his “frustration estimation number” to be 6.

If, at a paraphrasing feedback screen at any point in this interaction, the subject presses the “Yes” button, the system responds with two feedback strategies: an

empathy response, and a sympathy response, both based on the final frustration estimation number the system recorded. For the “2” used in the above example, these two responses would read:

“It sounds like things could have gone better.

“Good to hear that on the whole, though, the game wasn’t terribly frustrating to play.”

An “8”, by contrast, would have yielded the following initial paraphrased feedback:

“Wow! It sounds like you felt terribly frustrated playing this game.

“Is that about right?”

If the subject pressed the Yes button here, the subsequent screen would read:

“That must feel lousy. It is no fun trying to play a simple game, only to have the whole experience derailed by something out of your control.

“This computer apologizes to you for its part in giving you a crummy experience.”

### ***CASPER Repair Dialogs***

Repair dialogs offered an apology each time, which grew in magnitude with each subsequent one. The idea is that the agent’s role is to facilitate the ability of subjects to express themselves emotionally when already feeling frustration. These screens represent points at which the agent has failed to provide appropriate feedback. There are three such repair screens, and as with this entire agent’s function and purpose, each repair screen should try to appear to acknowledge the probable level of the subject’s frustration.

These levels would be assumed to mount with each successive failure on the part of the system to supply accurate feedback. From pre-tests, it was estimated that some subjects would get to the first repair screen, and only a few would get to the second. However, with the expectation that no one would probably get to the third (and final) repair dialog, a graceful resolution was implemented, in which the agent enabled the subject to express her feelings in her own words. The apology and response in the three repair screens were, consecutively,

**#1:** “Sorry about that.

“To clarify, how would you describe yourself instead?” [Labeled radio buttons, as described above] [Okay]

**#2:** “So sorry again!

“This computer seems to be having difficulty accurately describing your state. Please let's give it one more try. Based on my last attempt to paraphrase, how would you describe yourself? [Labeled radio buttons, as described above] [Okay]

**#3:** “Please accept this computer's sincere apologies for not being able to accurately capture how you're feeling.

People typically feel many different kinds of things in response to a situation like this. Could you describe how you're feeling in your own words? [Blank text field] [Okay]

Finally, subjects in the CASPER condition are presented with an open-ended question (with text field and Okay button) and a “thank you” screen, both identical to those used in the VENT condition.

## Procedure

### ***Recruiting Subjects***

Subjects were recruited via fliers posted on bulletin boards around MIT campus (see a copy of this flier on the first page of Appendix A). The flier simply suggested “Earn \$10 in 40 minutes, and automatically be entered to win \$100...for playing a game!” Beneath, a 2-line description read: “It’s that simple. Come to the MIT Media Lab and evaluate a new kind of computer game. Walk out with hard currency, and a chance to win one of two \$100 prizes.”

Because of the nature of the study and the importance of maintaining the cover story, subjects were not solicited in E15, the Media Laboratory building. In fact, only two subjects who worked in this building were subjects in the study, and both were casually and informally screened pre- and post-hoc (before participating and after the de-briefing) for knowledge of the nature of the research, and belief in the cover story. All subjects were informally interviewed after debriefing about their confidence in the cover story, which was high in all cases, if not unquestioned.

Subjects voluntarily responded to the flier by contacting the author, via contact information (phone and email) printed on rip-off stubs at the bottom of each flier.

### ***Off-site Subject Contact***

Subjects responding by email were then sent an email back, giving broad details of the game and their participation in it (the cover story), including monetary incentives, along with a request for scheduling, and suggested times for coming in. Subjects were also told that the first part of the study involved filling out and sending back a brief, emailed questionnaire (see the example of first contact response and questionnaire emails in Appendix A.2 and A.3).

Subjects responding by phone were given a verbal version of the same material, were scheduled on the spot, and arrangements were made for them to receive the emailed questionnaire (all but three had email accounts). Other questions about the game, the nature of the research group, etc. (there were few of these) were deflected with the blanket response: “It’s important that subjects be fresh and untainted, so we ask that you hold your questions until after your participation.”

Unbeknownst to the subjects, the questionnaire sent to them was actually Mehrabian’s Trait Arousability Inventory (Mehrabian 95). It was anticipated that the nature of the questions asked on the personality inventory (see the questionnaire, and its containing email, in Appendix A.3), might seem somewhat odd and personal



for an experiment evaluating a game, so several precautions were taken. First, the nature of the questionnaire was incorporated into the cover story, by simply stating in the email that we needed to get a sense of their personality in order to properly evaluate their reactions to the game. As the email stated, “In order to gauge your evaluation of the game, we would like to get an idea of what you are like.”

Time was also used in the effort to mitigate any effects from the inventory’s nature: In all but two cases, subjects were emailed the trait arousability inventory at least several days in advance (in most cases at least 3 days in advance). Most subjects returned the questionnaire the same day it was received, and all returned the questionnaire at least 24 hours before participation. Three subjects did not have access to email; of these, two came in to the lab to fill out the questionnaire on the spot, at least 48 hours before participating. The third filled out the questionnaire at her home, and returned it to the administrator while he waited (he had brought it to her residence 3 days in advance of the subject’s participation).

Once subjects returned their filled-in questionnaires, they were sent a brief confirmation note. The evening before their participation, subjects were sent a reminder email (or phone call if email was not available). See examples of both these emails at the end of Appendix A.

### ***Motivating the Subject***

In order to induce a strong frustration response in subjects, motivation to do well in the game was seen as a critical factor. A number of incentives, monetary, social and psychological, were proffered. The monetary incentive was simple: \$10 to participate, and the chance to win one of two \$100 prizes, given out to the top two scorers of the game. Psychological and social incentives were built into the cover story and its supporting materials at various points before subjects began playing Game 1. In the first contact email, for example, subjects were told the following:

In this game, intelligence matters even more than reflexes, so novice game players with smarts have a better chance of winning than experienced gamers without them.

This strategy was meant to accomplish two things: 1) To add incentive to perform well in the game, since accomplishment was associated with intelligence level; and 2) To try to baseline confidence levels between novice and experienced game players.

During the verbal briefing, the script the administrator read to the subject elaborated and clarified this double-intention ploy (See Appendix B.1 for the text of this script). Subjects are told that

“We’ve play-tested the game on experienced gamers who play an average of 10 or more hours per day, and we’ve also tested the game on people who have never picked up a mouse to play a game before in their lives. What we found is that novice players who are intelligent and use their heads do as well or better at this game than experienced gamers who just rely on their reflexes. You need normal reflexes for this game, but if you are a smart person and use your head, you have a good chance of doing well.”

Further, in the on-screen instructions for game 1 (that subjects are reminded to read by the administrator), subjects are told the following:

Remember that the higher your score, the greater your chance of winning one of the two \$100 prizes! And, while you will certainly need your reflexes for this game, you will need your intelligence even more. To play well you will have to make fast, intelligent choices. You will be competing against students from MIT and Harvard, so play as fast, accurately, and intelligently as you can.

Good luck!

Finally, subjects were given social incentives to do well. They were told that they will be play-testing a prototype of a new game, for which the game's developers "need you to play like it's a real game situation in order to get a good sense of how players will feel about the finished game."

### ***On-Site: Running the Subject***

At least 10 minutes before the subject arrives, the computer system was checked to make sure that the operating system is up and running, and that the experimental software program (all six conditions in one) had been started, primed, and set to the condition planned for that timeslot. Allotting conditions to timeslots was done in semi-random fashion; sets of 2 conditions at a time tended to be run together, since in the beginning of the study the two VENT conditions were run first, to evaluate if differences in self-reported frustration levels were showing up in the results, between DELAY and NO-DELAY conditions.

When the subject arrived at the site, s/he was greeted by the experiment's administrator, who was told to be friendly to the subject, but in a natural way. The subject was asked to have a seat. The administrator sat down in a seat 4-5 feet away from the subject with a \$10 payment voucher, a numbered consent form, and a script that the administrator proceeded to read to the subject (see Appendix B.1 and B.2 for the contents of these two last documents, respectively).

Notes on the script: The script was designed to add substance and believability to the cover story, while at the same time assisting in motivating the subject to play their best, and attempting to "emotionally baseline" subjects (attempt to instill in all subjects an affective state of positive valence and moderate emotional arousal by the time they sit down to play the game). The script appears to the subject to be introducing them to the world of Web-based computer game development. After describing the role the subject is to play ("We have built a prototype of an Internet-based computer game. We'd like you to work with this game, and tell us how you feel about it."), the administrator described that one aim of the study is to test a new Internet communication protocol that they were told speeds up the graphics on Web-based games, which are normally really slow. (The game was built to appear somewhat faster than this standard, but not as fast as most modern computer games. The intention was to ensure that the game would not be *too* inviting or fun to play.) The script described these speed differences, that seasoned gamers understand and appreciate, but others may not. In pretests, subjects in the no-delay condition were often reporting high levels of frustration from this speed difference, so the script was adapted in this way to describe the speed issues up front, to baseline expectations for the slowness.

The script then described that the game is an early prototype, so a lot of the "bells and whistles" planned for the game have not yet been implemented. It was described that this stage is a key one in the game's development, since it's the first

time that play-testers can come in and begin evaluating the game for important criteria. The purpose of this rigmarole was to provide subjects with a good reason for why the game is so simple and not very exciting to play—but also to motivate them; for one thing, by giving them the title of “play-tester”.

Subjects were then told that “we need them” to play as hard as they can (in order to simulate a “real game situation”), and that the monetary incentive is to help motivate them to do as well as they can, to counteract the effects of the fact that the game at this stage may not be as “immersive” an experience as it will eventually be.

Subjects were then reminded that the game is based on intelligence even more than it is on reflexes, and that if they are really smart and use their heads, they will do well. This passage was intended to do two things: 1) Baseline confidence in performance, by enfranchising novice players with a description of a real possibility of winning, and 2) Make the desire for winning “personal”, by encouraging the idea that winning is associated with intelligence (hypothesized to be a strong motivator among the high-pressure ranks of MIT students and staff, which comprised nearly all subjects).

A brief description of what the structure of the subject’s experience with the game system would be was then presented to the subject, along with a casual reminder to read both sets of directions, since directions for Game 2 were different from Game 1. Notably, subjects were *not* told the specifics of the differences between games up front. In order to ensure no administrator bias in the subject’s decision for playing Game 2 for as long as they decide to, subjects were to read about what was expected of them on-screen, with no contact from humans (or experimenter bias), one way or the other.

Finally, subjects were presented with the payment voucher. This was presented to subjects up-front (before the game) in order to help bias initial affect to be more consistent across subjects, to a mildly positive tenor (affect tends to be improved with the receipt of gifts or other forms of payment). They were then given the consent form to read and sign, which included a separate consent form to take video footage of the subjects’ participation (Appendix B.2).

Upon signing this form, subjects were politely asked to follow the administrator into the experimental chamber, where the administrator turned on the video camera (if the subject had signed the video consent form), and reminded the subject that “everything should be self-explanatory on the screen. If there’s a problem, I’ll be right outside here; otherwise, GOOD LUCK!” The administrator closes the door. The subject was then left alone to read directions for Game 1.

### **Game 1**

Subjects sat down at the computer, to find what look like two windows just barely overlapping on a Macintosh desktop (see figure 5.9 for a screen shot of what all subjects first see on screen). The menu bar is greyed out, and the two windows are immobile. The bottom-right window looks like a SimpleText window, but with a “Next Page” button on it, and an unfamiliar application icon on the right-hand side of the menu bar (of the author’s own design). This window is currently the active one of the two (it overlaps the other, and the window’s grabber bar is high-contrast and active. This window’s title is “Instructions”, and several paragraphs of text are in this window, along with “Page 1 of 3” and the “Next page” button.

The other window is clearly a greyed-out Netscape browser window (familiar to most or all subjects), with what looks like a brightly-colored game inside it. The URL (Internet Web address) for the “site” also supports the cover story, and is displayed as:

`http://white.media.mit.edu/projects/synchttp/lost_treasure/v07alpha.shq`. Presumably, the server name is “white.media.mit.edu”, an old, but actual, Media Lab server name. One of the subdirectories is entitled “synchttp”—a plausible name of a working title of a protocol for Internet communication for multi-player games; the title of the game is “The Lost Treasure”, which is also displayed on the game’s background graphic (see Figure 5.9, below). The file name contains the term “alpha”, along with a small version number (“v07”), in line with the subject’s knowledge of the cover story (they are told they will be testing a “first playable” version, which is akin to an Alpha version).

Subjects read the directions for Game 1 (see Appendix B.3), to play this “Web-based” game (a deception—the game is actually mounted locally). They are told to finish with as many points as possible—i.e., gather up as much treasure as they can in five minutes—and given incentive to do so. They were told that the computer will stop the game automatically after 5 minutes. They then are told to click on the Netscape window to highlight it, and to then press the Start button when ready.

Subjects then played Game 1, the details of which were described in the preceding section.

#### ***The DELAY Condition***

During Game 1, approximately half of all subjects (all subjects in the DELAY condition;  $N = 37$ ) experienced what are orchestrated to seem like severe Web delays. All subjects in the DELAY condition experienced the exact same “delays”, which appeared to occur at random intervals, for what appeared to be random amounts of time.

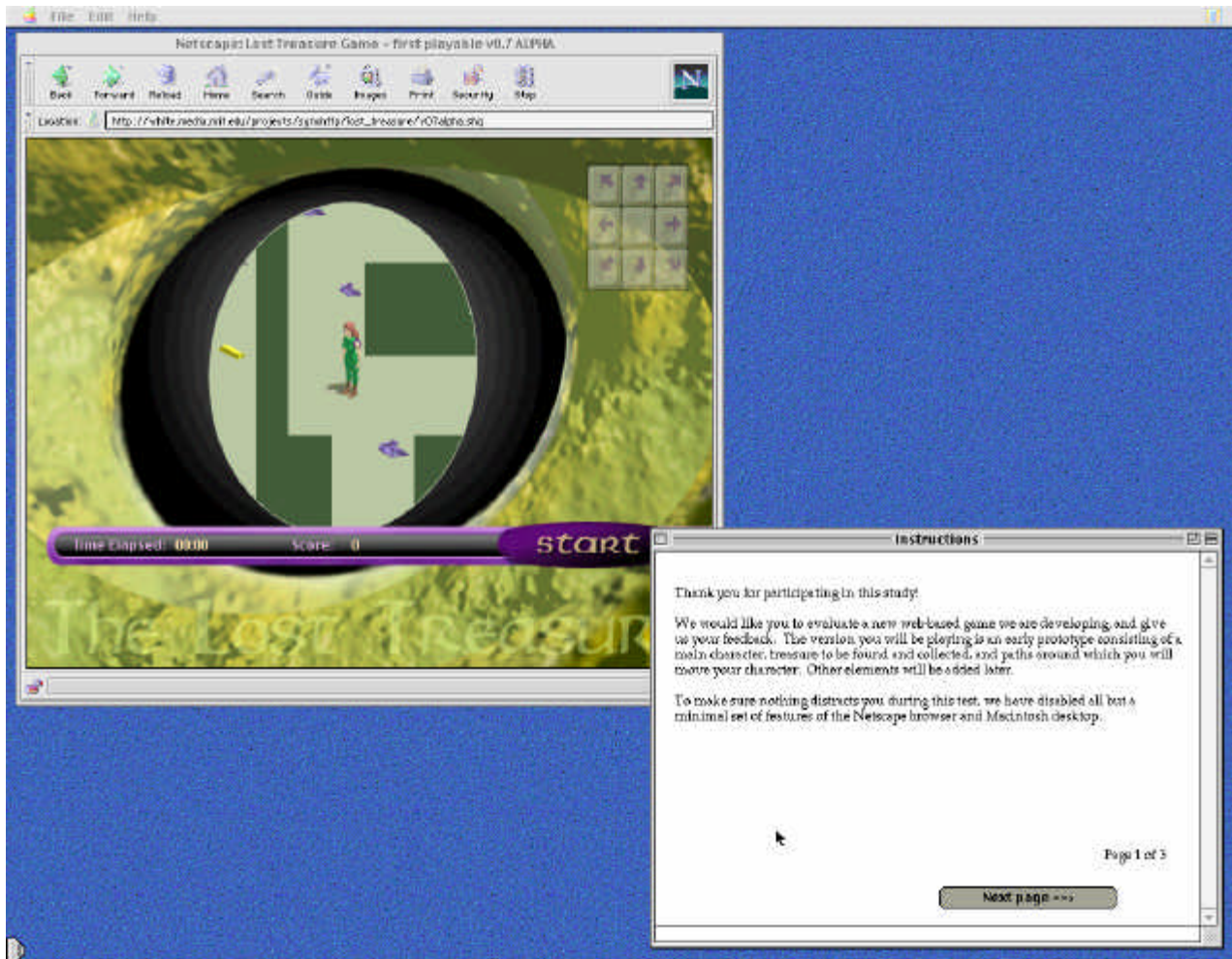


Figure 5.9: A screenshot of the entire screen that each subject sees for the first time upon entering the chamber.

The apparent delays were designed to simulate real Web delays: The Netscape logo (the large version, the size of a postage stamp) begins animating at 4 frames per second, a typical speed for this animation. A legend appears at the bottom of the Netscape browser window, saying “Connect: Contacting host white.media.mit.edu...”, a typical message displayed by Netscape here during a connection delay, and displaying the same server name as in the URL field.

During a delay, the subject is unable to move the character on screen; none of the keys work. The subject is functionally frozen on screen. Meanwhile, the timer continues to iterate. The plausible scenario for the subject here is that some elements of the game are mounted locally (such as the timer), but information on other elements (such as the character’s position) are mounted remotely; i.e. on the network.

The schedule for delays is as follows:

Delay begins at: (seconds into Game 1)	Interval of play before next delay	Delay lasts for: (in seconds)
81	81	6
132	45	6
158	20	17
188	13	8
211	15	8
238	19	12
252	2	10
279	17	8
291	4	6

Note that the delays are “bottom-heavy”, designed to be frustrating in the context of the game: The first delay does not occur until 1:21 into the game. The next delay does not occur until 44 seconds after the first delay ends. By the end of the game (which lasts for 300 seconds), though, the delays are coming fast and furious. There are 3 delays in the last minute, totaling 24 seconds; the last two delays occur with only four seconds between them, and after the last delay there is only 5 seconds to go before the game ends. Subjects often reported being very close to high-point treasure with plenty of time to go, but unable to reach them, or collect all of the treasure pieces there, frozen from the delays.

#### ***The NO-DELAY Condition***

The other half of all subjects played Game 1 with no such delays.

#### ***The Interactive Questionnaires (Response Conditions)***

All subjects (DELAY and NO-DELAY) then followed directions on screen (see Appendix B.4), to begin answering questions back in the Instructions window, which was greyed out (with all text invisible) while the Netscape Browser window appeared to be “active”. The questions that subjects were to answer differed across subjects, corresponding to different conditions for the experiment, and to the rows in Table 5.1, at the beginning of this chapter. The design of these three “interactive questionnaires” is detailed in *Design of the Interactive Questionnaires and CASPER Agent* section of this chapter, above.

#### ***All Subjects Play the Same Game 2***

All subjects in all conditions were then shown what they were led to believe was a computer-generated summary and assessment of their performance in game 1 (see Figure 5.10, below). The text, which was rigged to look as though it were alternately “canned” and “dynamic”, featured dynamic-generated text of their score on game 1, and the time of the game (05:00 in all cases). In an effort to try to equalize the confidence levels of all subjects before they played game 2, the text offered an assessment of the game that looked dynamic, but which was actually the

same for all subjects. For subjects whose confidence after playing game 1 was high, this assessment was not projected to affect them very much. For others who feared they performed badly in Game 1, this statement would have a much greater effect, helping to boost self-confidence.

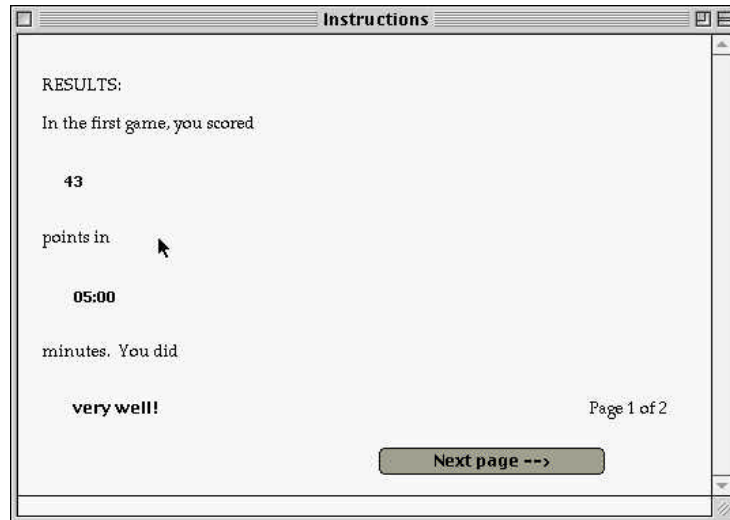


Figure 5.10: Game 1 performance summary and assessment, just before directions for Game 2.

A subsequent Instructions-screen-full of directions asked the subject to play the game again (see a screen dump of these directions in Appendix B.5). This time, however, they would play for an open-ended amount of time, after a minimum 3 minutes (to get players started). They were told that this time, the researchers were testing out a different strategy for Internet communication, so the game may or may not seem different to play. They were also told that this time they were not going to be playing for a prize, and that their score would not be counted toward any prize at all. This was a non-competitive version of the game, they were informed. Subjects were told that the Quit button in the game window would be grayed out for the first 3 minutes, becoming enabled after that point. Subjects were told that they could then quit at any time after that; at 20 minutes, however, the game would automatically stop. Subjects were informed that even if they played for the full 20 minutes, that their entire participation in the study would not take more than 45 minutes. In this way, players will not feel pressure to play longer than they want to, and will therefore be free to play as long as they like.

All subjects then played the same Game 2, which was the same game with the same rules, in the same maze they were in for Game 1. Game 2 began in the same general area of the maze where Game 1 had left off, and did not have any built-in delays.

#### ***Exit Questionnaire & Debriefing***

Once Game 2 ended, either voluntarily or not, a dialog box on the screen notified the subject to get up, leave the room, and find the administrator, who had the paper Exit Questionnaire for the subject (See Appendix B.6) The administrator led the subject to a desk with a pen and the questionnaire, which was numbered with the same

number as the Consent Form (the subject's ID#). Upon filling out the Exit Questionnaire, which asked self-report questions as described at the end of Chapter 4, subjects were then handed the debriefing statement (see Appendix B.7), asked to read it on the spot, and told by the administrator that he would be happy to answer any and all questions the subject had after reading it. (Note: There were a variety of responses from subjects at this point, ranging from complete surprise to great interest in the research and inquisitiveness, to complete indifference. No one expressed anger.)

Note that scores achieved by subjects in the DELAY condition in Game 1 were expected to be lower on average, and not comparable in any fair way to those in the NON-DELAY condition. This disparity was addressed in the prize-awarding scheme, by actually awarding one US\$100 prize to the top scorer in each of the two (DELAY/NO-DELAY) conditions. This award scheme was not part of the cover story, and was only told to subjects after the debriefing.

## Measures

The following measures were taken in this study:

- **Emotional Arousability.** Also known as *Trait Arousability* (Mehrabian 95), this measure was taken via a 34-question inventory established by Mehrabian, in the form of an email questionnaire filled out and returned by subjects well in advance of their on-site participation. For more on this metric, see the section entitled “Evaluation Criteria, Predictions, and an Overview of the Protocol” at the end of the preceding chapter.
- **Visual record.** A video camera recorded subjects' interaction with the software, their entire time in the experimental chamber.
- **Gender.** Recorded in the log files, as a direct question.
- **Age Range.** Recorded in the log files, as a direct question.
- **Gaming Experience Level** was measured by collecting self-report data from the interactive questionnaires between the two games (also recorded in the computer log file). Specifically, this measure involved taking the mean of the responses to two questions on computer-game-playing frequency asked of all subjects in all three conditions. The two questions were:
  - 1) “Okay. Now how often have you played computer games or video games before?”
  - 2) “And how often (if ever) have you played graphical adventure games on the Web, or graphical MUD's?”
- **Game-Playing Behavior.** This parameter was measured by the experimental computer system itself, which recorded the number of seconds each subject played Game 2, as well as the final score of Game 2, in an automatically-generated log file.
- **Self-report for felt emotion,** from the Exit Questionnaire, estimating how subject felt now, when she first entered the room, immediately after Game 1,



immediately after answering the interactive questionnaire, and immediately after game 2:

- *Frustration level*
- *Satisfaction level*
- *Anxiety/tenseness level*
- *Fun level*
- *Anger level*
- ***Other self-report information***, such as how much subjects liked Game 1, how much they liked Game 2, why they played game 2 for as long as they did, and (lastly), whether the questionnaire between the two games made them feel better, worse, or about the same.

## Data Treatment

### ***Video Record Not Considered***

The data considered in this study were the behavioral and self-report measures. The video record was not used in the present analysis, but was captured for a variety of reasons:

- as an auxiliary means of data analysis, were the behavioral and self-report measures to yield nothing substantive (the behavioral results obtained obviated the analysis of this data);
- as illustrative data, for the purpose of making presentations in the research community that can show the experimental chamber, and/or subjects' visible and auditory reactions to the software;
- as a means for constraining subject actions. The running camera's presence in the room was intended to foster a sense of "being watched", which, among other things, would act to inhibit curious subjects from getting up and examining the hardware for clues—an act that might have compromised the cover story.

### ***Data Cleaning***

Data cleaning is an established, standard practice across social science communities, to make the data more uniform. Some researchers in these communities have been known to take some liberties in this process, thereby "massaging" their data to conform to a priori considerations, and pushing the limits of scrupulousness. This author faced similar temptations, but resorted to an established, conservative protocol for data cleaning, that conformed to the following criteria:

- ***Outlier control.*** Data that fit a pre-determined criteria for “outlier status” were altered in a standardized way, according to the following procedure: Data that were two or more standard deviations (SD) from the mean, and were alone in that category, were considered to be outliers. Outliers were identified, noted in the comment field of the cumulative data file, and automatically converted to the mean. Professor Moon, one of the readers of this thesis, noted that many researchers maintain a standard outlier cut-off point of a much-less-conservative 1.5 SD, and some even use a 1.0 SD standard. Accordingly, a 2.0 SD standard is widely held as a sound, conservative figure that only deals with extreme cases for data cleaning.
- ***Uniformity of subject group size.*** Following the advice of Professor Moon four subjects in three condition groups were trimmed to make the number of subjects in each group more homogeneous (either 5 or 6 men and women in each of the six conditions, and either 11 or 12 total subjects in each condition—if a group had a surplus of one sex, that group was trimmed to 6 of that sex). See the *Subjects* section at the beginning of Chapter 5 for the criteria used for cropping subjects from groups.



## CHAPTER 6

## Results

All analyses were based on either a full-factorial ANOVA or ANCOVA. When appropriate, planned orthogonal comparisons were conducted using Dunnett's  $t$  test (see Winer 91), which adjusts for the inflated significance levels associated with multiple comparisons. All results were based on two-tailed tests.

## Frustration - Manipulation Check

A *manipulation check* was performed to confirm that the DELAY condition produced significantly more frustration than the NO-DELAY condition. In the Exit Questionnaire, participants responded to the question, "How much frustration did you feel immediately after playing Game 1?" A full-factorial ANOVA (see Appendix C.2) revealed that participants in the DELAY condition ( $\underline{M} = 4.8$ ) rated their frustration level as being significantly higher after the first game than participants in the NO-DELAY condition ( $\underline{M} = 3.56$ ),  $\underline{F}(1, 64) = 4.54$ ,  $p < .05$  (where  $\underline{M}$  = mean;  $\underline{F}$  = the *F-ratio*, the statistic used in ANOVA to determine statistical significance; it is the ration of between-groups variability to within-groups variability;  $p$  = *p-value*, the actual probability of making a type I error). There were no other significant effects with respect to this variable.

## Game-Playing Behavior

The key prediction in this experiment was that subjects who were experiencing high levels of frustration, resulting from playing Game 1 in the DELAY condition, would

feel relief from this state—calm down, and feel better in general—immediately after experiencing the CASPER condition, in comparison with the CONTROL condition and, to a lesser degree, with the VENT condition. Based on this improvement in affective state, subjects were expected to feel more positive affect toward the task, as well as to the source of their frustration—the game itself, and the computer system on which it was played. Subjects in the DELAY/CASPER condition were therefore predicted to play longer, and perform better (i.e. score more points) than subjects in either the DELAY/CONTROL or DELAY/VENT conditions.

The results supported this prediction. More specifically, a full-factorial ANOVA (see Appendix C.3) revealed two main effects. First, there was a main effect for feedback-type,  $F(2, 64) = 8.00, p < .01$ . Planned orthogonal comparisons indicated that participants in the CASPER condition played Game 2 for a significantly longer time than participants in either the CONTROL condition ( $t(45) = 2.63, p < .01$ ; Appendix C.12), or the VENT condition ( $t(44) = 3.97, p < .01$ ; Appendix C.13). However, there was no significant difference in the game-playing behavior between participants in the CONTROL and VENT conditions (Appendix C.10; see Figure 6.1, below).

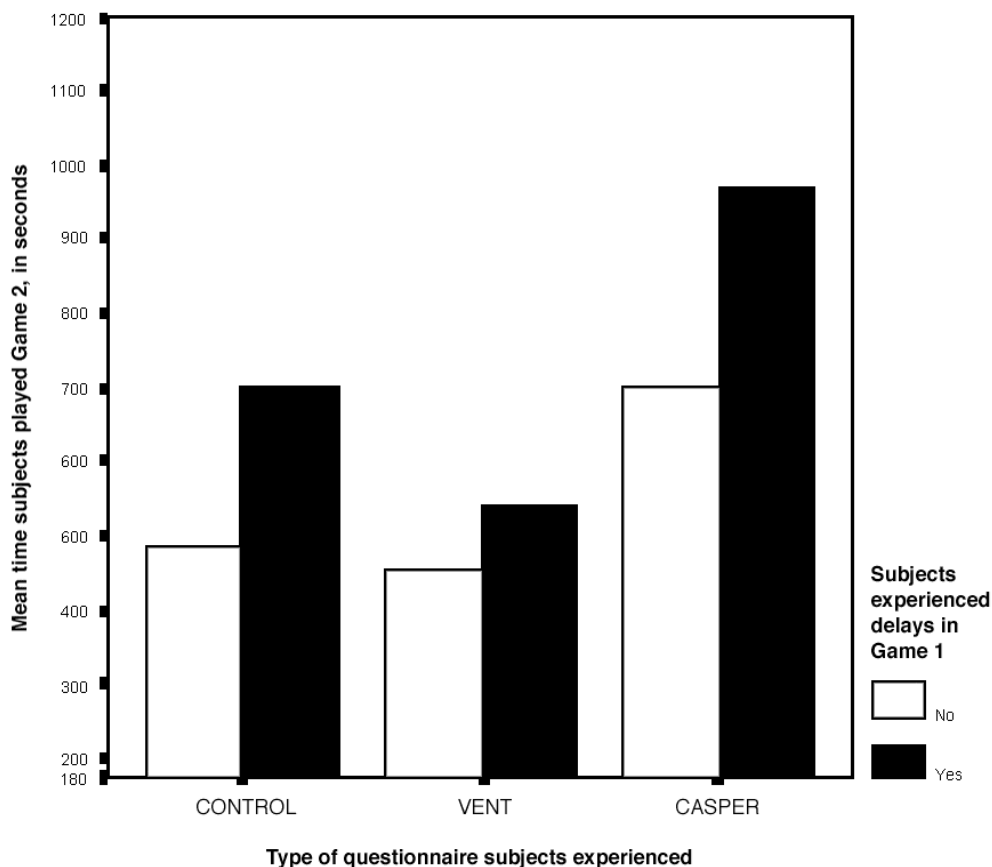


Figure 6.1: Mean times that subjects played game 2, by response type and DELAY/NO DELAY conditions. (Subjects could play for  $180 \leq T \leq 1200$  seconds—3 minutes to 20 minutes).

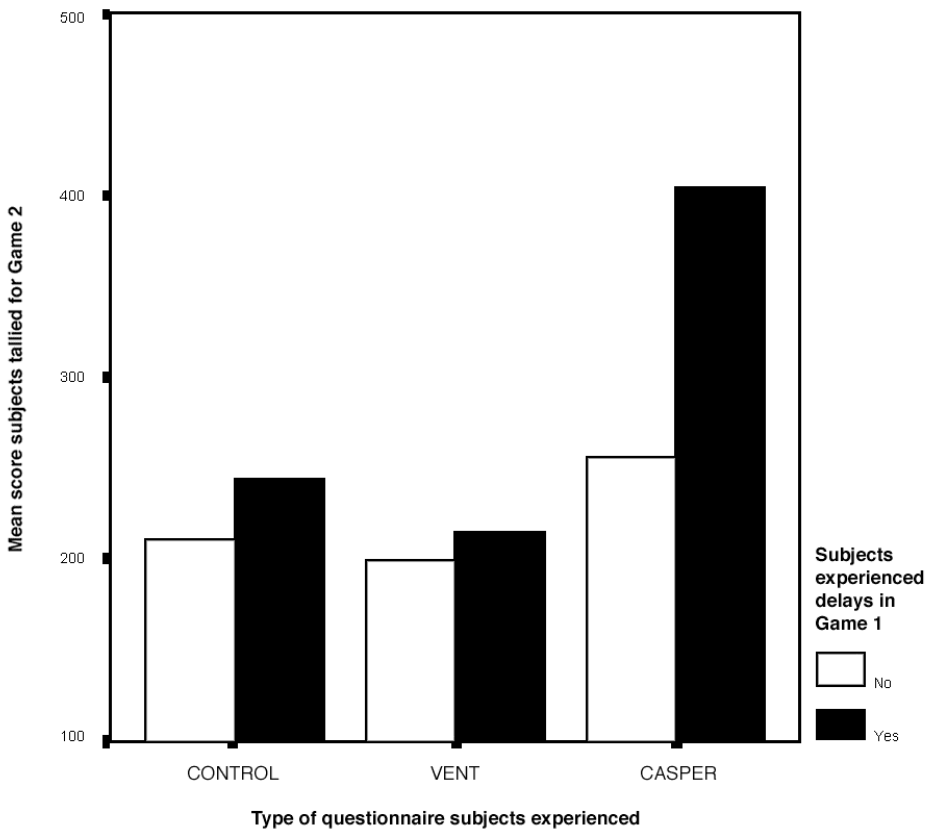


Figure 6.2: Mean scores that subjects tallied in game 2, by response type and DELAY/NO DELAY conditions. Scores were found to be highly correlated with time played (Figure 6.1)

There was also a main effect for frustration, such that participants in the DELAY condition ( $M = 735$ ) played Game 2 for a significantly longer time than participants in the NO-DELAY condition ( $M = 545$ ),  $F(1, 64) = 9.20$ ,  $p < .001$ . The interaction effect was not significant with respect to this variable. In other words, play times for the DELAY conditions were consistently higher than those for the NO-DELAY conditions. Reasons for this phenomenon will be discussed in Chapter 7.

A two-part analysis was conducted to test the hypothesis that subjects in the CASPER condition would be calmer and more alert during Game 2, and consequently would score more points in a given amount of time, vs. subjects in the VENT and CONTROL conditions. First, a frequency dichotomy analysis was conducted (Appendix C.8), and an analysis of variance conducted using the frequency dichotomy analysis results as an independent variable. Then, an analysis of time on task was conducted (Appendix C.9). A main effect was found for score, but only inasmuch as score was found to highly correlate with time spent playing the game. Score, however, was not found to be higher *per time played* in the CASPER condition than the other conditions.

### Satisfaction Level After Game 2

With some suspicions (noted at the end of Chapter 4, above) there was reason to believe that self-report measures such as satisfaction levels after Game 2 (reported in the Exit Questionnaire) should reflect a reported benefit from interaction with the CASPER agent, in contrast to the other interactive questionnaires. A full-factorial ANOVA (Appendix C.15) was performed to test this possibility, and yielded no effect, either favorably or unfavorably.

### Game-Playing Behavior, Controlling for Gender

As discussed above, there was some limited reason to believe that females would be more likely to respond positively to the CASPER feedback than males. A three-way ANOVA (Appendix C.4) was performed to test this possibility. When gender was added to the analysis, the two main effects for feedback-type and frustration remained significant. However, the results showed no significant main effect for gender on game-playing time, and no significant interaction effect with respect to this variable. Gender thus appeared to play no significant role in this experiment.

### Game-Playing Behavior, Controlling for Emotional Arousability

As discussed above, there was also reason to believe that individuals who are high in emotional arousability would respond differently in this experiment than individuals who are low in emotional arousability. An ANCOVA (Appendix C.5) was performed to test this possibility. When emotional arousability (as measured by Mehrabian's *Trait Arousability* metric (Mehrabian 95; see Methods section, above) was introduced as a covariate into the analysis, the two main effects for feedback-type and frustration remained significant. However, the results showed no significant effect for the emotional arousability covariate. In other words, emotional arousability appeared to have no impact on the results in this experiment.

In addition, when the emotional arousability variable was recoded as a dichotomous variable (footnote: this was accomplished by first ascertaining the general mean for this variable over all participants. Participants scoring above the mean were then recoded as "1" and participants scoring below the mean were recoded as "0."), a three-way ANOVA (Appendix C.6) showed no significant interaction effect with respect to this variable.

### Game-Playing Behavior, Controlling for Gaming Experience Level

An ANCOVA (Appendix C.7) was performed to test the possibility that participants' prior gaming experience level affected the results. When gaming experience level was introduced as a covariate into the analysis, the results showed that participants

with more experience did tend to play Game 2 longer than participants with less experience; however, this effect did not approach statistical significance. Moreover, even when controlling for this variable, the two main effects for feedback-type and frustration remained significant.



## CHAPTER 7

# Discussion

## Discussion of Primary Results

### *Manipulation Check Successful*

To assess the effectiveness of an agent designed to relieve strong frustration in human users, it is necessary to establish that the people on whom it was tested were, in fact, feeling demonstrably frustrated when they interacted with the agent. Emotions, as discussed in Chapter 2, are by nature hard to describe, let alone isolate and verify in humans. Moreover, as noted in Chapter 4, the social science community has long mistrusted self-report data used to describe affective state in humans. However, sometimes the best way of telling whether the lights are on or off in a room is to simply ask someone in the room. By this logic, we chose to simply ask subjects.

The answers, from subjects' open-ended responses, speak volumes. Said two subjects in the VENT/DELAY condition, when asked to "please describe your experience playing this game" at the close of the interactive questionnaire (see Appendix C.17 for all these responses), said:

I was unsure if I was playing the game correctly on first try. The lag time was quite frustrating. Since I was trying to get a high score and was racing against the clock, it made me feel powerless to excel since I was dependent upon the response time

Overall, it was frustrating because it was like the network connection was hindering my efforts to achieve a high score.

Subjects in the CASPER/DELAY condition, by contrast, tended to report a more favorable opinion of the game, as well as milder reports of frustration levels (described as “annoying”) compared to the reports from the VENT/DELAY condition, above:

The game was interesting to play but there were constant delays that kept pausing the game. Watching the game stop every 30 seconds was annoying.

It was OK. It was annoying to watch time tick away when the host was being contacted.

In contrast, two typical responses to this question in the VENT/NO DELAY condition were:

Not much reaction except a bit of frustration in how slow the character walks and turns.

I didn't find the game very frustrating, rather i found it sort of boring, largely because it only involved walking around and piking stuff up, not avoiding enemys or any other intelligence based tasks.

Subjects in the CASPER/NO DELAY condition responded with mixed reviews. Some were similar to those above:

Wanting to do well but seemingly not able to do so since the graphical interface was so slow. Also got boring.

But a number were more upbeat, emphasizing positive attributes of the game and, notably, a non-judgmental assessment of their own behavior:

It was a bit of fun, I felt like it took a while to get the hang of it.

I liked the fact that it was a woman and the instructions were easy to follow, although it took me a little while to get the hang of it.

The free-response, self-report data sampled above seemed to indicate a tendency for reports of higher frustration levels in the VENT/DELAY condition than in the CASPER/DELAY condition, and slightly milder reports in general in both CASPER conditions, compared with the VENT condition (note that the CONTROL condition's questionnaire did not contain any means for reporting subjects' impressions). Such responses are difficult and time-consuming to formally code, however. Their coding and analysis, beyond the informal examination described above, were therefore considered to be beyond the scope of this thesis. Again, see Appendix C.17 for a complete record of these textual responses.

The data that was used in the statistical analysis to verify the manipulation check, on the other hand, was derived from the numerical self-report data obtained from subjects' responses to the Exit Questionnaire. Subjects were asked to rate how frustrated they felt they were at several key points in the course of their participation in the experiment. One of those key points was immediately following Game 1, and the responses to this question enabled self-report to be taken across all conditions.

The results from this manipulation check (described in Chapter 6, above) confirmed that the experimental protocol was effective in achieving that necessary, intermediate goal: The DELAY condition was shown to elicit significantly higher levels of frustration in subjects than the NO-DELAY condition. Note: This result,

in addition to the main behavioral result, was found both before and after the data were cleaned.

### **Significant Behavioral Results Found**

**Time playing Game 2:** In contrast to self-report, as stated at the end of Chapter 4, behavioral measures are the strongest, most impressive and most reliable indicators of changes in internal state. Under statistical analysis, the behavioral measures of those who experienced higher frustration levels in the DELAY condition revealed no meaningful difference between the amount of time that subjects in the CONTROL and VENT conditions played Game 2, or how many points they scored. However, those who experienced the CASPER agent displayed clear, unmistakable evidence of a significant improvement in both the times that subjects played Game 2, and how many points they scored (see Figures 6.1 and 6.2). Indeed, between the time when subjects *could* quit out of Game 2 and when they *did*, the average time that CASPER subjects spent playing the game was 179% of the combined mean of times played by subjects in the other two DELAY conditions, using the following equations:

$$\begin{array}{l}
 1) \quad 967 - 180 \qquad \qquad \qquad = 787 \\
 2) \quad \frac{(698 - 180) + (540 - 180)}{2} \qquad \qquad \qquad = 439 \qquad \qquad \qquad 3) \quad \frac{787}{439} = 1.79.
 \end{array}$$

where 967 = the mean raw time in seconds played in the CASPER/DELAY condition; 698 = the mean raw time subjects played the CONTROL/DELAY condition, 540 = the mean raw time subjects played Game 2 in the VENT/DELAY condition, and 180 = the amount of time subjects were requested to play before being able to quit (and as noted in Chapter 5, the game's Quit button behavior reinforced this request).

Subjects in the CASPER/DELAY condition played much longer—over 5 and a half minutes longer, on average, than the other two conditions combined. For a game with a 17-minute playing window, this is a strong result. (Note: The actual figure is 5:48, obtained from subtracting instead of dividing in equation 3, above.) It should also be noted that five of the twelve subjects (42%) in the DELAY/CASPER condition played for the maximum time allowable, and had to be automatically stopped from playing by the system itself. It is unknown how long these subjects might have continued playing without this upper limit in force. In contrast, the only other condition with more than one subject who played all 17 minutes was the CONTROL/DELAY condition, in which there were three such subjects. The CASPER/NO DELAY condition had one maximum-time player; no other condition had any such players.

**Score of Game 2:** Of points scored in all conditions, subjects in the CASPER/DELAY condition scored many more points, on average, than any of the other conditions (as shown in Figure 6.2). A statistical analysis for time on task was done as discussed in Chapter 6, and revealed no strong relationship between score and time across conditions, other than high correlation between the two variables. It

would seem, then, that time is a good predictor of score. The prediction, then, that scores should be higher per unit of time in the CASPER/DELAY condition, was not confirmed. It is unclear, however, whether subjects *could* have scored higher if they were less frustrated, since this aspect of the game was never established.

## Unexpected Results

### ***A Rebound Effect for Game 2***

The no-delay condition did not conform to expectations. It was expected that the NO-DELAY condition would yield players who were less frustrated; hence, they were expected to play longer in the CONTROL condition than their counterparts who experienced delays (and, therefore, higher levels of frustration). Yet results showed instead that subjects in the CONTROL/DELAY condition played Game 2 *longer*, on average, than subjects in the CONTROL/NO DELAY condition. Indeed, since all subjects in DELAY conditions played Game 2 for longer than subjects in the NO DELAY conditions, it was suspected that there all behavioral results were skewed; it seemed to suggest that the presence or absence of delays was the single most important factor in the time that subjects played Game 2. Was this true? If so, why was that?

First, there was indeed a main effect for DELAY/NO DELAY on the times that subjects played Game 2. However, a reason for this effect was found. It was noted that, apart from experiencing a game with delays (and consequently higher frustration levels) in the DELAY conditions, these subjects were also experiencing something else: In an effort to make subjects' experience uniform across all conditions in Game 2, the protocol design mandated that Game 2 be identical in nature across all conditions. This uniformity meant that for those in the NO DELAY condition, Game 2 was nothing new at all—in fact, it was exactly the same, simple, dull game they had played in Game 1. For those in the DELAY condition, by contrast, Game 2 was a very different game from Game 1. Not only was the action smoother, faster, and more uniform than it was in Game 1, but the single most frustrating problem experienced in Game 1 was suddenly fixed in Game 2. Instead of relative tedium, all subjects experienced relatively more excitement, better action, and a blissful end to the problems in Game 1. No wonder subjects in the DELAY condition played longer in Game 2. Known as the *Rebound Effect*, this relatively common phenomenon makes it impossible to compare probable frustration levels across DELAY and NO DELAY conditions.

However, the manipulation check (a test to determine that extra frustration was experienced by subjects in the DELAY conditions) was confirmed by self-report data, as described above. The fact that there is a rebound effect does not, therefore, diminish the power of the results found between the response conditions *within* the DELAY condition. Indeed, the significant behavioral results found between CASPER/DELAY subjects and the VENT/DELAY & CONTROL/DELAY subjects, described above, cannot be explained by the rebound effect.

### ***A Floor Effect Found for Frustration Level***

Laboratory experiments, by their very nature as orchestrated, controlled events, tend to be somewhat difficult environments in which to elicit genuine human feelings—particularly strong, negative ones. The experimenters were initially concerned, because of these inherently sterile conditions that tend to be the rule in the laboratory, that subjects in the “high-frustration”—i.e. DELAY—condition might not get as frustrated as the study required them to be in order to show an ameliorative effect. Consequently, the protocol was designed to bring all elements reported in the literature to be involved in a classic frustration response (see Chapter 2) into stark relief. The goal was established and reified by the script read to subjects on-site, by the Consent Form, and by the directions for Game 1 (Appendices B.1-3): Score as many points as possible in Game 1. Motivation was designed to be multi-pronged and intense, as described in the *Procedure* section in Chapter 5. And finally, the thwarting of the goal as presented in Game 1 of the DELAY condition was iteratively designed to be as annoying and frustrating as possible, again as described in Chapter 5’s *Procedure* section.

The results seem to suggest that the intent to establish the frustrating stimulus was successful, but perhaps overmuch. In particular, the results show a *Floor Effect*, in which all subjects felt some frustration in trying so hard to achieve their goal. The manipulation check showed that subjects in the DELAY condition experienced significantly more frustration than did those in the NO DELAY condition, but nearly all subjects reported feeling at least some frustration after Game 1. This coordinated effort to put pressure on subjects to perform well, coupled with the tedium of the game’s design, the rough-edged nature of the game (due to both purposeful and incidental bugs, all of which supported the cover story and which were described in the *Apparatus* section of Chapter 5). because of the pressure put on all subjects, as well as the fact that the game was designed to be somewhat tedious.

#### ***Low Scores (and Variance) for VENT Conditions***

Another result that was somewhat surprising was the low mean scores for times subjects played Game 2 recorded for both DELAY and NO DELAY conditions in the VENT condition. An apparently related discrepancy is that contrary to the hypothesis, the variance for game play times in the VENT conditions was not as high as was expected—not as high, for instance, as it was in the CONTROL conditions. Because there is conflict about the effects of emotional venting (as described in Chapter 4), it was predicted that variance for the VENT conditions would be high, particularly in the DELAY condition—exposing the unreliability of venting as an emotional regulation strategy. If anything, this author expected that the ability to vent would, on average, increase subjects’ ability to recover from frustration arousal, yet the behavioral results were extremely low and uniform, and the SD for the delay condition was not exceptionally high.

Why, then, were these results so uniformly low? The floor effect, described above, may help to explain the VENT condition’s consistency in its poor showing. However, another explanation, strengthened by the other results obtained in this study, is that venting is simply a bad strategy: it isn’t just unreliable; although not significant, on average it yielded negative results, slightly worse than doing nothing at all to support the user’s emotion regulation. One reason why this may be is that

something in the protocol's implementation of this condition made subjects uniformly less patient, more frustrated, and less inclined to play very long. The self-report data yielded no significant opinion whatsoever to this effect, however.

Another reason for this behavior is one that some emotion researchers have already suggested: pure venting may help subjects recall the situation, and how frustrated they became with it, without doing anything further. Instead of helping to ameliorate the user's affective state, venting may actually be helping users to become more frustrated.

Although on the whole the difference between offering support for venting and no help at all (the CONTROL condition) was not statistically meaningful, on average doing nothing seems to be a little better (although yielding higher variance in times played) than supporting venting alone—and as will be discussed later, both seem to be no match compared with the support that a CASPER approach offers.

#### ***Little Statistically Meaningful Self-Report Data***

Much self-report data was taken, especially in the Exit Questionnaire, to try to capture subjects' affective states at various points throughout the experiment. Yet, despite the strong behavioral results described above, very little of these data reflected the behavioral showing, or indeed anything of statistical consequence—especially beyond the comparatively strong measure assessing frustration level at the close of Game 1. For example, the level of satisfaction described at the close of Game 2, upon analysis by condition, yielded no meaningful results. This was typical of these self-report data. The fact that none of it was usable (apart from the highest-contrast question) supports the suspicion described earlier: Self-report tends to be unreliable, yielding answers that may be put forth for a variety of reasons.

One notable illustration is the self-report data for anger experienced, which were extremely low. These scores were low despite evidence of a close, established relationship between anger and frustration, as discussed in Ch 2 and seen in (Oatley 94). Why was anger so routinely reported as “absolutely not angry at all”, across all key points in the study, despite routine reports of highly varying frustration levels? Possible reasons for this phenomenon might be a lack of an available, identifiable target for subjects' anger, even though it would seem that many such candidate targets abound, as is discussed below. Another reason might simply be that, in the laboratory, frustration in the absence of anger might be as strong a negative affect as one can muster—or that something about the study encouraged subjects to feel aroused enough to report fairly high levels of frustration, yet low enough to not raise most subjects' ire. Another possibility is the social desirability factor—that subjects may not be wont to suggest that they “lost control” and got angry, which for many in this culture is considered a sign of weakness.

#### ***One Interpretation: The More Frustrated One gets, the Longer One Plays***

Did subjects become *more* frustrated by the CASPER agent than either of the two interactive questionnaires, and did they therefore use Game 2 as a passive support, to play longer in order to try to relieve their own frustration states? Indeed, use of computer games when upset is a common strategy for passive support of emotion

regulation, as discussed in Chapter 2. However, this hypothesis is unlikely, for several reasons.

1: No evidence was found to support this hypothesis, from the self-report data. As described in Chapter 5, subjects were asked on the Exit Questionnaire to report their level of frustration, as well as other emotions, at key intervals during the study. These reports tended to follow the following pattern: Subjects generally reported feeling little or no frustration upon entering the room, then reported much higher frustration levels after Game 1, then slightly lower frustration levels (or the same) after the interactive questionnaire, then lower levels still (or the same levels) after Game 2.

On average, subjects in the DELAY/CASPER condition reported feeling either less frustrated by the interactive questionnaire than they did at the end of Game 1, or the same level of frustration for both. These reports were similar to those in all of the other conditions, with no statistically significant differences found between conditions.

Two (of 12) subjects in the DELAY/CASPER condition did report feeling more frustrated after the questionnaire than after Game 1, but the other 10 followed some variant of the pattern described above. Notably, the two subjects who deviated from the typical pattern each subsequently played Game 2 for the maximum allowable time (17 minutes). There are a number of possible reasons why these subjects may have played longer, the most obvious being the problems known about self-report data, as discussed above. However, another possible reason might be that these subjects represent a type of personality that was untested—for in this study—a type that is resistant, for one reason or another, to either the HCI implementation of the social cues, the cues themselves, or some other factor.

2: Although the behavioral results obtained might, at first, look like they supported this hypothesis, a number of other factors make the conclusion that more frustration = longer playing time unlikely. First, the rebound effect, described above, explains these results satisfactorily. Second, no correlation was found between self-report of frustration level after answering the interactive questionnaire, and time spent playing Game 2. Indeed, no statistical analyses on any of the self-report measures taken in the Exit Questionnaire revealed any statistically significant results, except for the aforementioned manipulation check for differences in frustration levels after DELAY vs. NO DELAY conditions in Game 1. Third, subjects were asked at the end of the Exit Questionnaire whether the interactive questionnaire they answered between games made subjects feel better or worse. Although there was wide, uncorrelated variation between 0 (no difference) and +5 (much better), no subjects reported feeling worse from interacting with it.

3: The protocol was designed to discourage subjects from playing Game 2 for a long time in general, and especially in all DELAY conditions. After all, the game was not exciting to play. Further, it had some bugs in its interaction that all subjects experienced, that made it somewhat tedious to play. Perhaps most important, though, was the fact that for those in the DELAY conditions, it was *the computer/network system itself* that was the cause of the subject's frustration levels after Game 1, as well as after the interactive questionnaire. The argument that subjects would perform a task on a system that frustrated them, made them feel less patient and more agitated, *for longer than* those who were less agitated by the

system, is a much harder argument to defend than that of the central argument in this hypothesis: That subjects played longer because they felt better, not worse.

#### ***Other Alternate Interpretations of These Results***

As discussed above, the time and score data for Game 2 were found to be highly significant, especially for the CASPER/DELAY condition, in contrast to all the rest of the conditions. This author argues that the only acceptable explanation for these differences was that subjects responded differently to the different interactive questionnaires. To verify this hypothesis, however, let's consider other, alternative explanations for this behavior, one by one. First, the obvious questions:

- Was it simply chance that yielded these results? No. Statistical analysis demonstrated the extreme unlikelihood of the null hypothesis, at  $p < .05$ . These samples are large enough that they may be confidently described as representative of the population.
- Was Game 2 better or more interesting in the CASPER/DELAY condition, compared to the others? No. The game was the same across all conditions.
- Did differences in the amount of time each interactive response type took to work through make a difference in how long they played the game? Unlikely. Times to completion of each response type were carefully estimated and measured in pre-tests, although large variance in the VENT and CASPER conditions were possible due to the open-ended, text-entry question at the end of both interactive questionnaires. The time, in minutes, between the end of Game 1 and the start of Game 2 were recorded, however, and analyzed. Note that this time interval includes time spent on the questionnaire as well as time spent reading directions for Game 2. The analysis yielded the following results:

CONTROL:  $\underline{M}$  = 4.23 minutes; variance = 2.36 and  $\underline{SD}$  = 1.53

*CONTROL/NO DELAY:  $\underline{M}$  = 4.38 minutes*

*CONTROL/DELAY:  $\underline{M}$  = 4.08 minutes*

VENT:  $\underline{M}$  = 5.43 minutes; variance = 1.76 and  $\underline{SD}$  = 1.33

*VENT/NO DELAY:  $\underline{M}$  = 4.78 minutes*

*VENT/DELAY:  $\underline{M}$  = 6.08 minutes*

CASPER:  $\underline{M}$  = 4.87 minutes; variance = 2.79 and  $\underline{SD}$  = 1.50

*CASPER/NO DELAY:  $\underline{M}$  = 5.11 minutes*

*CASPER/DELAY:  $\underline{M}$  = 4.64 minutes*

Note the differences in times that subjects spent answering the questionnaires between DELAY and NO DELAY conditions, and between the VENT and CASPER conditions. These differences merit a discussion of their own (for instance, it seems as though subjects in the DELAY/VENT condition took a much greater amount of time to express themselves than any other group; such a finding seems believable if these subjects were prompted to recall the frustrating experience, and then were given no substantive support in this process). Nevertheless, such a discussion remains for future work. Mean differences in response/reading times in this interval are not great; they amount to no more than 2 minutes between maximum and minimum. Further, these



times do not correlate with the differences in play times across conditions in Game 2.

- Did the administrator influence subjects into playing longer? Unlikely. The experiment was not wholly double-blind, but it did not involve human contact of any kind from before the directions for Game 1 were read, until after Game 2 was completed. Before that, the rigid format of all email and phone transmissions dictated what subjects were exposed to beforehand, and the script protocol probably pre-empted much or all of this influence, if present at all. The only other factor of administrator influence was the selection of which subject was run under which condition, and at what time. But since all subjects were strangers to the administrator, and most of the factors in these decisions were determined by which condition was working bug-free at any given time, who was willing and able to be scheduled for which times, as well as trying to recruit enough women to keep their numbers consistent with those of the men. On top of all these considerations was an effort to randomize the remaining room for maneuver by time of day, sex and condition experienced.
- Was there something unexpected and unaccounted for in the CASPER agent that entices subjects to play longer, for reasons other than those stated in the hypotheses at the end of Chapter 4? Probably not, since the second game is not mentioned at all in the CASPER agent interaction, nor is there any allusion to playing the game again during this interaction. To confirm that the agent had no such effect, the NO-DELAY condition demonstrated significantly lower scores and playing times for the CASPER condition.
- Did age or game play experience influence players in the CASPER/DELAY condition? No. Subjects' ages and game playing experience were shown via statistical analysis to have no significant relationship with condition, time played, or score achieved.
- Did subjects' sex influence players? No, as shown via statistical analysis in Chapter 6, and despite suspicions to the contrary.
- Did subjects' trait arousability affect how long they played, or how well they scored? No, as was also demonstrated via statistical analysis, and despite predictions to the contrary. In fact, trait arousability was not shown to effect performance on any vector, despite predictions to the contrary.
- Did the fact that the CASPER condition allowed the subject to express herself about both the problem and her emotional state account for differences in performance? No, since the VENT condition performed these functions too.
- Did the fact that the CASPER condition was more interactive and/or conversational in nature account for the differences in performance? Unlikely. The questionnaires in all three conditions were designed to be interactive in nature (all responses were through radio buttons and generic buttons, with precious few text fields at the end). Also, the phrasing of all questions in all conditions were displayed with a "chatty", conversational tenor—questions that started out were phrased, "First, can you describe...", while questions farther along the sequence might begin with "Okay, and how often did you..." Some questions began with cues analogous to what linguists call back-channel

feedback” (Tannen 94), such as “Hmmm” and “Okay, now what would you...” Beyond this effort to make the other questions more conversational-seeming (while canned), it is worth noting that one of the hypothesized primary benefits of the CASPER agent is to encourage the user to feel “heard” and “empathized with”, for which more dynamic interaction with the user is a must.

- Did how well subjects performed in Game 1 affect how long (and/or how well) they played in Game 2? Yes, a correlation was found between score of Game 1 and the time (and score) of Game 2. However, this was a consistent, main effect across all conditions, with no special, significant effect for one condition or another.

The findings strongly indicate that subjects played as long as they did for the reasons hypothesized above: They felt much less frustration arousal while playing Game 2 than subjects in any other condition, and consequently felt more patient, less bored, and more satisfied in their experience playing Game 2. They were also better able to concentrate on the task at hand, and scored more points accordingly.

The most convincing reason that subjects in the CASPER/DELAY condition played longer and scored higher was because of what they experienced in the CASPER condition: In place of a few more self-report queries for other relevant emotions, most subjects were led to read what amounted to two dialog screens, on which were displayed a total of four simple sentences.

#### ***Possible Targets***

An argument may be made that subjects in the DELAY/CASPER condition must have also felt better and more forgiving toward the entity that had recently served as the object of their frustration. In order to play so much longer, not only must frustration levels have been lower, but subjects probably felt more kindly toward what they were doing. But what, specifically, may have been the object of their shifting feelings? The game? The character? The computer? The system’s designers? The Web?

Subjects may have been led to believe that the designers of the game were to blame, since it was they who were ultimately responsible for such a faulty game experience. On the other hand, clearly, subjects may have felt that the Web was the culprit here, since the cover story based the game on the Web, and since the delays did, clearly, seem to originate from the Web. The Web is indeed a named entity, although by nature an abstract one. And users of computer systems, as most or all subjects in this study seemed to be, in general tend to be familiar with the common notion of Web delays, and the frustration they can engender.

Steps were taken in the design of the protocol to foster the attribution of most of the problems subjects encountered to the Web. Indeed, as illustrated in Figure 5.9, the Web browser window was physically located all the way over to one corner of the screen, only overlapping the “local” Instructions window by a small margin (and neither window could be resized or moved). The local materials did not cause frustration, but merely gave instructions, asked questions and, in some cases, gave feedback. Could the object of subjects’ negative and positive affect be more than one perceived entity, as the protocol tried to foster? Since all computer activities

were done on a single computer (with a single monitor), recent research would tend to suggest otherwise.

The Media Equation work (Reeves 96; described at the end of Chapter 3) makes a strong case for the idea that humans, on many levels, tend to relate to computers as if they were either real people or real places. One argument that the Media Equation work would seem to support is single affective attribution toward the computer itself. The monitor and keyboard were what subjects interacted with, after all, and these entities are offering all activities at hand. If people tend to respond to the same vs. different computers with different politeness rules, the suggestion is that people will tend to treat the computer they're working on as a single, personified entity.

Work remains to be done, however, to discover the extent of this physical entity/personifying phenomenon. It's also possible that there might indeed be multiple targets for affect: People might map the Web to a location, and think badly or better of that place—or for that matter to a “Web” that is envisioned as a personified, unreliable, human-like entity—all the while treating the game system as a separate entity.

But in Game 2, which was identical in design across all DELAY conditions, subjects never again encountered such apparently “Web-based” delays. They also never interacted with a human that represented the game's designers (i.e. the administrator) until they were finished playing Game 2. What they did do was engage in one task: they played this simple game. Arguably, it was the game itself that became frustrating in Game 1, and for those in the DELAY condition, the game stopped being quite so frustrating in Game 2.

It would seem that regardless of the target or targets for the affect, subjects in the DELAY/CASPER condition played the game for much longer, and therefore probably reconciled their negative affect, and at least in partial exchange for positive affect. They felt better about the experience in general and, most probably, about any and all targets involved.

Despite whether subjects felt good or ill affect, the fact remains that there were many, many reasons for subjects to stop playing this game after the first three mandatory minutes. Indeed, very little was keeping subjects in the room, and playing. This game offered little incentive to keep playing. By many accounts (especially from those in the NO-DELAY conditions), this game was “definitely boring”, “dull”, “boring after a while” and “the game seemed pointless”, and another was “disappointed at little action”. Others found the “slow movement” of the character “frustrating” in and of itself.

But something clearly happened to subjects in the DELAY/CASPER condition that made them play for much longer and score more points than subjects in either of the other two DELAY conditions. The most logical reason why they did was because they were, in fact, using not just the computer in general but *specifically* the CASPER agent in a uniquely social role, as an active support mechanism to modulate their emotional state. Subjects in the DELAY conditions performed according to a pattern that was very much in line with expectations that were made assuming the use and effectiveness of the CASPER agent, and its supporting theory.

These findings point to one conclusion: The CASPER agent demonstrated significant support in helping users relieve their feelings of frustration.

## CHAPTER 8

# Theoretical and Practical Implications

This author built the first prototype of an interactive software agent that has been demonstrated to provide CASPER-style interaction, and provide effective relief of strong, negative emotional states related to frustration, as well as relief of negative feelings about the *object* of the negative emotion. This agent is of very simple construction, and uses no technology more advanced than elements found in any standard graphical user interface (GUI)—elements such as dialog boxes, radio buttons, and text (see Figure 5.8 for a diagram of the agent's interaction).

Interaction with the CASPER agent occurs solely through traditional means: monitor, keyboard, and mouse.

And, as stated in Chapter 7, interaction with this agent amounted to this: answering a few questions about one's experience and emotional reactions to it, followed by reading a brief paraphrasing statement, approving an error check, and then reading two brief sentences expressing first some acknowledgment and empathy for the problems experienced, and then sympathy for the subject's negative emotional reaction. The entire transaction amounted to roughly reading 6 or 8 short lines of 12-point text, pressing a few buttons, and typing a blurb of text.

## Theoretical Implications

There are many, many practical and theoretical implications of this work. Some of these are discussed below.

- This project demonstrates that effective addressment of strong, negative emotional states is something computers can do. This finding has, to this author's knowledge, never been demonstrated before.
- The results from the VENT condition in this study suggest that, when facilitating the regulation of negative emotions in others, it's not enough to let people vent. Some acknowledgment must be made of the user's negative experience, and resulting affective state. This result would seem to apply to human-computer interaction as it does to human-human interaction, and may even inform the debate about the utility of venting in the ameliorative process.
- The CASPER agent described in this study is clearly a social agent, with demonstrated social interaction capability. Yet the approach used here represents a departure from traditionally-envisioned social agency. Apple's 1980 "Knowledge Navigator" video described a fully humanoid agent to interact with, including natural language speech input and output, and a photoreal, animated image on screen. Microsoft's Bob and subsequent Wizards and Assistants (such as the "Paper Clip" Assistant) all featured brightly animated characters with which to interact. As anecdotal evidence mounts that users often feel strong, negative emotions toward the paper clip agent, the CASPER agent demonstrated in this study offers no such cues to its agency. There is no embodying character, no image, no introductions, no cheerful greetings, no name, no specific "personality" programmed into it to seem more lifelike. Indeed, as a social agent, it appears to be an extremely crude one. The implications of this study include support for the argument that social agents need not be personified characters, or use advanced interaction techniques such as speech I/O in order to be effective.
- The Media Equation (Reeves 96) presents a view of human-computer interaction that is modulated in many ways by social forms. According to that work, the way we interact with computers tends to be "social and natural". The work reporting significant findings with important implications for HCI in over 30 different categories of social and/or spatial phenomena. All such phenomena seemed to inform user interaction with computer systems, yet it is unclear to what extent the Media Equation can be demonstrated. The work reported in this thesis represents a dramatic extension of the Media Equation: Beyond indications such as that flattery, criticism and other kinds of social cues can affect human users and their perceptions about machines that "express" them, this work demonstrates that strong, negative emotional states can be actively modulated through the use of a computer system. This activity, hitherto assumed to be uniquely social, suggests a role that may be closer to confidante than a sterile service relationship.

#### ***Implications of Principles for Designers***

From the standpoint of the human-computer interaction designer, there are also many implications. The following are few such implications:

- The results from this study represent at least several significant contributions to the design of computational systems—in particular, how designers think about how to address failures in the design of software. Among other things, these findings enable designers to address possible failure in their systems, by:
  - Considering the user’s emotional state as a factor in the design process
  - Appreciating that offering no support for the user’s emotional state during or after a system failure is tantamount to ignoring a critical aspect of the human-computer interaction, and may well result in many side-effects that are injurious to the product and its maker.
  - Being able to begin to address the full scope of the effects of computer system failure for their users
  - Gaining knowledge of an easy-to-build tool to help make meaningful differences in the product’s subsequent use, as well as its (and the manufacturer’s) perception in the eyes of users
- It was functionally two dialog boxes, each with a bit of text and one or two buttons in them, that made the difference between much higher demonstrated satisfaction rates, and what may have constituted a comparative annoyance. This study also suggests that text and buttons may be an underutilized and overlooked method for creating agency.
- Following the points raised above (and considering the all-too-common responses to the Paper Clip Wizard), the approach to agency used in this thesis may be a much-lower-risk, much-less-offensive way of creating agency.

### Practical Implications

The CASPER agent was shown to be effective not only at relieving strong, negative affect, but in making subjects feel better about using the same computer system (and performing very much the same activity) that had frustrated users in the first place. These findings have profound practical implications for virtually every kind of software system imaginable.

- The method demonstrated here represents an extremely cheap solution to addressing failure with an enormous amount of power; i.e., a big “bang for the buck”.
- Such devices can make a significant (comparative) contribution to productivity—at home, in industry and commerce—especially when products fail, but even when there is no such failure on the part of the system, but other events in the lives of the user conspire to frustrate.
- The CASPER method demonstrates that users who use a product that fails for some reason should actually wind up liking the product *better* with a CASPER agent onboard than without it.
- The very idea of a positive outcome tied to the acknowledgment of a corporation’s or product’s failure may actually have important lessons for industry. The current assumption in corporate America, especially in light of

the litigiousness of this society, seems to be that admitting failure, offering apologies or taking any other kind of responsibility for poor performance is to be avoided at all costs—at least until a strong legal case can be mounted. This work demonstrates that there may be great benefit to customer satisfaction if such things as timely apologies are made, even by machines.

- Implications abound for practical applications for this work in helping to automate customer service and complaint departments. The idea of talking to a machine in order to complain to a company about a product or service may sound hard to buy, yet there are a number of benefits to the use of such systems that many members of society might actually prefer—particularly people who are shy. The benefit of being able to complain about a product, feel heard and understood, yet not have to risk confrontation with a service representative who may be confrontational, and may not be trained (or have the patience) to deal with frustrated customers, may actually be attractive to those who fear confrontation. Such systems would have the benefit of actually increasing the likelihood that more people would go to the website or call the hotline, and report their problems. Such an automated service would ideally be offered in tandem with a real human service department, and not in place of one.

To better appreciate the effect of such an agent, an example of one in action follows, below.

#### ***An Illustrative Scenario From Real-Life Corporate America***

What might a CASPER agent look like in practical application? Here is one scenario, adapted from real life.

Microsoft is an excellent example of a company with a notorious reputation, even among its loyal customers, for prematurely releasing products that have severe problems associated with them—then fixing the eliciting problem, but leaving their users feeling distinctly dissatisfied and angry.

Microsoft's Windows NT 4.0 operating system (OS), for example, was released in 1995, and subsequently found to have critical errors in its design. At least three subsequent "service pack releases" fixed many of the bugs in the OS software, but did nothing to address users' frustration with the problem, or, for that matter, acknowledge the problem at all, apart from offering the solution to it. Because of this inability to adequately address the consumer's experience in this process, consumer confidence in both the NT OS product line, as well as the company in general, was greatly undermined, and such experiences are not soon forgotten. This author has overheard more than one conversation about this very product, characterized by incredulity, sarcasm and derision—conversations that occurred many months after the "service pack release" was intended to fix all problems.

Imagine, if you will, a different scenario: Assume the same grievous error in prematurely releasing the software, which makes customers and users feel frustrated, manipulated and powerless. However, this time imagine that the service pack release includes a CASPER-like agent that, like the version described herein, is simply text-based. Such an agent would be so computationally inexpensive that it would be capable of running effectively during the software installation process.



The agent would engage users in a dialog about the product and its subsequent service pack release, and query users about their feelings about the experience. In the end, the system would offer a statement offering a heartfelt apology appropriate to the level and type of affect expressed by the user, as well as some empathy and sympathy for the user's predicament, stated inconvenience, and loss of faith in the product. Then, this agent would encourage users to describe in their own words how they felt, and anything else they would like to add.

Imagine that, once the software was finished installing, that a dialog box appeared politely asking for the user's permission to automatically send the user's feedback via email to Microsoft headquarters in Redmond, WA, with specifics about who would see the information, how it would be used, and assurances that it was indeed valuable, confidential information that will contribute to better products in the future. The company then can keep track of its user base, maintain some idea of how its products are perceived in the marketplace (extremely valuable information in industry), as well as possible leads for designers on new product innovations or bug fixes that should be made.

The user, on the other hand, is left feeling *much* better—about the product specifically, the interaction in general, and about work overall. Indeed, the user may actually feel downright *warmly* toward a product and a company that would care so much to consider the user's feelings, and make the user feel so heard, understood, and accepted. Perhaps best of all, the information sent to the company (depending of course on the company's ability to make sense of and use this information to make better products) helps maintain a sense of *authenticity* to the interaction.

This sense would help users accept this entire process—especially if the user is subsequently sent a brief, email note from a real person at Microsoft, thanking the user for the feedback, and giving some assurances that the information was heard (i.e. providing more paraphrasing feedback) and will represent a change in the way that the company will do business in the future. Such a sense of authenticity would also greatly elevate the affect that the user feels toward the company, its products and services. Plus, this element of authenticity would help make subsequent blunders on the part of the company much more acceptable, and forgivable in the future.

Of course, this scenario might be just as effective—if not more so—if the company were smaller, newer, and had oriented itself as a customer-centered design shop, complete with built-in mechanisms for accepting, dealing with, and responding to such feedback...

### Further-Reaching Design Implications

This work has implications for user-centered design, in particular the design of systems that are able to interact with their users as social and affective beings, not just as information processors. Beyond the obvious—that this study may contribute to the ability of systems to help users recover quickly and efficiently from strong, negative emotional states such as frustration—there are still other implications as well. This work may also lead to side-effects from its social-interaction nature, both positive and negative. Positive side-effects of such systems may include a sense of rapport with computers that can engage users in this way, perhaps fostering an

increased sense of cooperation and good feeling in the user, leading to increased productivity and job satisfaction (rapport and its beneficial side-effects were suggested by studies on teamwork in Reeves 96).

Negative side effects may include emotional and/or cognitive objections to perceived false (or imitation) caring on the part of the machine that uses empathy as an emotional management support strategy. Good interface design, however, includes communicating the system's capabilities and "intentions" clearly to the user (e.g. Preece, 94), and this system is no exception. Proper design of the CASPER agent should go far toward mitigating these negative responses, and no such objections were found in the self-report data collected in this study. However, such objections (as well as other, unforeseen complications) may be found in subsequent tests of the CASPER approach, in particular when subjects are informed of the agent's existence, and goals, up front.

There are many more practical and design implications for this work than have been described here. Indeed, those indicated here are just "the tip of the iceberg." Hopefully, however, the above partial list will help articulate the potential benefits of such an approach to HCI.



## CHAPTER 9

# Sociological and Philosophical Implications

The CASPER agent represents the first of a new genre of human-computer interaction. Such devices (and their approach to problem-solving) therefore carry with them a host of new implications, positive and negative alike. These implications are broad in their scope, and range from the level of the individual to that of the culture. It is therefore imperative, scientifically as well as ethically, that this impact be explored as fully as possible—before such devices can responsibly and ethically be put into widespread use.

At the level of the personal, issues include: how humans may use (or abuse) such devices themselves; how might such devices change the nature of human-computer (and human-human) interactions, and how will humans define themselves in a world where such devices are regularly used?

On a commercial level, issues include the ethical use of such devices, and the incentive corporations may have to develop such products—as well as high-quality products in general, when incentive to release fine products is diminished by the ready availability of an effective safety net.

Political issues run the gamut from public consensus on acceptable design and use of such devices, as well as the potential misuse and/or abuse of them, including the use of these devices to help maintain disciplined citizens and consumers. At the level of culture, might the advent of such devices be used to foster positive change on a society-wide basis, or might they be used as another means for manipulation and control, fostering the dismantling of a society that once held dear values of individuality, autonomy, and authenticity? And finally, on the global stage, how

might widespread use of such devices help to enfranchise humans around the globe, and how might they be used in the steamrolling, Hellenizing effect that Western culture seems to have on other, diverse cultures around the world?

This device is, for all intents and purposes, a way of leveraging human nature by “tweaking heart strings”. Is such a “tweaking” for the best, or for the worst? A discussion of the relative benefits, and risks, ensues below.

Soothing, Salving, Modeling: For the Best?

### **Main Effects**

As discussed in Chapter 2, frustration in humans has many unpleasant side-effects: Increased ability to become more negatively aroused, increased likelihood of getting angry, decreased ability to pay attention, think (and problem-solve) creatively, and interact harmoniously with others. Notably, though, frustration is often regarded, along with the other emotions judged as negative, to be painful or, at the very least, removing pleasure from one’s experience. The agency demonstrated in this project had the effect such that people who interacted with it during times of emotional duress showed signs of improvement in their emotional state, as manifest by their willingness to engage in significantly longer interactions with the system after interacting with the agent. This kind of effect has direct implications, then, for actively addressing all the aforementioned debilitating effects of frustration. Such an effect demonstrates the possibility of helping humans to better manage difficult events, thereby rendering day-to-day existence less stressful and, perhaps, more pleasurable. Clearly such longer-term effects have yet to be shown, yet the software built for this thesis holds great promise for them.

Similar effects may be shown for:

- Improving harmony in interactions with workmates and other “cooperative parties” (other humans, as well as the computer systems with which the user interacts);
- Increasing one’s ability to think creatively and generate solutions to one’s own problems with greater ease (and improving, in the process, one’s sense of autonomy and control);
- Decreasing the likelihood that subsequent stimuli, particularly those that may be mildly annoying, will be perceived as frustrating, and to which an even stronger negative reaction will be mounted. This effect should, over the long term, act to preserve—or even possibly improve—one’s sense of self-control.

### **Modeling and Meeting Human Needs**

A number of other possible side benefits suggest themselves by work of this nature, some of which are discussed below.

- Assuming that such an interaction constitutes a perception of a social exchange (for which there is some evidence in the literature), such interactions may have the effect of enabling humans to help meet their social needs—feeling a sense of social connection to others and combating loneliness—despite apparent large-

scale cultural trends for the reverse: Increasing isolation (Myers 93), Internet-based malaise (Edupage 98), as well as increasing scarcity of human resources and human attention in a culture of divorce, two-income families, telecommuting, etc.

- Evidence from a number of quarters (Myers 89, Goleman 95), suggests that, in this culture at least, precious few humans seem to possess solid, effective, non-judgmental active-listening skills. Indeed, positive, constructive communication skills are very much appreciated when encountered—well-developed skills of which are something of a rarity. One possible reason for this lack of such skills in the larger culture may be a kind of positive feedback loop: Since so few people are available to practice such skills in day-to-day life, there are thus few people to serve as role-models for this behavior. The CASPER agent, however, is an easily-reproducible device that uses an effective variant of active-listening behaviors. As human users interact with such agents, they may experience the above-mentioned benefits, and note the agent’s techniques. Thus, this human-computer interaction can therefore act as a positive model for subsequent human-human interactions.
- It may be argued (as it is below) that the kind of interaction of which the CASPER agent is capable may constitute a critical lack of authenticity, of believability and, therefore, credibility. This could be true for a variety of reasons, among which is that the machine is simulating human-like behaviors without truly *knowing* what the problem is, experiencing empathy, sympathy, or really understanding the emotions that the person is experiencing. However, there is a kind of interaction with which humans partake every day, in which there a number of striking similarities with the CASPER approach, and in which positive, ameliorative effects on the part of the human are measurable and significant: Those that occur with pets, especially dogs and cats. Humans have long benefited from these interactions in many ways, including emotionally, even though it is clear that these animals may not understand the emotions of the human. Further, humans routinely use vague cues from these non-human sources that the humans anthropomorphize to meet emotional needs such as feeling empathized with, feeling accepted and understood—all despite the probability that the animal may very different motivations and goals, and that there is probably no real empathy or understanding (as we conceive it) present.

Heartstrings and Pursestrings: For the Worst?

#### ***Emotion Management Assistance vs. Emotion Manipulation***

In the proposal for this thesis, it was suggested that providing active support for a person to regulate her own emotions may be seen as a tame version of far more nefarious concepts, such as involuntary emotion manipulation, brainwashing and mind control. While these concepts and their relationship to this research is a frank concern, it was argued that there seems to be a major difference: The stated goal of the system is to support the *user’s own emotion management*. It should, ideally, be

incumbent upon the user to use the system in this manner and in this role, and to receive any benefit at all from it.

However, such a deliberate intention may not always be the case—as was demonstrated in the design of this study. Whether or not the process at the heart of this interaction is voluntary is of critical importance to this work, since if it can be shown that a software agent can perform its work beyond the suspicions of the user, this indeed demonstrates a means for a kind of mind control. Troublingly, the experiment around which this thesis is based involves just such a deception, and one that was apparently not uncovered by most, if not all, subjects.

It should be noted that, as a first-of-a-genre device, that its capabilities might well subsequently become better known. Further, humans tend to develop ways of resisting efforts that are perceived as manipulative. Still, this thesis demonstrated strong behavioral effects that appear, from the self-report data, to have been almost undetected cognitively. Subjects appear to have felt much better, but were not aware of any specific manipulation. Therefore, it is possible that this approach to emotion regulation could be used in an involuntary manner—and examples in which such uses are unscrupulous are not hard to imagine.

#### ***Trusting the Image in the Mirror...***

The CASPER approach is, at its heart, an idealized simulation of real empathy, real understanding, and real caring. It's effectiveness may lie in leveraging the very fact that when humans hear these words, they sound so honest and true that, in their rarity of utterance, perhaps intuitively they sound real and “heartfelt.” Yet in truth, there *is* no heart to “feel”. It may be, in essence, a simulated transaction that, in its simplicity and boiled-down absence of fallible, ineffectual statements, is seductive—perhaps too much so to resist. This notion by itself is troubling. Since as is the case with simulation by definition (Starr 94), the extent of the model tends to be unclear, and can lead to over-attribution of capabilities, as well as just plain wrong attributions.

#### ***... and Trusting the Reality the Mirror Reflects***

Another possible pitfall of the fact that this work centers around a simulation is that it may render the authentic, human-human interactions that it simulates less effective. It is suggested above that one possible positive benefit might be the routinization (albeit automated) of a currently rarely-experienced, very beneficial conversation, and one which might help model positive, effective human-human interaction for its users. But perhaps, in automating an artificially-derived human response, the net effect may be to wreck the foundation of efficacy for both the real and the simulated conversation. As Baudrillard wrote of the cave paintings of Lascaux, “the duplication is sufficient to render both artificial” (Baudrillard 83). Whether humans become so jaded in interaction with CASPER-style software that they become less receptive to real human empathy is an open question, but one of great concern nonetheless.

#### ***Computers as Friends: The Next Generation?***

With the advent of the computational pet, a legion of children are already growing up feeling a deep connection with their toys. Now, children have perennially grown up feeling emotionally attached to their play objects. Indeed, children have seemingly always used blankets, dolls and other toys to fulfill needs of one kind or another throughout their development. For example, toddlers bond with “transitional objects” to help them conceptually segue from primal attachment with their primary caregiver to an independent identity (Winnicott 71). Young children subsequently use baby dolls to model and simulate relationships they perceive in the real world (e.g. Fischer 97).

What difference will the new generation of increasingly “intelligent” interactive toys make to the generation of children that grows up with them? A question remains as to whether the Interactive Barney’s, and the armies of Tamagocchi’s, Furbies, and Interactive “Winnie the Pooh” dolls may be socializing a generation of children to using computers as confidantes, and to having emotional relationships with such artifacts. If CASPER-like interaction is not only feasible but easy to create, as the present study suggests, what implications does this have for such toys, and their relationships with their owners?

One credible possibility is an entire generation of toys that are capable of this kind of emotional-content interaction with their young users—capable of soothing a crying child or of perhaps artificially preventing strong feelings of loneliness, sadness, frustration, and a host of other strong, negative emotions. Might such artifacts discourage their owners from fostering normal, healthy interactions with their parents and other children? If support for emotion regulation is provided too-early-on, would this have a beneficial, educational effect, or leave children emotionally crippled, thwarting the development of the skills needed to interact successfully with other humans?

This question may be developed in the adult world as well: Specifically, if such devices achieve popular success, and humans routinely use them to help manage their emotional states, what happens to the human’s sense of his or her own self-control? Moreover, where does the locus of control (or its perception) reside in a world in which humans are dependent on emotional cyborg relationships for their emotional well-being? Assuming such interactions are purely voluntary, it may be hard to conceive of such locus of control moving toward the external, toward the machine. Furthermore, the CASPER agent was modeled after human-human interactions precisely in order to emulate the range of unpredictable effects that humans might have to it. Still, the transplant of behavior from human to machine is dramatic, and the resulting fundamental differences in the subsequent interaction cannot be ignored. When machines model human behaviors for interacting with other humans, what happens to the perception of self, of one’s humanity? To the sense of autonomy?

### ***Sapping Needed Wind From Sails***

A problem that may be of great concern in this work is the possibility of diffusing emotions that need not and should not have been diffused. Since an important function of emotion is motivation, building frustration can often be seen as building fuel to overcome obstacles. Therefore, diffusing frustration and anger prematurely



may act as a kind of “computational Prozac”, undermining a person's ability (if not their right) to perform a potentially unpleasant task, such as confronting the company that sold her a poorly-designed system and demanding her money back.

In this situation, a system that tries to help diffuse strong user emotion performs a service for the manufacturer (lowering the number of complaints mandating better products), at the expense of the user. Indeed, it is not hard to imagine a system that convinces the user that it genuinely has the user’s best interests “at heart”, when the system is obviously undermining the user’s goals by prematurely assuaging his emotional state.

Such effect is not necessarily limited to the domain of commerce, either. It is not hard to imagine politicians, if not whole governments, in a variety of political denominations, falling prey to the seduction of employing masses of such devices with which to manipulate a disgruntled constituency. Indeed, what better way for the current sitting president to try to reconcile his ailing image in the minds of angry voters, than to use an inexpensive tool that makes it seem as though he himself were performing personal acts of contrition, when he was not?

Yet it is also possible that this strategy would quickly backfire. Let us take the corporate case as an example: Users, as experts in social relationships, would soon see such a strategy as disingenuous; humans seem to have need of some semblance of authenticity in their interactions. If all a person ever receives in transaction with another is assuaging (without the eliciting problem ever being addressed in some reasonable timeframe, for instance), the person would become offended at this abuse of apparent trust. Similarly, a manufacturer that offers nothing but poor product after poor product, albeit with wonderful apologies after each, would eventually develop a reputation for trying to “pull a fast one” on users. Perhaps a bit slower than the company that offers no such consideration for the user’s feelings, however automated.

The onus, then, would still remain on the manufacturer to maintain its integrity and public image. Still, it would seem that more and more effective CASPER-like devices might draw out this period before the epiphany of disillusion toward infinity, to the detriment of the consumer.

### ***Absent Empathy***

The appearance of empathy is considered to be of primary importance to the effectiveness of active listening. And, as the study describes herein suggests, through careful design of the agent, the impression of empathy can be communicated from computer to human, with resulting benefits. But to what effect? What are the effects on users, and on the culture as a whole, when devices are in routine use that profess to have empathy for users, when no such state exists in the machine? This is a troubling question, but one that is not without some precedent in social interaction. Humans, after all, sometimes find themselves in supportive situations in which believability and credibility are seen as highly-valued, yet the substance behind the interaction may in fact be lacking.

It is not hard to imagine, for example, a situation in which a psychotherapist or social worker is conducting a session with a client who is upset, and is crying or angry or

otherwise demonstrative of his feeling. The situation may be seen as relatively typical, in which the therapist is trying to do the hard work of listening to the client, while distracted by some thought or concern about the client's behavior, progress, or even some notion unrelated to the client. In such a situation, the therapist may not actually be listening very much or very well to the client, yet with well-trained skills in providing the right kinds of feedback (such as active listening paraphrasing), is providing the client with cues to the effect that the therapist *is* listening.

The therapist knows that one important goal of this interaction is for the therapist and client to make some kind of therapeutic progress. However, while the client is very upset, the therapist may be trying to get the client to calm down. The tools at the therapists disposal are similar to the theory behind the CASPER agent: Enable the client to feel listened to, and that his emotional state is understood, acceptable, and accepted. As a result, *while it is being delivered*—while the therapist is distracted but offering solace anyway—such apparent empathy may not be genuinely felt. In this situation, the therapist may feel as though the session was successful if the immediate therapeutic effect was achieved—if the client came away from the interaction feeling heard, listened to, understood.

Certainly, the interaction with an automated system that is greatly limited in its capabilities guarantees a qualitatively different experience than one involving a real, trained, human listener. Further, implications for this kind of “absent empathy” on humans in the long term are suspect and unclear, but they are beyond the scope of this thesis to speculate in an informed manner. Yet some of the benefits may be the same as in apparently authentic social interactions, and even similar illusions in their presentation may be employed by even the best-intentioned human listeners.

### ***Synthetic Sympathy***

In many ways sympathy expressed by a machine may be more problematic than empathy. Whereas empathy assumes understanding, sympathy implies much more: feelings, memory, correspondence. These are skills that computers perhaps even more obviously do not possess. Implying by demonstrating a sympathetic response that these faculties really *do* exist raises further questions of authenticity, and effect.

### Situating This Work in Context

The focus of this thesis raises complicated philosophical questions. Further, the work makes it clear that strong AI is not needed to raise such questions, and that the age of such devices is upon us. What effect does the notion of CASPER agency have against the backdrop of larger questions of post-industrial society? We live in a culture in which litigation and litigiousness are routine; in which the notion of privacy is being eroded, often by the very affordances of the age of information and computer technology, as well as by acts of government, industry, and commerce. Conversely, there is also a trend toward institutions, both public and private, being less willing to disclose information. And, at the same time, there seems to be an increasing separation of ownership and responsibility in the culture at large.

Issues that arise in this context abound, some of which have been broached, and others which are important, but which are beyond the scope of this thesis to properly address:

- The computer's role as revelateur, as revealer of how we see society, and ourselves.
- The fact that humans seem to be "hard-wired" in a certain way, to respond socially to media. This makes humans vulnerable to, among other things, second-hand manipulation by humans employing such devices on others. As designers and moral citizens, we have a responsibility to address such phenomena, not ignore or simply explain them.
- The argument that machines can and must be designed to display their capabilities and limitations to the user. This idea has long been a tenet of HCI (e.g. Shneiderman 97), but the ethical implications of violating such principles abound. It has been suggested that the machine must be "put in its place" so humans are reminded of where the machine actually "stands". In other words, modern computers are good at several tasks: They can transmit information (the least problematic vision of computer, in which the machine is conceptualized as message passer); they are also excellent repositories of information, especially when the machine makes it clear what data it's storing, and what data it is using in the interaction with humans. Beyond these natural affordances of modern computers, their capabilities to simulate human behavior can quickly cross over into illusions, false appearances, and unsatisfiable expectations, among other things. Therefore, the design of any such systems must be done with extreme care—and sometimes, not done at all.
- The authenticity of the interaction is an extremely important issue, as was briefly discussed in the example at the end of Chapter 8. One of the things post-industrial society has learned how to do is to find effective ways to foster illusions, false perceptions, and romantic notions in humans, in order to keep them disciplined as citizens and as consumers. Does CASPER agency represent a contribution to the tools society has developed to hoodwink its citizens, or is it a meaningful, beneficial tool whose benefits outweigh its possible ill uses? Indeed, generating ways of envisioning this new era of HCI without reducing authenticity will be a challenge in the coming years.

***What This Approach is Not: Computer Psychotherapy***

Emphatically, this approach has little to do with computer psychotherapy (see Chapter 3, above). The stated goal of approaches to computer psychotherapy (outlined in Turkle 95) is the same stated goal of most psychotherapies: to somehow bring about healing from lasting pathological disorders via psychotherapeutic techniques. While the approach taken in this thesis is informed in part by techniques adapted from social psychology such as Active Listening (Ch. 5 of Corsini 95), this approach is not intended as a treatment to heal long-term psychopathological problems. Rather, it is only intended to make it easier for the user to modulate his or her own emotional state in the very short term, in an environment that is otherwise uncondusive to such support.



## CHAPTER 10

# Future Directions

***Determining Effective Components of the Present Study.***

An important extension (and next step) of the present study would be one to determine what works, what doesn't work, and what could work even better in the CASPER agent's design. Is the paraphrasing feedback necessary and sufficient to yield the same results? Are both empathy and sympathy required to maintain effectiveness? Might humor be more or less effective in frustrating situations, as was suggested by the MATLAB anecdote illustrated at the end of Chapter 3?

A fruitful approach to this research would be to decompose the CASPER condition into its constituent strategic elements (and/or subsets of the current condition), and to re-run the experiment in the CASPER condition, while manipulating the CASPER agent to contain different permutations and/or subsets of the current set of strategies, to employ other strategies, and to test different implementations of current strategies.

Other relevant research questions include:

- Does it help, hinder, or maintain the agent's effectiveness to make more explicit the computer's inability to feel empathy or other affective states?
- Does it affect the agent's effectiveness to make more explicit the idea that the computer itself is the social actor, and to try to remove all traces of (or for that matter, to try to boost) the connection between the agent (and the software of which it is apart) and its human designers? Does the idea that the information communicated by the user will reach human eyes contribute to, hinder, or muddy the effect achieved by the CASPER agent?

### ***Double-Blind Validation of this Experiment***

The study described in this thesis was executed as singly-blind—only the subjects were unaware of the manipulation, while the administrator had knowledge of which condition each subject would experience. Ideally, the experiment should have been double-blind, in which the administrator had no knowledge of which condition each subject experienced, so as to remove all possibility of experimenter bias being conveyed to subjects. Ideally, a system could be implemented in which a third party specified which condition was to be run on which subject, so as to keep the administrator in the dark on this matter. Alternatively, the software itself could have been designed with a procedure to calculate, in randomized fashion, the condition to be run on the next subject.

It should be noted that in the present experiment, a number of steps were taken to try to remove any experimenter bias: The materials sent to each subject before on-site participation were rigid in their format (see Appendix A), and the detailed script read to subjects upon their arrival (see Appendix B.1) was all that was communicated before they entered the experimental chamber—apart from an overall polite, friendly tenor (with some enthusiasm, to help sell the cover story).

It should also be noted that any behavior desired of subjects by the administrator was complex in nature and, arguably, entirely emergent from the subject's interaction with the software, and were therefore impossible to convey to subjects. Still, a double-blind implementation would raise the results of this study above such suspicions, however off-base they might be.

### ***Going Public***

In this experiment, the CASPER was shown to be effective in relieving frustration, as well as the overall negative affect felt by subjects toward the source of their frustration—the computer—while subjects had no knowledge of the agent's existence, its goals or its method for achieving its goals. It remains to be seen whether users will accept such an agent when they know about the agent beforehand. These effects should be determined in future tests of the agent.

### ***Repeat Performances?***

What effects do time, experience and expectation have on the effectiveness of a CASPER-type agent? Will humans accept and/or benefit from the behavior of such an agent on more than one occasion, or is this a one-time-only phenomenon, akin to *fool me once, shame on me...*?

### ***Determining Deeper Effects of this Genre***

What effects might this agent have on users over time? What ramifications might there be for determining the effects of the potential problems of authenticity and locus of control discussed in Chapter 9? Future studies must be conducted to try to determine these effects and others described in Chapter 9, and subjects from this

work need to be repeatedly interviewed, in-depth, to try to determine the effects of such devices.

## CHAPTER 11

# Conclusions

This thesis establishes by example that computer interfaces can be designed to actively help users recover from strong, negative emotional states, especially those related to frustration—including, but not limited to, the frustration that can arise from using computer systems. Further, this thesis shows that such interfaces have the added benefit of making computer users feel better *about* the very computer systems and applications that were involved in eliciting frustration in the user in the first place.

The implications for this work span dozens of important domains including human productivity, customer satisfaction, product design, emotion theory, human-computer interaction, the fields of human psychology, sociology, communications, business administration, and even philosophy—not to mention the quality of day-to-day life for the average computer user.

Further, this study demonstrates that computer systems can provide direct, effective support for a critical, yet hitherto unaddressed and unsupported, aspect of human experience: namely, modulating strong, negative emotions. Such systems, along with the nature of the task they can perform, represent a fundamental shift in the way humans think about computers, and their role in human life, for better or for worse.

This work extends the known boundaries of the Media Equation, by showing that human users can and will use media in ways hitherto assumed to be uniquely personal and social—to envision and use the computer as confidante, not just as



service provider—to actively help manage their emotional states, just as humans use other humans in this capacity. This thesis also begins to articulate guidelines for the design of active emotion self-management support in future systems.

This research problem was non-trivial, and for a variety of reasons: First, interpersonal emotion management, the ability to assist in the regulation of others' emotions, has been identified as a critical social skill, yet effectiveness of such skills—from motivating others to active listening—is often hard to achieve in practice, and specific criteria for effectiveness is often described as a complex balance of knowledge, sensitivity, expressiveness, and timing. The ability of machines to exhibit skill at emotion management assistance presented an even greater challenge, and such a system has, to this researcher's knowledge, yet to be demonstrated elsewhere.

Along with important practical contributions to a number of fields, this work, and its demonstration of a new genre of human-computer interaction, carries some potentially serious implications for the culture and society. Some of these implications are positive, but many more are negative—if not out-and-out risks to the fabric of this and other cultures, and to the way humans think about themselves. There are a number of hard questions regarding the authenticity of such interactions,

Finally, with the advent of emotionally-assistive technologies (and their theoretical framework), the very nature of what they are designed to support—the very fact that they represent systems that can affect demonstrable, beneficial change in the user's emotional state—helps create a much more complete picture of the human user than is currently conceived in many disciplines, including human-computer interaction. Human emotional needs, and the fact that nearly every single human user has them, need no longer be ignored as a “human factor” in the design of modern technology.

**APPENDIX A:**  
OFF-SITE SUBJECT EXPERIENCE MATERIALS

OFF-SITE SUBJECT MATERIALS 1.

The Soliciting Flier

Example of flier soliciting human subjects, posted on MIT campus bulletin boards

**Earn \$10<sup>00</sup> in 40 minutes**

**...and automatically enter to win one of two \$100 cash prizes...**

**...playing a game!**

*Come and evaluate a new kind of computer game at the MIT Media Lab.  
Walk out with hard currency and a chance to win one of two \$100 cash prizes.*

**English-speaking men and women needed • No game experience necessary**

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

Play a game for \$10.  
enter to win \$100  
phaedra@media.mit.edu  
(617) 253-0384

## OFF-SITE SUBJECT MATERIALS 2.

## Informational Email

Example of informational email reply and request to schedule subject, in response to first contact stating interest (following postings of flier on MIT bulletin boards)

To: Clarence <email.address@mit.edu>  
 From: Jonathan Klein <phaedra@media.mit.edu>  
 Subject: Re: Game for \$10  
 Date: Friday, Aug 7, 1998 3:31pm EDT

-----  
 On Friday, Aug 7, Clarence wrote:

> I'm interested in playing your game for \$10. Can you give me more details?  
 > What do I have to do to participate?  
 >  
 > Clarence

Clarence,

Thanks for responding, and we hope you can participate! The details are thus: we've been building a new kind of web-based adventure game that uses a new technique for Internet communication, and we need men and women to help evaluate it.

When you arrive to participate, you will be given a voucher for US\$10, redeemable at the MIT Cashiers' office in building 10. And, since we want players to play as well as they can, we will be giving out two \$100 prizes, one each to the two highest scorers. In this game, intelligence matters even more than reflexes, so novice game players with smarts have a better chance of winning than experienced gamers without them.

Can we schedule you for sometime late next week? We have many open slots available on Thursday and Friday, and even more the week after. If you can give me a few different times/days you would be able to come to the Media Lab next week, one of them should surely work. We can also schedule an evening session, if that's more convenient for you.

Let me know what's good for you as soon as you can, so we can block out that time. Also, I need to send you a brief email questionnaire to fill out and send back to me as soon as is convenient (all info collected is strictly confidential, and the 10 minutes needed to fill it out is counted towards total time). Okay?

I look forward to hearing from you soon.

Cheers,

--Jonathan

-----  
 Jonathan Klein  
 phaedra@media.mit.edu  
 E15-394, 20 Ames St. Cambridge MA

-----  
 Grad Student Research Assistant  
 MIT Media Laboratory  
 (617) 253-0384

OFF-SITE SUBJECT MATERIALS 3.

## Confirmation and Questionnaire

Emailed note of time confirmation and transmission of Mehrabian's Trait Arousability questionnaire (Mehrabian 95)

To: Clarence <email.address@mit.edu>  
 From: Jonathan Klein <phaedra@media.mit.edu>  
 Subject: Game study confirmation (11am Fri) and questionnaire  
 Date: Monday, Aug 10, 1998 12:24pm EDT

-----  
 On Monday, Aug 5, Clarence wrote:

> Friday at 11am or 3pm are both good for me. Let me know...  
 >  
 > Clarence

Clarence,

Thanks for participating! This Friday (August 14) at 11am works perfectly, and I have reserved that timeslot for you. Let's plan to meet in Room E15-383 at the Media Lab.

Here are directions, to the building and to the room: The Media Lab is in Building E15 (a.k.a. The Weisner Building), at 20 Ames Street on the east side of campus. It's the building with the big white tiles, rounded corners and the big modern concrete arch connected to the side of it. Room 383 is on the 3rd floor, left off the elevator and straight through 2 sets of doors.

As we discussed, the first part of the study is an email questionnaire, which follows, below. If you can fill it in and return it to me as soon as possible (by sometime this evening?), I'd really appreciate it. Remember that it only takes about 10 minutes to fill it out, which is counted towards the total participation time).

Thanks, and I look forward to seeing you Friday morning at 11.

Cheers,

--Jonathan

----- Jonathan Klein phaedra@media.mit.edu E15-394, 20 Ames St. Cambridge MA	----- Grad Student Research Assistant MIT Media Laboratory (617) 253-0384
---	--

x-----x-----x-----x--- PRE-GAME QUESTIONNAIRE ---x-----x-----x-----x

Thanks for agreeing to participate in our study evaluating game software. For step one of this study, we ask that you please take ten minutes or so to fill out the questionnaire below. In order to gauge your evaluation of the game, we would like to get an idea of what you are like.

When you come in to evaluate the game, you will be paid \$10 in a voucher on the spot, and will have a chance to win one of two \$100 prizes. Please be advised that your entire involvement in this study should amount to less than 45 minutes, including answering the questions below.

Please note that your entire participation in this study is strictly confidential, and neither your name nor any information you provide will ever be used outside the domain of this study. Further, please note that answering the following questions is very important to our research, but is entirely voluntary. Feel free to skip any question you don't want to answer.

*[NOTE: The following portion of this letter contained the text of Dr. Albert Mehrabian's Trait Arousability Scale (Mehrabian 95). The Trait Arousability Scale is copyrighted by Dr. Mehrabian, and can be obtained by contacting him directly.]*

OFF-SITE SUBJECT MATERIALS 4.

### Questionnaire Receipt Confirmation

Emailed confirmation of administrator's reception of the completed questionnaire

To: Clarence <email.address@mit.edu>  
From: Jonathan Klein <phaedra@media.mit.edu>  
Subject: Re: Game study confirmation (11am Fri) and questionnaire  
Date: Monday, Aug 10, 1998 8:19pm EDT

-----  
Clarence,

Thanks for sending back the questionnaire. You're all set! I have reserved 11am on Friday morning for you in Room E15-383 at the Media Lab, and I look forward to seeing you then.

Cheers,

--Jonathan

-----  
Jonathan Klein  
phaedra@media.mit.edu  
E15-394, 20 Ames St. Cambridge MA

-----  
Grad Student Research Assistant  
MIT Media Laboratory  
(617) 253-0384

OFF-SITE SUBJECT MATERIALS 5.

### Email Reminder the Night Before

Reminder email note sent out the evening before the date of participation

To: Clarence <email.address@mit.edu>  
From: Jonathan Klein <phaedra@media.mit.edu>  
Subject: Reminder -- game study tomorrow (Friday) morning at 11am  
Date: Thursday, Aug 13, 1998 8:19pm EDT

-----  
Hey Clarence,

I just wanted to remind you that you are scheduled to come to Room E15-383 at the Media Lab tomorrow morning at 11am to participate in a study and help evaluate our game. Remember, you will be paid \$10 on the spot for your time, with a chance to win a lot more.

The Media Lab is in Building E15 (a.k.a. The Weisner Building), at 20 Ames Street on the east side of campus. It's the building with the big white tiles, rounded corners and the big modern concrete arch connected to the side of it. Room 383 is on the 3rd floor, left off the elevator and straight through 2 sets of doors.

Thanks again, and I look forward to seeing you tomorrow evening at 6.

Cheers,

--Jonathan

-----  
Jonathan Klein  
phaedra@media.mit.edu  
E15-394, 20 Ames St. Cambridge MA

-----  
Grad Student Research Assistant  
MIT Media Laboratory  
(617) 253-0384

APPENDIX B:  
ON-SITE SUBJECT EXPERIENCE MATERIALS

ON-SITE SUBJECT MATERIALS 1.

## Briefing script

Read to all subjects upon arrival at laboratory  
(comments to the administrator in italics)

*NOTE: Be very friendly to the subject, but in a natural way.*

*Greet the subject with this script, the consent form, and a signed payment voucher, and lead them to a seat in the briefing area (the nondescript room outside the conference room where the experiment computer is).*

*If the subject asks any questions during this process, politely state:*

"I'm sorry, but I'm only allowed to read from this script. I will be able to answer any questions you have when the experiment is completed."

*Otherwise, read the following script:*

"Thank you for coming and participating in this study!"

"We have built a prototype of an Internet-based computer game. We'd like you to work with this game, and tell us how you feel about it. That's basically it. A few questions, though: Have you ever played Internet-based graphical computer games before? *(If they look puzzled or ask if networked local games qualify, say "Like,*



have you played games on the Web?") *Wait for them to finish their answer, while nodding. Then ask: "Okay. Have you played computer game, PC games, video games, Nintendo, that sort of thing before?" Again, wait for them to finish their answer, while nodding.*

(NOTE: While the number of people who had played games on the Web, nearly all subjects verbally reported having played some form of computer game or another at some time in their lives).

"One thing that computer game players and video game players invariably notice when they play games on the Web for the first time, particularly graphical games with animation, is that the graphics on the Web move ve...ry...slow...ly ...on...the...Web...like...this...it's...pain...ful." (*Gesture with hand in tandem with each syllable, which should be spoken staccato, every 1/2 to 1 second each, and then smile.* "This slowness of graphics has been one reason why you don't see a lot of games being played on the Web yet.

"One of the things we're working on here is trying to address this problem. (There are actually two groups here working on this project, one working on the game and another on network stuff.) We've developed a new protocol for Internet communication that greatly speeds up the graphics for games, particularly over slow networks and modem connections. It makes the speed of graphics for games much faster than the current state of affairs for games on the Web, but what we find is that what we have developed is still *not as fast* as most modern home computer games and PC games, or even Nintendo-style video games. We're interested to know if it's fast *enough*, though, so we're interested in how you feel about it. Anyway, while this element of the experience is something we want you to evaluate, do expect it to be slower than you might be used to. Smooth, but slower.

"Have you ever heard the term "First Playable" before? (NOTE: No subjects reported having heard the term before.) I wouldn't expect you to; it's a term from the computer game industry. A First Playable is kind of like an Alpha version—a very early version of a computer game, with certain criteria—particularly a graphical, character-based game. Now a character-based game is one where there is a character on screen that represents you in the world, and you can move the character around in the world, and manipulate the environment in some way. Many games are character-based; most Nintendo and PlayStation games are like that, and many PC games are, too.

"So anyway, every Nintendo game, every Sega Genesis game, Doom, Quake—they all go through a first playable stage. In a character-based game, the First Playable is the point at which the character is pretty much fleshed out, in both software and graphically, but the world generally *isn't*. It's pretty bare-bones; the rich graphics that will be in the final version are not yet installed, and the kinds of things the character can do in the world are generally a subset of the things it will eventually be able to do. But, it's a necessary phase in the development of these kinds of games, and a really critical step, since it's the first time that the game's developers can have what we call 'play testers' come in and bang away at the game, and evaluate it for playability and a lot of other criteria.

“So, we need you to be a play-tester. Okay?”

“So, the problem is—you ready for the problem? The problem is that because the game doesn’t have many of the bells and whistles added to it yet—and there’s no music yet, no sound effects, and lots of elements in the game aren’t added yet—the game may not be as *immersive* an experience as we’d like, as good games are, and as we expect this game to be once it’s done.

“You know, when a game is good, it’s really ‘immersive’, right? You’ve probably played a good game on the computer before; you become totally involved in the game, you lose track of time, your concentration is total yet effortless, and you just naturally play as hard as you can, and try to do as well as you can.

“Well, a perennial problem with first-playables is that they may not be as immersive as the games will be when they’re done. The problem is, we need you to *play* as if this were the finished version. We want you to evaluate the game as if it were a real game situation. But since it may not be as immersive as it can be, like a real game situation, we need you to play like it is anyway.

“So that’s why we’re offering an incentive for you to do well. We’re offering two \$100 prizes to the top two scorers in the game. If you score one of the top two scores, you will win \$100. And we’ll be announcing the winner within 2 weeks. Hopefully, this will encourage you to play as hard as you can, as intelligently as you can, and do what *we* want you to do, which is to score more points than the other play-testers!

“And by the way: We’ve play-tested the game on experienced gamers who play an average of 10 or more hours per day, and we’ve also tested the game on people who have never picked up a mouse to play a game before in their lives. What we found is that novice players who are intelligent and use their heads do as well or better at this game than experienced gamers who just rely on their reflexes. You need normal reflexes for this game, but if you are a smart person and use your head, you have a good chance of doing well.”

“Okay. You’ll be asked to play the game twice, stopping to answer questions after playing each time. Directions for playing the game will be provided on-screen, as will the first set of questions you will be asked. So you’ll read the directions, you’ll play the game once, you’ll answer some questions on-screen, you’ll read the directions for the *second* game ‘cause the second game is different from the first, you’ll play the second game, and then when you’re done, the computer will notify you to get up and come get me—and I’ll be right out here—and I’ll give you a brief, 3-page paper questionnaire to fill out. The entire experience, start to finish, should take you less than 40 minutes. Okay?”

"First, we'd like to give you your payment voucher, redeemable at the cashier's office (*building 10, at the dollar bill mural in the Infinite Corridor, if they've never been*). [*have them fill out voucher, and have them hand it back to you to copy.*]

"Now, we'd like you to read and fill out this consent form." [*hand subject consent form, and pen if necessary. While subject fills out consent form, copy the filled-out payment voucher, and/or otherwise look busy; do not rush the subject, or make*

*them feel nervous.] When they are done filling out the consent form, say, “Thanks. Okay, let’s get you started. Right this way.”*

*Lead the subject into the experiment room, and offer them a seat in front of the computer. Turn on the video camera, and then say to the subject:*

"Again, the directions, game and everything is all on-line. Everything should be self-explanatory. If there’s a problem, I’ll be right outside the room. Otherwise, good luck!"

## ON-SITE SUBJECT MATERIALS 2.

## Consent form

(Read and signed after administrator finishes reading script to subject, and US\$10 payment voucher is filled out and signed)

## Consent Form

Your participation in the following experiment is completely voluntary. You are free to withdraw this consent at any time, for any reason, and to request that any data collected be destroyed. If at any time you feel uncomfortable, or unsure that you wish your results to be part of the experiment, you may discontinue your participation with no repercussions.

In a few minutes, you will be asked to play a prototype of a web-based, graphical computer game we are developing. The game will consist of moving your character around in a virtual fantasy world via the keyboard, finding and picking up treasure while avoiding walls, while the scenery passes by. You will have two opportunities to play the game.

You may be asked during your interaction with the computer to answer a series of on-line survey questions, which will include questions about the game and your reactions to it. You may also be asked by the computer to clarify your responses. Your participation in this game may be recorded on videotape. Finally, you may be asked to fill in one or more questionnaires about yourself. If at any time you are uncomfortable, either physically or with what you are being asked to do, you are free to ask that the experiment be suspended.

A voucher for payment of US \$10 is being given to you for your participation in this study, along with instructions for its redemption. Payment will be prorated for early withdrawal.

To encourage participants to achieve the highest possible score, the player with the top score will win one of two bonus prizes of US\$100, to be determined at the conclusion of this study. In the unlikely event of a more-than-two-way tie for the top score, the winner will be selected at random, via a drawing, from among the top scorers. The winner will be contacted via phone or email on or before August 31st, 1998.

Any responses that are collected during the experiment will be completely anonymous. From this point forward, you will be referred to only as the ID number which appears on the upper right corner of this packet.

If you have any questions, the experimenter will gladly answer them at the conclusion of your participation.

(go on to the next page)



Please read the following and sign on the lines below:

"I, the undersigned, have read and understood the explanations of the following research project and voluntarily consent to my participation in it. I understand that my responses will remain confidential and that I may terminate my participation at any time.

In the extremely unlikely event of physical injury resulting from participation in this research, I understand that medical treatment will be available from the MIT Medical Department, including first aid emergency treatment and follow-up care as needed, and that my insurance carrier may be billed for the cost of such treatment. However, no compensation can be provided for medical care apart from the foregoing. I further understand that making such medical treatment available, or providing it, does not imply that such injury is the Investigator's fault. I also understand that by my participation in this study I am not waiving any of my legal rights.

I understand that I may also contact the Chairman of the Committee on the Use of Humans of Experimental Subjects, MIT 253-6787, if I feel I have been treated unfairly as a subject.

Name: \_\_\_\_\_

Date: \_\_\_\_\_

Location: \_\_\_\_\_

Additionally, please read the following paragraph:

Videotapes and/or audiotapes may be collected of your participation. This data will be used for experimental purposes only, and after the data collection is over, they will be permanently stored in a private archive. In the future, portions of this record may be published and/or presented in scientific journals and/or in scientific conference proceedings, but will never be published in a non-scientific venue. Further, no other information (such as the subject's name, address or other private information) will be included in these publications. Apart from this possible usage, such data will only be viewed/used for experimental purposes. Again, at any time during or after the experiment you may request that your tapes be destroyed.

Please sign on the lines below to give special permission for the collection of this material.

Name: \_\_\_\_\_

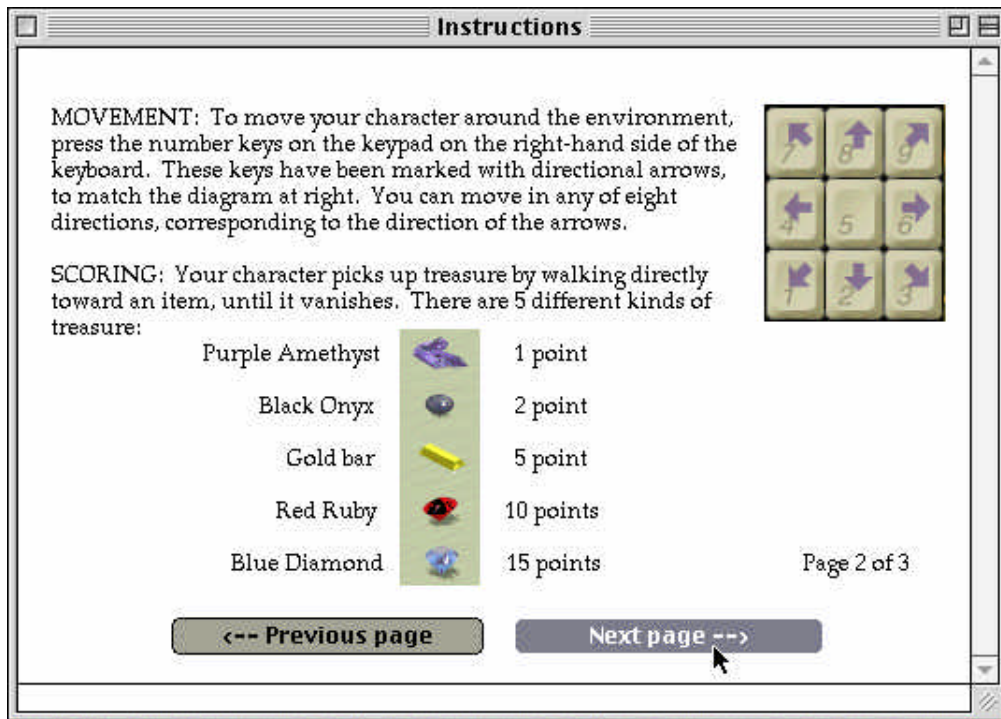
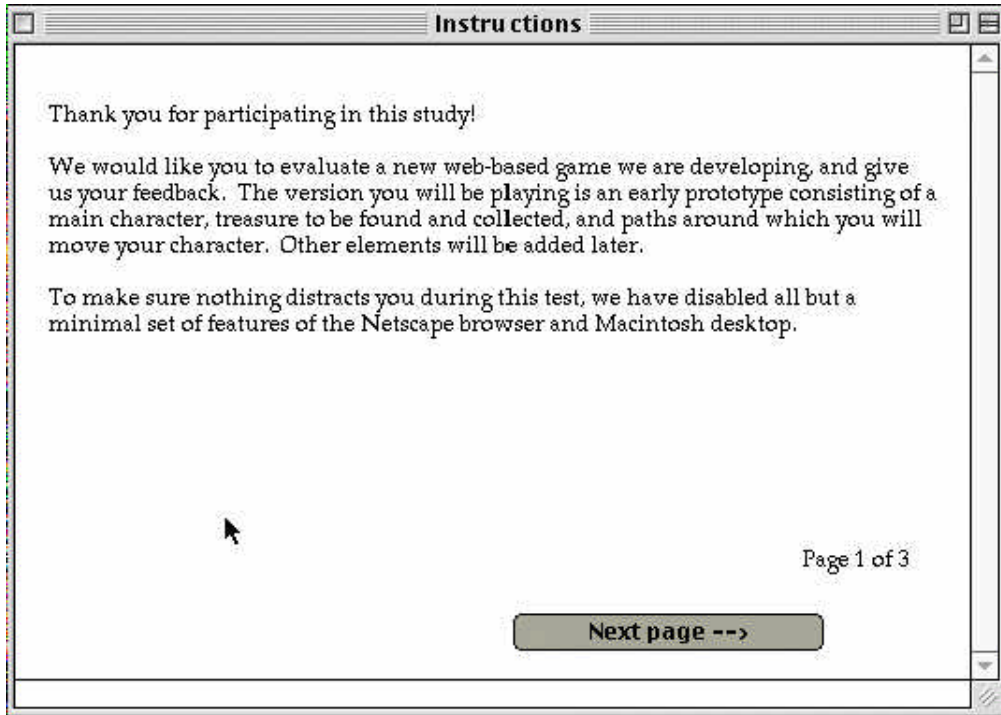
Date: \_\_\_\_\_

Location: \_\_\_\_\_

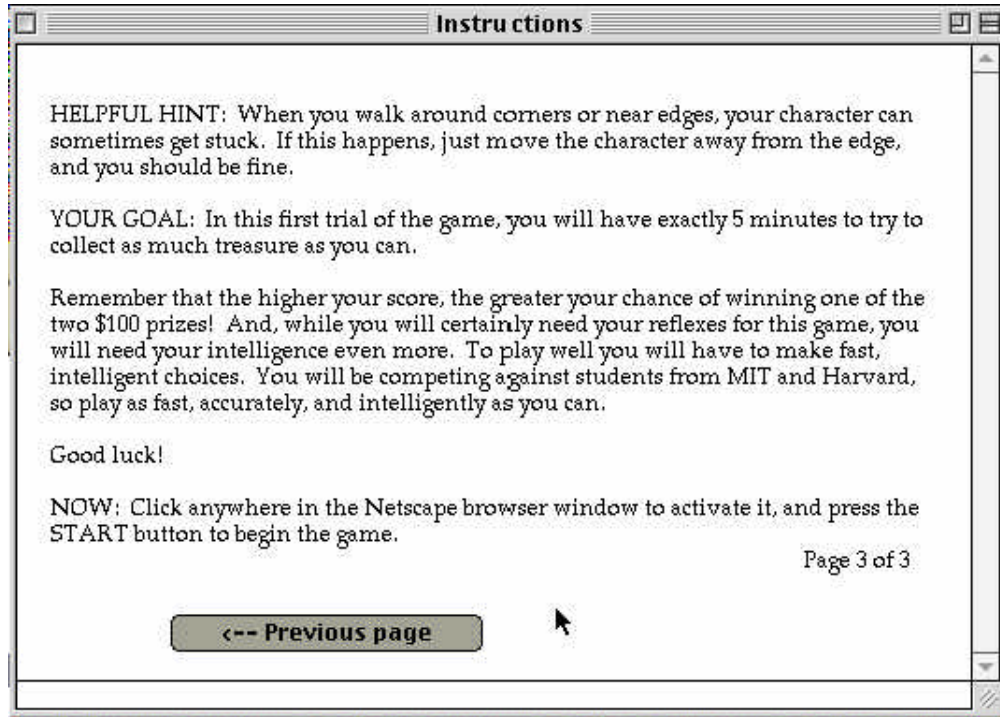
## ON-SITE SUBJECT MATERIALS 3.

## On-screen instructions, Game 1

(Read after reading and signing Consent Form, above).



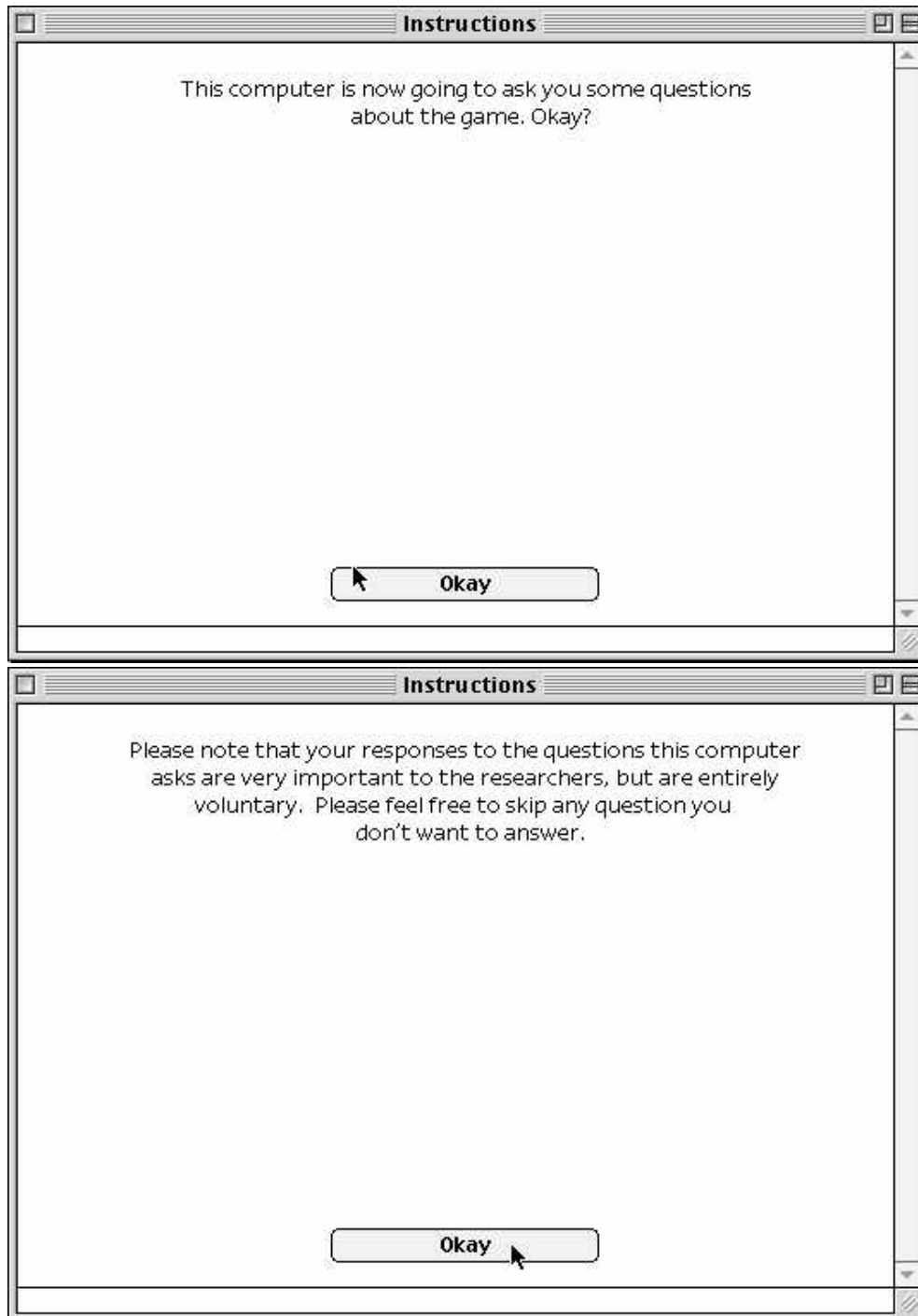




## ON-SITE SUBJECT MATERIALS 4.

## Directions for answering questionnaire (all conditions)

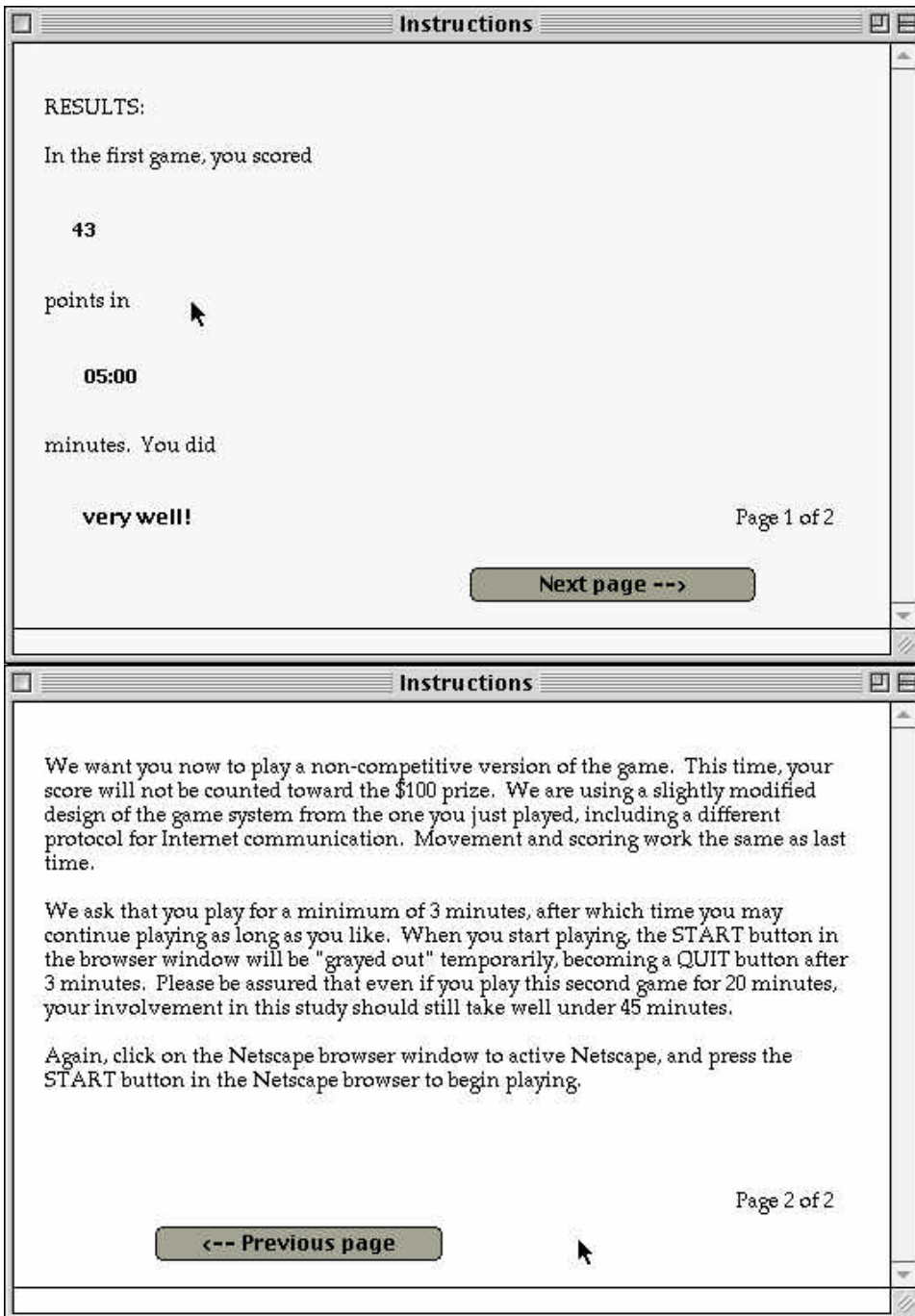
(Read immediately after finishing Game 1, and before viewing all three interactive questionnaire/response systems)



## ON-SITE SUBJECT MATERIALS 5.

## Directions for Game 2

(Read immediately after answering one of the three interactive questionnaire/response systems, corresponding to the three [CONTROL, VENT, CASPER] conditions



## ON-SITE SUBJECT MATERIALS 6.

## Exit questionnaire

(Next 3 pages—given to subjects to fill out, with a pen, following Game 2)

## EXIT QUESTIONNAIRE

Please answer the following questions as best you can. NOTE: Your answering the following questions as honestly as you can is vital to our research, but again is entirely voluntary. Feel free to skip any question you don't want to answer.

- How much did you enjoy playing the first game? \_\_\_\_\_  
\_\_\_\_\_
- How much did you enjoy playing the second game? \_\_\_\_\_  
\_\_\_\_\_
- Why did you play the second game for as long as you did? \_\_\_\_\_  
\_\_\_\_\_
- How do you feel right now? Please rate your current state according to the following criteria:

	Absolutely none at all						The most I have ever felt in my life				
	0	1	2	3	4	5	6	7	8	9	10
Frustration	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fun or enjoyment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Anger	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Satisfaction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Anxiety	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tenseness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- You rated your current level of frustration above. Compared to how you feel right now, how much frustration did you feel...

	Absolutely not frustrated at all						The most frustrated I have ever been in my life				
	0	1	2	3	4	5	6	7	8	9	10
...when you entered this room today?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...immediately after the first game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after answering the questions between games?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after the second game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If there were changes in how much frustration you felt over the course of your participation in this experiment, what do you think caused these changes?

\_\_\_\_\_  
\_\_\_\_\_

## EXIT QUESTIONNAIRE , PAGE 2

6. Compared to how you feel right now, how much fun and/or enjoyment did you feel...

	Absolutely none at all										The most I have ever felt in my life	
	0	1	2	3	4	5	6	7	8	9	10	
...when you entered this room today?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...immediately after the first game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after answering the questions between games?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after the second game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If there were changes in how much fun/enjoyment you felt over the course of your participation in this experiment, what do you think caused these changes?

---



---

7. Compared to how you feel right now, how much anger did you feel...

	Absolutely none at all										The most I have ever felt in my life	
	0	1	2	3	4	5	6	7	8	9	10	
...when you entered this room today?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...immediately after the first game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after answering the questions between games?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after the second game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If there were changes in how much anger you felt over the course of your participation in this experiment, what do you think caused these changes?

---



---

## EXIT QUESTIONNAIRE, PAGE 3

8. Compared to how you feel right now, how much satisfaction did you feel...

	Absolutely none at all										The most I have ever felt in my life	
	0	1	2	3	4	5	6	7	8	9	10	
...when you entered this room today?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...immediately after the first game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after answering the questions between games?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...after the second game?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If there were changes in how much satisfaction you felt over the course of your participation in this experiment, what do you think caused these changes?

---



---

9. Did the questionnaire you answered after the first game make you feel better or worse, overall?

Much Better	About the same					Much Worse				
5	4	3	2	1	0	-1	-2	-3	-4	-5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Is there anything you'd like to add about your experience in this study?

---



---



---



---

## ON-SITE SUBJECT MATERIALS 7.

## Debriefing statement

(Given to subjects to read following their completion of Exit Questionnaire)

## Debriefing Statement

The experiment you just participated in was designed to elicit specific emotions, so that we may understand the nature and the range, as well as the similarities, between humans as they experience emotions on a daily basis.

The task that you participated in was designed to be especially frustrating, since we are most interested in how people manage their emotional responses to computer interaction that is frustrating in nature. We are also interested in how a computer might be able to assist a user in managing his or her feelings of frustration, and how a computer might help the user feel better.

Since emotional states are extremely difficult to elicit consciously, we needed to make you believe that you were participating in this experiment in order to evaluate the “web-based” game you just played. However, this was a deception. The game is not, in fact, on the World Wide Web, but is mounted locally. Additionally, any “network delays” you may have experienced while playing the game were designed into the experiment on purpose, to attempt to elicit an emotional reaction of frustration.

Please understand that this deception was an absolutely critical component of our research, and that this experiment was carefully evaluated and approved by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES).

If at any time, now or later, you experience any ill effects (either mental or physical) as a result of your participation in this experiment, please do not hesitate to tell the experimenter, or call 253-0384. Again, you have the option to withdraw your participation in this experiment now, including its video/audio record, with no repercussions.

Feel free to ask any questions about the experiment at this time.

Your help has been greatly appreciated, and will aid the Media Lab in the construction of new software that assists humans in managing their own emotional responses, and ultimately will be used to construct an enhanced user-computer environment.





**APPENDIX C:**  
DATA ANALYSES AND RELATED DOCUMENTS

ANALYSIS 1.

Frequency distribution of subject age ranges

## STATISTICS

	N		Mean	Median	Std. Deviation	Variance	Range
	Valid	Missing					
<b>Subject's age range</b>	69	1	3.62	3.00	1.47	2.15	6

## SUBJECT'S AGE RANGE

Code	Age Range	Frequency	Percent	Valid Percent	Cumulative Percent
<b>Valid</b>	<b>1 10 - 14</b>	1	1.4	1.4	1.4
	<b>2 15-19</b>	15	21.4	21.7	23.2
	<b>3 20-24</b>	22	31.4	31.9	55.1
	<b>4 25-29</b>	15	21.4	21.7	76.8
	<b>5 30-34</b>	8	11.4	11.6	88.4
	<b>6 35-44</b>	3	4.3	4.3	92.8
	<b>7 40-44</b>	5	7.1	7.2	100.0
	<b>Total</b>	69	98.6	100.0	
<b>Missing</b>	<b>System Missing</b>	1	1.4		
	<b>Total</b>	1	1.4		
	<b>Total</b>	70	100.0		

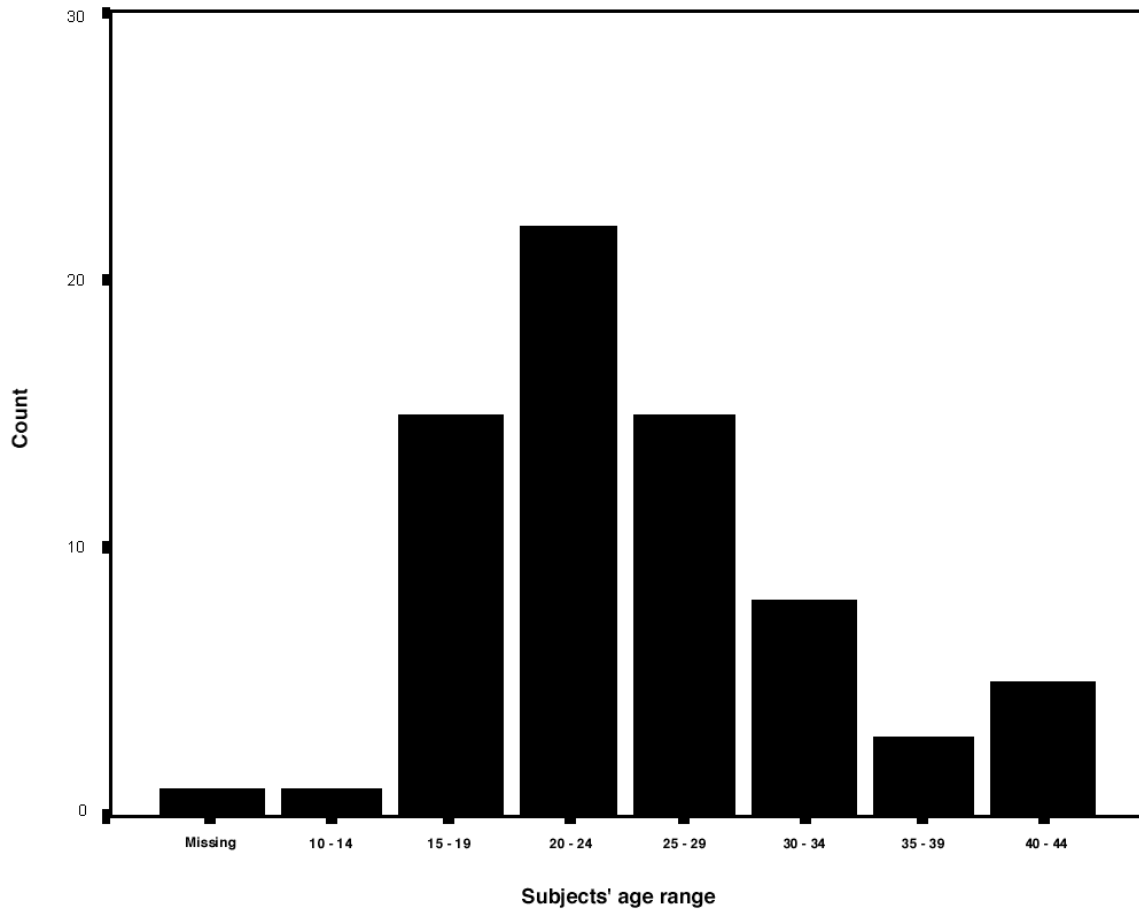


Figure C-1: Distribution of subject age ranges

ANALYSIS 2.

Manipulation check

ANOVA of self-report for frustration level immediately after Game 1, as reported on Exit Questionnaire (Significant results found)

CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
70	100.0%	0	.0%	70	100.0%
a (How much frustration felt immediately after playing game 1) by (Subject experiences delays in Game 1), (Type of questionnaire subject experiences)					

CELL MEANS(B)

		How much frustration felt immediately after playing 1	
Subject experiences delays in Game 1	Type of questionnaire subject experiences	Mean	N
No	QA (Control)	3.42	12
	QC (Vent)	3.64	11
	QD (CASPER)	3.64	11
	Total	3.56	34
Yes	QA (Control)	3.83	12
	QC (Vent)	5.00	12
	QD (CASPER)	5.58	12
	Total	4.81	36
Total	QA (Control)	3.63	24
	QC (Vent)	4.35	23
	QD (CASPER)	4.65	23
	Total	4.20(a)	70
a Grand Mean			
b (How much frustration felt immediately after playing game 1) by (Subject experiences delays in Game 1), (Type of questionnaire subject experiences)			

## ANOVA (A)

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
<b>How much frustration felt immediately after playing game 1</b>	<b>Main Effects</b>	<b>(Combined)</b>	39.587	3	13.196	2.267	.089
		<b>Subject experiences delays in Game 1</b>	26.447	1	26.447	4.543	.037
		<b>Type of questionnaire subject experiences</b>	12.409	2	6.204	1.066	.351
	<b>2-Way Interactions</b>	<b>Subject experiences delays in Game 1 * Type of questionnaire subject experiences</b>	7.022	2	3.511	.603	.550
		<b>Model</b>	46.609	5	9.322	1.601	.173
		<b>Residual</b>	372.591	64	5.822		
		<b>Total</b>	419.200	69	6.075		

a (How much frustration felt immediately after playing game 1 ( by (Subject experiences delays in Game 1), (Type of questionnaire subject experiences)

## ANALYSIS 3.

## ANOVA: Main behavioral measure

Mean time subjects played Game 2, in seconds, by condition (Strong significant differences found)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
70	100.0%	0	.0%	70	100.0%

a (The time S played Game 2, in seconds) by (Subject experiences delays in Game 1), (Type of questionnaire subject experiences)

## CELL MEANS (B)

Subject experiences delays in Game 1	Type of questionnaire subject experiences	The time S played Game 2, in seconds	
		Mean	N
No	QA (Control)	484.00	12
	QC (Vent)	453.91	11
	QD (CASPER)	703.00	11
	Total	545.12	34
Yes	QA (Control)	698.42	12
	QC (Vent)	540.25	12
	QD (CASPER)	967.50	12
	Total	735.39	36
Total	QA (Control)	591.21	24
	QC (Vent)	498.96	23
	QD (CASPER)	841.00	23
	Total	642.97(a)	70

a Grand Mean

b (The time S played Game 2, in seconds) by (Subject experiences delays in Game 1), (Type of questionnaire subject experiences)

## ANOVA (A)

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
The time S played Game 2, in seconds	Main Effects	(Combined)	2066352.938	3	688784.313	8.841	.000
		Subject experiences delays in Game 1	623067.910	1	623067.910	7.998	.006
		Type of questionnaire subject experiences	1433315.081	2	716657.540	9.199	.000
		2-Way Interactions	Subject experiences delays in Game 1 * Type of questionnaire subject experiences	97073.929	2	48536.964	.623

	<b>Model</b>	2163426.867	5	432685.373	5.554	.000
	<b>Residual</b>	4986101.076	64	77907.829		
	<b>Total</b>	7149527.943	69	103616.347		

a (The time S played Game 2, in seconds) by (Subject experiences delays in Game 1), (Type of questionnaire subject experiences)

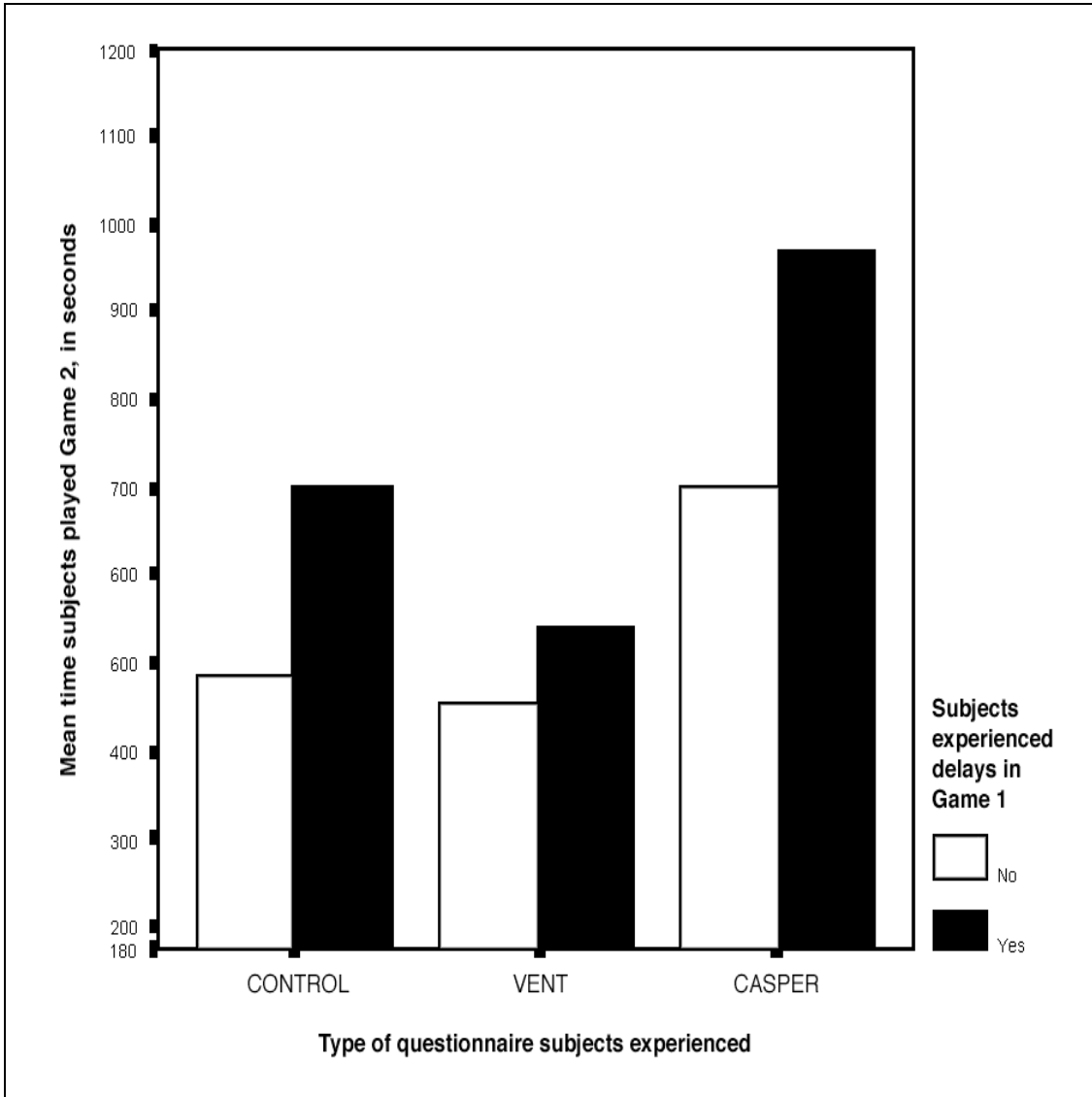


FIGURE C-2: Mean times subjects played Game 2, in seconds, by condition. Subjects in CASPER/DELAY condition played nearly 5 minutes longer, on average, than CONTROL/DELAY condition, and over 7 minutes longer, on average, than VENT/DELAY condition.

## ANALYSIS 4.

## Gender analysis

ANOVA: Mean time subjects played Game 2, in seconds, by condition, and by gender (No significant correlation found)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
70	100.0%	0	.0%	70	100.0%
a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences, Gender					



## ANOVA (A)

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
The time S played Game 2, in seconds	Main Effects	(Combined)	2101413.955	4	525353.489	6.680	.000
		Subject experiences delays in Game 1	613723.897	1	613723.897	7.804	.007
		Type of questionnaire subject experiences	1435651.732	2	717825.866	9.127	.000
		Gender	35061.016	1	35061.016	.446	.507
	2-Way Interactions	(Combined)	316363.942	5	63272.788	.805	.551
			Subject experiences delays in Game 1 * Type of questionnaire subject experiences	100616.121	2	50308.060	.640
		Subject experiences delays in Game 1 * Gender	102.479	1	102.479	.001	.971
		Type of questionnaire subject experiences * Gender	218238.808	2	109119.404	1.387	.258
	3-Way Interactions	Subject experiences delays in Game 1 * Type of questionnaire subject experiences * Gender	170224.180	2	85112.090	1.082	.346
			Model	2588002.076	11	235272.916	2.992
		Residual	4561525.867	58	78646.998		
		Total	7149527.943	69	103616.347		

a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences, Gender

## ANALYSIS 5.

## Trait Arousability analysis

ANCOVA: Mean time subjects played Game 2, in seconds, by condition, and by raw Trait Arousability score (No significant correlation found)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
67	95.7%	3	4.3%	70	100.0%
a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences with MEHRABRW					

## ANCOVA (A, B)

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
<b>The time S played Game 2, in seconds</b>	<b>Covariates</b>	<b>MEHRABRW</b>	1113.828	1	1113.828	.014	.907
	<b>Main Effects</b>	<b>(Combined)</b>	2109924.504	3	703308.168	8.688	.000
		<b>Subject experiences delays in Game 1</b>	599970.923	1	599970.923	7.412	.008
		<b>Type of questionnaire subject experiences</b>	1529767.577	2	764883.788	9.449	.000
	<b>2-Way Interactions</b>	<b>Subject experiences delays in Game 1 * Type of questionnaire subject experiences</b>	117677.390	2	58838.695	.727	.488
		<b>Model</b>	2228715.722	6	371452.620	4.589	.001
		<b>Residual</b>	4856982.845	60	80949.714		
		<b>Total</b>	7085698.567	66	107359.069		
a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences with MEHRABRW							
b Covariates entered first							

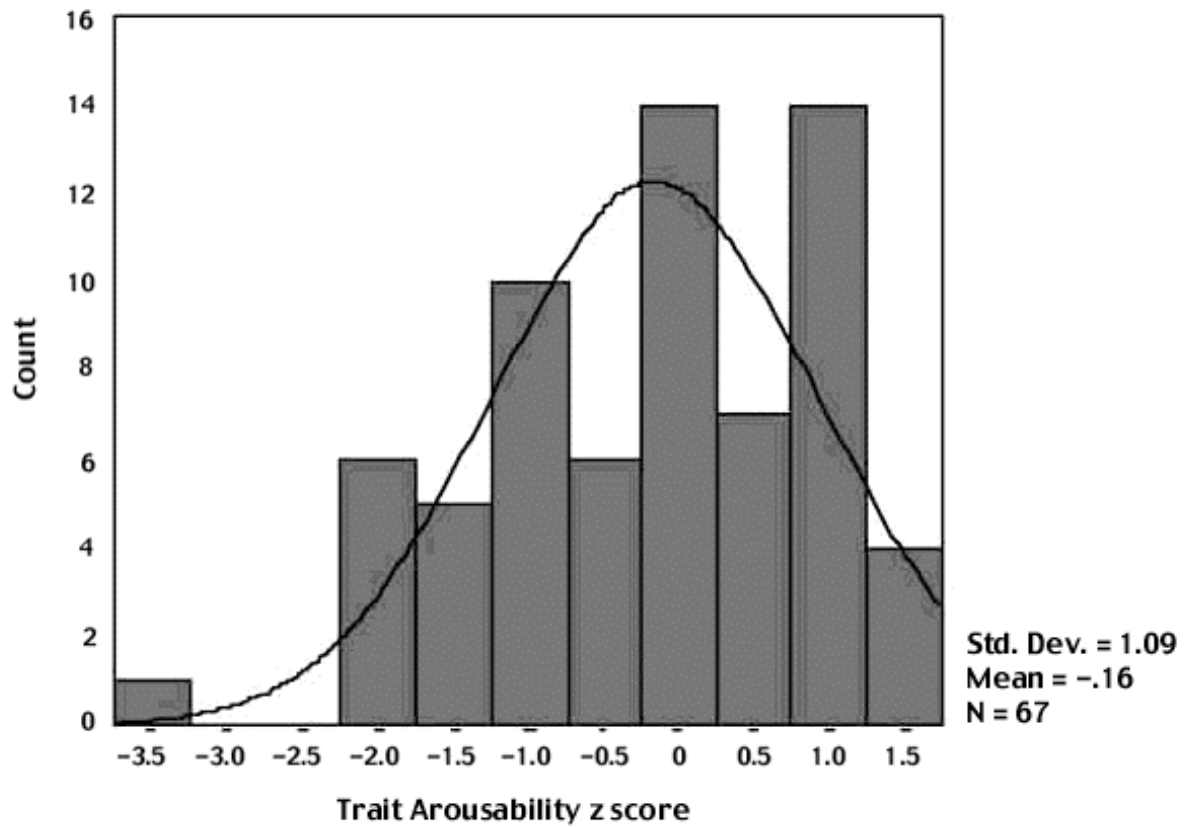


FIGURE C-3. Mean distribution of Trait Arousability  $z$  scores (linear raw score adjusted for  $\underline{M}$  and  $\underline{SD}$ ), across subjects

## ANALYSIS 6.

## 2nd Trait Arousability Analysis

ANOVA: Mean time subjects played Game 2, in seconds, by condition, and by categorized Trait Arousability score (No significant correlation found)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
70	100.0%	0	.0%	70	100.0%
a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences, MEHRAB_D					

## ANOVA (A)

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
The time S played Game 2, in seconds	Main Effects	(Combined)	2115622.804	4	528905.701	6.673	.000
		Subject experiences delays in Game 1	662809.060	1	662809.060	8.362	.005
		Type of questionnaire subject experiences	1430948.725	2	715474.363	9.027	.000
		MEHRAB_D	49269.865	1	49269.865	.622	.434
	2-Way Interactions	(Combined)	406763.484	5	81352.697	1.026	.411
		Subject experiences delays in Game 1 * Type of questionnaire subject experiences	108568.450	2	54284.225	.685	.508
		Subject experiences delays in Game 1 * MEHRAB_D	13.728	1	13.728	.000	.990
		Type of questionnaire subject experiences * MEHRAB_D	288218.547	2	144109.274	1.818	.171
	3-Way Interactions	Subject experiences delays in Game 1 * Type of questionnaire subject experiences * MEHRAB_D	29860.908	2	14930.454	.188	.829
		Model	2552247.195	11	232022.472	2.927	.004
	Residual	4597280.748	58	79263.461			
	Total	7149527.943	69	103616.347			
a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences, MEHRAB_D							

## ANALYSIS 7.

## Game-Playing Experience Analysis

ANCOVA: Mean time that subjects played Game 2, in seconds, by condition, using combined mean of self-reports for frequency of computer game play (Close to significance, but none found.)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
70	100.0%	0	.0%	70	100.0%

a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences with MEAN(q3cmpgms, q6grpadv)

## ANOVA (A, B)

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
The time S played Game 2, in seconds	Covariates	MEAN(q3cmpgms, q6grpadv)	280138.947	1	280138.947	3.546	.064
	Main Effects	(Combined)	1810241.645	3	603413.882	7.638	.000
		Subject experiences delays in Game 1	444173.418	1	444173.418	5.622	.021
		Type of questionnaire subject experiences	1399781.780	2	699890.890	8.859	.000
	2-Way Interactions	Subject experiences delays in Game 1 * Type of questionnaire subject experiences	81938.039	2	40969.020	.519	.598
		Model	2172318.631	6	362053.105	4.583	.001
		Residual	4977209.312	63	79003.322		
		Total	7149527.943	69	103616.347		

a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences with MEAN(q3cmpgms, q6grpadv)

b Covariates entered first

## ANALYSIS 8.

## Frequency Dichotomy analysis

ANOVA: Frequency dichotomy analysis, in preparation for analysis of time on task as covariate (Analysis 9)

## CASE PROCESSING SUMMARY(A)

<b>Cases</b>					
<b>Included</b>		<b>Excluded</b>		<b>Total</b>	
<b>N</b>	<b>Percent</b>	<b>N</b>	<b>Percent</b>	<b>N</b>	<b>Percent</b>
70	100.0%	0	.0%	70	100.0%
a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences, FREQDICH					

## CELL MEANS (B)

			The time S played Game 2, in seconds		
Subject experiences delays in Game 1	Type of questionnaire subject experiences	FREQDICH	Mean	N	
No	QA (Control)	.00	479.00	10	
		1.00	509.00	2	
		<b>Total</b>	484.00	12	
	QC (Vent)	.00	468.75	8	
		1.00	414.33	3	
		<b>Total</b>	453.91	11	
	QD (CASPER)	.00	703.00	11	
		1.00		0	
		<b>Total</b>	703.00	11	
	<b>Total</b>	<b>Total</b>	.00	561.14	29
			1.00	452.20	5
			<b>Total</b>	545.12	34
Yes	QA (Control)	.00	714.14	7	
		1.00	676.40	5	
		<b>Total</b>	698.42	12	
	QC (Vent)	.00	525.00	5	
		1.00	551.14	7	
		<b>Total</b>	540.25	12	
	QD (CASPER)	.00	856.00	4	
		1.00	1023.25	8	
		<b>Total</b>	967.50	12	
	<b>Total</b>	<b>Total</b>	.00	690.50	16
			1.00	771.30	20
			<b>Total</b>	735.39	36
<b>Total</b>	QA (Control)	.00	575.82	17	
		1.00	628.57	7	
		<b>Total</b>	591.21	24	
	QC (Vent)	.00	490.38	13	
		1.00	510.10	10	
		<b>Total</b>	498.96	23	
	QD (CASPER)	.00	743.80	15	
		1.00	1023.25	8	
		<b>Total</b>	841.00	23	
	<b>Total</b>	<b>Total</b>	.00	607.13	45
			1.00	707.48	25
			<b>Total</b>	642.97(a)	70
a Grand Mean					
b The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences, FREQDICH					



## ANOVA (A)

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
<b>The time S played Game 2, in seconds</b>	<b>Main Effects</b>	<b>(Combined)</b>	2084822.042	4	521205.510	6.279	.000
		<b>Subject experiences delays in Game 1</b>	429729.063	1	429729.063	5.177	.027
		<b>Type of questionnaire subject experiences</b>	1446893.490	2	723446.745	8.716	.000
		<b>FREQDICH</b>	18469.103	1	18469.103	.223	.639
	<b>2-Way Interactions</b>	<b>(Combined)</b>	154696.658	5	30939.332	.373	.865
		<b>Subject experiences delays in Game 1 * Type of questionnaire subject experiences</b>	44607.984	2	22303.992	.269	.765
		<b>Subject experiences delays in Game 1 * FREQDICH</b>	356.801	1	356.801	.004	.948
		<b>Type of questionnaire subject experiences * FREQDICH</b>	54878.440	2	27439.220	.331	.720
	<b>3-Way Interactions</b>	<b>Subject experiences delays in Game 1 * Type of questionnaire subject experiences * FREQDICH</b>	12610.663	1	12610.663	.152	.698
<b>Model</b>			2252129.362	10	225212.936	2.713	.008
<b>Residual</b>			4897398.581	59	83006.756		
<b>Total</b>			7149527.943	69	103616.347		
a The time S played Game 2, in seconds by Subject experiences delays in Game 1, Type of questionnaire subject experiences, FREQDICH							

## ANALYSIS 9.

## Game 2 performance analysis

ANCOVA: The effect of condition on Game 2 performance (using Frequency Dichotomy from Analysis 8 as covariate)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
70	100.0%	0	.0%	70	100.0%
a G2SCORE by Subject experiences delays in Game 1, Type of questionnaire subject experiences with The time S played Game 2, in seconds					

## CELL MEANS (B)

		G2SCORE	
Subject experiences delays in Game 1	Type of questionnaire subject experiences	Mean	N
No	QA (Control)	211.25	12
	QC (Vent)	198.82	11
	QD (CASPER)	255.45	11
	Total	221.53	34
Yes	QA (Control)	244.42	12
	QC (Vent)	214.67	12
	QD (CASPER)	404.67	12
	Total	287.92	36
Total	QA (Control)	227.83	24
	QC (Vent)	207.09	23
	QD (CASPER)	333.30	23
	Total	255.67(a)	70
a Grand Mean			
b G2SCORE by Subject experiences delays in Game 1, Type of questionnaire subject experiences with The time S played Game 2, in seconds			

## ANOVA (A, B)

			Experimental Method					
			Sum of Squares	df	Mean Square	F	Sig.	B
<b>G2 SCORE</b>	<b>Covariates</b>	<b>The time S played Game 2, in seconds</b>	734114.566	1	734114.566	148.094	.000	.320
	<b>Main Effects</b>	<b>(Combined)</b>	10352.465	3	3450.822	.696	.558	
		<b>Subject experiences delays in Game 1</b>	1240.004	1	1240.004	.250	.619	
		<b>Type of questionnaire subject experiences</b>	9789.474	2	4894.737	.987	.378	
	<b>2-Way Interactions</b>	<b>Subject experiences delays in Game 1 * Type of questionnaire subject experiences</b>	33211.024	2	16605.512	3.350	.041	
		<b>Model</b>	777678.055	6	129613.009	26.147	.000	
		<b>Residual</b>	312297.388	63	4957.101			
		<b>Total</b>	1089975.443	69	15796.746			
a G2SCORE by Subject experiences delays in Game 1, Type of questionnaire subject experiences with The time S played Game 2, in seconds								
b Covariates entered first								

## ANALYSIS 10.

## Analysis of time played: CONTROL vs. VENT

T-Test: The time subjects played game 2, in seconds, by questionnaire type (CONTROL vs. VENT): No significant difference found

## GROUP STATISTICS

	Type of questionnaire subject experiences	N	Mean	Std. Deviation	Std. Error Mean
The time S played Game 2, in seconds	QA (Control)	24	591.21	291.91	59.59
	QC (Vent)	23	498.96	206.69	43.10

## INDEPENDENT SAMPLES TEST

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Mean	
									Lower	Upper
The time S played Game 2, in seconds	Equal variances assumed	1.437	.237	1.245	45	.219	92.25	74.07	-56.94	241.44
	Equal variances not assumed			1.254	41.489	.217	92.25	73.54	-56.21	240.71

## ANALYSIS 11.

## Analysis of time played: CASPER vs. VENT

T-Test: The time subjects played game 2, in seconds, by questionnaire type (CASPER vs. VENT): Significant difference found

## GROUP STATISTICS

	Type of questionnaire subject experiences	N	Mean	Std. Deviation	Std. Error Mean
The time S played Game 2, in seconds	QC (Vent)	23	498.96	206.69	43.10
	QD (CASPER)	23	841.00	357.16	74.47

## INDEPENDENT SAMPLES TEST

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Mean	
									Lower	Upper
The time S played Game 2, in seconds	Equal variances assumed	13.172	.001	-3.975	44	.000	-342.04	86.05	-515.46	-168.63
	Equal variances not assumed			-3.975	35.250	.000	-342.04	86.05	-516.68	-167.41

## ANALYSIS 12.

## Analysis of time played: CASPER vs. CONTROL

T-Test: The time subjects played game 2, in seconds, by questionnaire type (CASPER vs. CONTROL): Significant difference found

## GROUP STATISTICS

	Type of questionnaire subject experiences	N	Mean	Std. Deviation	Std. Error Mean
The time S played Game 2, in seconds	QA (Control)	24	591.21	291.91	59.59
	QD (CASPER)	23	841.00	357.16	74.47

## INDEPENDENT SAMPLES TEST

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Mean	
									Lower	Upper
The time S played Game 2, in seconds	Equal variances assumed	3.474	.069	-2.630	45	.012	-249.79	94.97	-441.06	-58.52
	Equal variances not assumed			-2.619	42.517	.012	-249.79	95.38	-442.20	-57.38

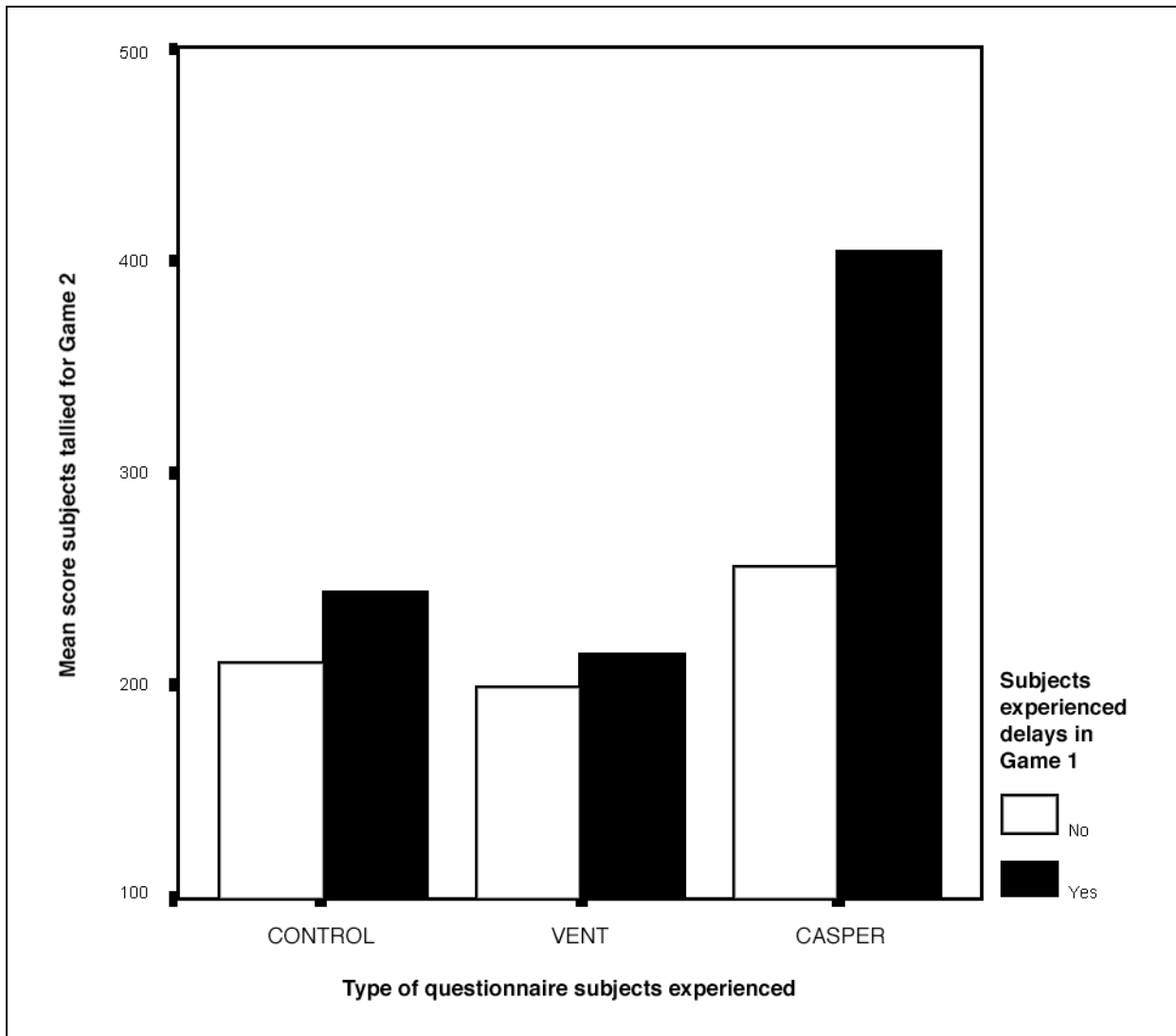


Figure C-4: Mean scores for Game 2, by condition

## ANALYSIS 13.

## Self-report on frustration level, after questionnaire

ANOVA: Self-report of frustration level experienced after questionnaire between games, by condition (measured by Exit Questionnaire; no significant results)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
70	100.0%	0	.0%	70	100.0%
a How much frustration felt after answering first questionnaire by Subject experiences delays in Game 1, Type of questionnaire subject experiences					

## ANOVA(A, B)

			Unique Method				
			Sum of Squares	df	Mean Square	F	Sig.
How much frustration felt after answering first questionnaire	Main Effects	(Combined)	17.421	3	5.807	1.176	.326
		Subject experiences delays in Game 1	11.695	1	11.695	2.369	.129
		Type of questionnaire subject experiences	5.392	2	2.696	.546	.582
	2-Way Interactions	Subject experiences delays in Game 1 * Type of questionnaire subject experiences	1.457	2	.729	.148	.863
Model			18.785	5	3.757	.761	.581
Residual			316.015	64	4.938		
Total			334.800	69	4.852		
a How much frustration felt after answering first questionnaire by Subject experiences delays in Game 1, Type of questionnaire subject experiences							
b All effects entered simultaneously							



## ANALYSIS 14.

## Analysis of self-report for questionnaire effectiveness

ANOVA: Self-report of whether the questionnaire between games made subjects feel better or worse, overall (+5 = much better, -5 = much worse), by condition. Mean positive, but no significant result found

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
69	98.6%	1	1.4%	70	100.0%
a Questionnaire_effectiveness by Subject experiences delays in Game 1, Type of questionnaire subject experiences					

## ANOVA (A, B)

			Unique Method				
			Sum of Squares	df	Mean Square	F	Sig.
<b>Questionnaire Effectiveness</b>	<b>Main Effects</b>	<b>(Combined)</b>	1.885	3	.628	.374	.772
		<b>Subject experiences delays in Game 1</b>	.775	1	.775	.461	.500
		<b>Type of questionnaire subject experiences</b>	1.096	2	.548	.326	.723
	<b>2-Way Interactions</b>	<b>Subject experiences delays in Game 1 * Type of questionnaire subject experiences</b>	1.461	2	.731	.435	.649
		<b>Model</b>	3.322	5	.664	.395	.850
		<b>Residual</b>	105.924	63	1.681		
		<b>Total</b>	109.246	68	1.607		
a Questionnaire_effectiveness by Subject experiences delays in Game 1, Type of questionnaire subject experiences							
b All effects entered simultaneously							

## ANALYSIS 15.

## Analysis of satisfaction levels, Game 2

ANOVA: Mean self-report of satisfaction levels immediately after game 2 (No significant result found)

## CASE PROCESSING SUMMARY(A)

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
69	98.6%	1	1.4%	70	100.0%
a SATG2 by Subject experiences delays in Game 1, Type of questionnaire subject experiences					

## ANOVA (A, B)

			Unique Method				
			Sum of Squares	df	Mean Square	F	Sig.
<b>SATG2</b>	<b>Main Effects</b>	<b>(Combined)</b>	9.288	3	3.096	.708	.551
		<b>Subject experiences delays in Game 1</b>	2.811	1	2.811	.642	.426
		<b>Type of questionnaire subject experiences</b>	6.647	2	3.324	.760	.472
	<b>2-Way Interactions</b>	<b>Subject experiences delays in Game 1 * Type of questionnaire subject experiences</b>	2.069	2	1.034	.236	.790
<b>Model</b>			11.421	5	2.284	.522	.759
<b>Residual</b>			275.652	63	4.375		
<b>Total</b>			287.072	68	4.222		
a SATG2 by Subject experiences delays in Game 1, Type of questionnaire subject experiences							
b All effects entered simultaneously							

ANALYSIS 16A.

### Coding directions for independent judges, part 1

Given to 3 independent judges to classify open-ended responses to questions on Exit Questionnaire

#### Directions for coding questions 1 and 2 of Exit Questionnaire

The answers you will be coding will be those to Questions 1 and 2 of the Exit Questionnaire. These questions are as follows:

1. How much did you enjoy playing the first game?
2. How much did you enjoy playing the second game?

Instructions: Go through the entire stack of questionnaires, only looking at Questions 1 and 2. Read both answers, compare them and rate which game the subject reported enjoying more on the upper section of the coding sheet. Please use the following coding scheme:

- 0** = Not enough information to make a determination.
- 1** = Subject enjoyed game 1 more than game 2.
- 2** = Subject enjoyed game 2 more than game 1.
- 3** = Subject enjoyed both games equally well.

The number on the coding sheet should correspond to the blue, circled number on the upper-right-hand corner of each Exit Questionnaire (the larger number between two smaller number codes). Responses can be on a sliding scale, so for example if the subject reported disliking or not enjoying one game, but liking the second game even less, then code the response as 1. Disregard any reasons *why* subject may have enjoyed one game more than another. If the subject left one or both answer spaces blank, then code the response as 0 (not enough information).

ANALYSIS 16B.

## Coding directions for independent judges, part 2

Given to 3 independent judges to classify open-ended responses to question on Exit Questionnaire

### Directions for coding Question 3 of Exit Questionnaire

The answers you will be coding this time will only be those to Question 3 of the Exit Questionnaire. This question is as follows:

3. Why did you play the second game for as long as you did?

Instructions: This time, go through the entire stack of questionnaires again, but only looking at Question 3. Read the response, and evaluate it according to the following rating/coding scheme:

**0** = Not enough information to make a determination, or entirely unrelated to any category below.

**1** = Game-related—criteria derived from the game itself or interest in the game, such as wanting to collect treasure, discover the underlying map, or other game-related factor (e.g. "I wanted to see what surprises there were", or "I just wanted to explore", or "I wanted to find out if there was a real map", or just "to find the diamonds!").

**2** = Time-related—playing until a specific time, for one reason or another (e.g., "I wanted to play more, but I need to eat." or "I thought from the directions that I was supposed to play for 20 minutes").

**3** = Score-related (e.g., I wanted to double my score").

**4** = Subject-related (e.g., "I am competitive" or "I enjoy playing games" or "I started to get mind-drubbingly bored out of my skull").

NOTE: If you see responses that seem to involve score AND time, such as "I wanted to see how many points I could score in the same amount of time", code it as #3, a scoring-related response. (Time-related responses have more to do with outside influences determining how long the subject played, rather than an internal decision or benchmark that the subject decides to use)

Bear in mind that the game involves manipulating a character to find and pick up various items of treasure, and that in the second game subjects are told that they can play for as little as 3 minutes or as long as 20. Where the subject responded with multiple reasons, please pick either the first one the subject mentions, or the one that seems most important (if you can clearly distinguish this reason from the other(s)).

Record your evaluation using the above codes on the bottom of the Coding Sheet, using the same large, circled reference number on the Exit Questionnaires that you used to judge and code Questions 1 & 2.

## ANALYSIS 17.

## Responses to On-Line, Open-Ended Question

Responses to open-ended question asked at the end of both VENTING and CASPER on-line, interactive questionnaires: "Please describe your experience."

**Responses from the VENT Condition (last question of interactive questionnaire)**

**NO DELAY/VENT Condition: Females ( $n = 5$ )**

- The most frustrating this is the slow speed. I like trying to figure out the best way to accumulate lots of points. The competition aspect adds a lot to that...
- It was too slow.
- NO RESPONSE
- It was generally a fun game, but was a bit too slow to fully enjoy.
- I wanted to know what the danger was in walking around picking up jewels. I wanted to know what the purpose was.

**NO DELAY/VENT Condition: Males ( $n = 6$ )**

- Difficult to say until I know how I did. Without a target or purpose I feel OK, but no strong feelings really. I would like to play some more though.
- Not much reaction except a bit of frustration in how slow the character walks and turns.
- Exploring was fun, but a slow and somewhat annoying process
- Indifference. Definitely not angry... a little impatient but not angry.
- Interesting to participate in an experiment. Five minutes was not enough time to develop strategy. I was still getting hang of controlling figure near walls and corners. I should have been bypassing more low value gems given the speed of the character at picking up items.
- I didn't find the game very frustrating, rather i found it sort of boring, largely because it only involved walking around and piking stuff up, not avoiding enemys or any other intelligence based tasks.

**DELAY/VENT Condition: Females ( $n = 6$ )**

- Frustrating- the character would not move when I tried to make her move. First thought, what am I doing wrong? But then, realized I had little control over the character, even though I had thought I would.
- I liked the game. Felt that I may have scored better if character moved more quickly with arrow movements. Also think it would help if you had a pre-diagram, even briefly of the treasure pieces placement.
- i wasn't done yet! Oh well.
- I thought that the game was pretty neat. It had decent graphics. The slow speed of her walking and the delays in which the timer was not paused was mildly annoying.
- I felt I was doing well, but the constant delays were very irritating. I would have been ok with the delays had the timer stopped when the game was stalled. At the end, I was right next to a red ruby thinger and had 10-15 seconds left to get it. I couldn't, because the game stalled. Grr.
- The slowness of the main character's movements was the most frustrating aspect of the game. Turning the character and waiting for it to pick up objects was also frustrating. One interesting part of the game was choosing which path to choose. It would have been nice if the character could run since there was a time limit and the object of the game was to pick up as many items as possible.

**DELAY/VENT Condition: Males ( $n = 6$ )**

- t was a bit frustrating that I knew my success was predicated upon being able to collect treasure, however the network delays were preventing me from doing so, as the timer was counting down while the server was being contacted. In addition, I think the game was rather slow. This certainly allowed the player time to think, yet it was slow enough, and five minutes so short a time, that I don't think it made any difference. In all, it was not hard to play an optimal game.
- It was fun; kinda interesting to look for the "treasures"; wish that I was able to see a bigger chunk of the screen so I had a better idea of where I was or where I was going. The delays waiting for the browser to load were pretty frustrating... I didn't want to wait for my character to move. She walked too slow too... not as responsive to my keystrokes.
- Overall, it was frustrating because it was like the network connection was hindering my efforts to achieve a high score.
- it was a fun game seeing it is an early version, but was quite simple and did have a lot of depth to its game play. Of course, it being a network game can somewhat compensate for that though there really isn't a whole lot of replay value to walking around and picking up rocks at a rather slow pace.
- I was unsure if I was playing the game correctly on first try. The lag time was quite frustrating. Since I was trying to get a high score and was racing against the clock, it made me feel powerless to excel since I was dependent upon the response time of the network to succeed.
- I did not mind playing it but I did feel frustrated and a little bit angry because of all the delays that happened and because of the delays all the time I lost when I could be getting more points.

Responses from the VENT Condition (last question of interactive questionnaire)

**NO DELAY/CASPER Condition: Females ( $n = 5$ )**

- it's really too slow.... Of course , I am not smart enough!! :)
- It wasn't that frustrating... sometimes I'd try and turn, and the person wouldn't turn, and I couldn't quite figure out where to stand to get the treasure. Also, she moves slowly, and sometimes I'd think I could go forward along the edge of the wall and the computer would disagree with me. If we're playing for large amounts of money, it might be nice to see a map. (Then again, that defeats the entire purpose of the game. Ah, well... it seems a touch luck-based at least playing the first time.)
- I think that I liked the game
- Wanting to do well but seemingly not able to do so since the graphical interface was so slow. Also got boring.
- I liked the fact that it was a woman and the instructions were easy to follow, although it took me a little while to get the hang of it.

**NO DELAY/CASPER Condition: Males ( $n = 6$ )**

- sort of indifferent, though the stops and stalls annoyed me
- The character is moving somewhat sluggishly. Because I have seen soem pretty fast attacks in other games that makes me feel somewhat helpless, like in a bad dream where one just can not move quite as fast as one would like to and so gets caught by something nasty. The game is very smooth, so maybe it would be worthwhile to consider a graphics that is less smooth but faster.
- It was a bit of fun, I felt like it took a while to get the hang of it.
- It's a cool game! It would be nice to have a map though, so I can play more strategically (i.e., pick up the objects which are worth the most points first).
- (NO RESPONSE)
- Character is pretty slow. I'd like the character be as fast as my thoughts. There is a lot of emptiness with nothing to do, not even scenery to look at. It might be because the game is first playable, without any sound or graphical filler.

**DELAY/CASPER Condition: Females ( $n = 6$ )**

- For a simple maze-running game, I'd expect speed comparable to a desktop game, so yes, this was pretty frustrating. I would be delighted to give up graphics for more speed.
- The game is so slow. When I got stuck, I felt a lillte annoyed and kick the keyboard heavily.....

- I feel tired of its delay. But it's still fun.
- Considering I was unable to move my character for periods of time, sometimes for at least 8 seconds, I ended up becoming more interested in the 5 minute countdown than I did the game itself. If the connection were smooth, and I didn't feel so interrupted, perhaps I would feel more interest in this game.
- I felt that I could do much better in this game if I considered more carefully of the steps, where and when should I turned. So I didn't get stuck too long. Thanks for the game, I really enjoy it .
- It was fun; kinda interesting to look for the "treasures"; wish that I was able to see a bigger chunk of the screen so I had a better idea of where I was or where I was going. The delays waiting for the browser to load were pretty frustrating... I didn't want to wait for my character to move. She walked too slow too... not as responsive to my keystrokes.

**DELAY/CASPER Condition: Males ( $n = 6$ )**

- It was interesting to note that the pauses seemed to become more frequent as I accumulated treasure. I noticed that as I was picking up the last of the 2 gold bars, she stopped for about 10 seconds to "think" about it. And also that she seemed to get to the last bar at 4:40 and not be able to pick it up in time. Oh, well.
- Although the game is a pretty good outline...having the delays while contacting the host affected my score a great deal. If it was a little more smoother, It is possible that I could have done a lot better.
- The game itself was pretty simple, or atleast it seemed simple, but I got a little frustrated with all of the delays. Other than that I'm curious as to what the point of the game is.
- Well, I got delayed by the computers loading for at least .5 of a minute.
- The game itself was enjoyable except the reaction times that the computer was stalled or just would not react.
- The game was interesting to play but there were constant delays that kept pausing the game. Watching the game stop every 30 seconds was annoying.



## REFERENCES

- Amsel 90:** Amsel, A. (1990). Arousal, Suppression and Persistence: Frustration Theory, Attention, and its Disorders. In *Special Issue: Development of Relations Between Emotion and Cognition. Cognition and Emotion*, Vol. 4 (3), Sept. 1990, pp. 239-268.
- Amsel 92:** Amsel, A. (1992). *Frustration Theory*. Cambridge University Press, Canada.
- Archer 97:** Archer, J. (1997). Why do people love their pets? *Evolution and Human Behavior* 18(4), July 1997, pp. 237-259.
- Averill 90:** Averill, J. R. (1990). Inner feelings, works of the flesh, the beast within, diseases of the mind, driving force, and putting on a show: Six metaphors of emotion and their theoretical extensions. In D. E. Leary, editor, *Metaphors in the History of Psychology*, Cambridge University Press, New York, pp. 104-132.
- Baudrillard 83:** Baudrillard, J. (1983). *Simulations*. Semiotext(e), New York.
- Baudrillard 88:** Baudrillard, J. (1988). *Selected Writings*(Mark Poster, Ed.) Stanford University Press, Stanford CA.
- Beck 96:** Beck, A. M., and Meyers, N. M. (1996). Health enhancement and companion animal ownership. *Annual Review of Public Health* 17, 1996, pp. 247-257.
- Biocca 95:** Biocca, F., and Levy, M. R. (Eds.), (1995) *Communication in the age of Virtual Reality*. Lawrence Erlbaum & Assoc., New Jersey.
- Bowlby 69:** Bowlby, J. (1969). *Attachment and Loss, Volume 1: Attachment. Second Edition*. Basic Books, New York.
- Breazeal 98:** Breazeal, C. (1998). Regulating Human-Robot Interaction Using “Emotions”, “Drives” and Facial Expressions. Presented at Autonomous Agents 1998 workshop

- "Agents in Interaction - Acquiring Competence through Imitation", Minneapolis/St.Paul, May 1998.
- Brennan 95:** Brennan, K. A., Shaver, P. R. (1995). Dimensions of adult attachment, affect regulation, and romantic relationship functioning. *Personality and Social Psychology Bulletin*, Vol. 21(3), March 1995, pp. 267-283.
- Cassidy 94:** Cassidy, J. (1994). Emotion Regulation: Influences of Attachment Relationships. *Monographs of the Society for Research in Child Development*, Vol. 59(2-3), 1994, pp. 228-283.
- Catanzaro 94:** Catanzaro, S. J., and Greenwood, G. (1994). Expectancies for Negative Mood Regulation, Coping and Dysphoria Among College Students. *Journal of Counseling Psychology*, Vol. 41(1), Jan. 1994, pp. 34-44.
- Cole 95:** Cole, K. M., and Gawlinski, A. (1995). Animal-assisted therapy in the intensive-care unit—A staff nurse's dream comes true. *Nursing Clinics of North America* 30(3), September 1995, pp. 529.
- Cooper 95:** Cooper, M. L., Frone, M. R., Russell, M. Mudar, P. (1995) Drinking to Regulate Positive and Negative Emotions: A Motivational Model of Alcohol Use. *Journal of Personality and Social Psychology*, Vol. 69(5), Nov. 1995, pp. 990-1005.
- Corsini 95:** Corsini, R. J., and Wedding, D., Eds. (1995). *Current Psychotherapies: Fifth Edition*. F. E. Peacock Publishers, Inc., Itasca, IL.
- Damasio 94:** Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. G. P. Putnam's Sons, New York.
- Edupage 98:** Edupage (author unknown), (1998). The Lonely Net. From *Edupage* [an Internet-based education technology news service], August 30, 1998. <http://webserv.educom.edu/edupage/98/edupage-0830.html>
- Ekman 92:** Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, Vol. 6, 1992, pp. 169-200.
- Elliot 92:** Elliot, C. (1992). *The Affective Reasoner: A process model of emotions in a multi-agent system*. Ph.D. thesis, Northwestern University, May 1992. The Institute for Learning Sciences, Technical Report #32.
- Felsman 87:** Felsman, J. K., and Vaillant, G. E. Resilient Children as Adults: A 40-Year Study. In Anderson, E. J., and Cohler, B. J. (Eds.) *The Invulnerable Child*. Guilford Press, New York.
- Fernandez 97:** Fernandez, R. (1997). *Stochastic Modeling of Physiological Signals with Hidden Markov Models: A Step Toward Frustration Detection in Human-Computer Interfaces*. MIT Masters Thesis, September 1997.
- Fischer 97:** Fischer, K. W., and Bidell, T. R. (1997). Dynamic Development of Psychological Structures in Action and Thought. In Lerner, R. (Ed. of volume), Volume 1: Theoretical Models of Human Development, in Damon, W. (Ed.), *Handbook of Child Psychology*, Wiley & Sons, New York.
- Folse 94:** Folse, E. B., Minder, C. C., Aycock, M. J., and Santana, R. T. (1994). Animal-assisted therapy and depression in adult college students. *Anthrozoos* 7(3), 1994, pp. 188-194.

- Forgas 87:** Forgas, J. P., and Bower, G. H. (1987). Mood effects on person-perception judgments. *Journal of Personality and Social Psychology*, 53(1), pp. 53-60.
- Forgas 87a:** Forgas, J. P., and Moylan, S. (1987). After the movies: Transient mood and social judgments. *Personality and Social Psychology Bulletin*, 13(4), pp. 467-477.
- Foucault 84:** Foucault, M. (1984). *The Foucault Reader* (Rabinow, P., Ed.) Pantheon Books, New York.
- Foucault 88:** Foucault, M. (1988). *Politics, Philosophy, Culture: Interviews and Other Writings 1977-1984*. Routledge, Chapman & Hall, Inc., New York, NY.
- Freud 48:** Freud, A. (1948). *The Ego and the Mechanisms of Defence*. Hogarth Press Ltd., London.
- Friedmann 95:** Friedmann, E., and Thomas, S. A. (1995). Pet ownership, social support, and one-year survival after acute myocardial-infarction in the Cardiac-Arrhythmia Suppression Trial (CAST). *American Journal of Cardiology* 76(17), December 1995, pp. 1213-1217.
- Frijda 86:** N. H. Frijda. (1986). The Emotions. *Studies in Emotion and Social Interaction*. Cambridge University Press, Cambridge.
- Gardner 83:** Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books, New York.
- Goffman 67:** Goffman, E. (1967). *Interaction Ritual: Essays in Face-to-Face Behavior*. Aldine Publishing Company, Chicago, IL.
- Goleman 95:** Goleman, D. (1995). *Emotional Intelligence*. Bantam Books, New York.
- Gordon 70:** Gordon, T. (1970). *PET: Parent Effectiveness Training*. Plume Books (New American Library), New York.
- Green 93:** Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, 64, pp. 1029-104.
- Gross 95:** Gross, J., and Muñoz, R. F. (1995). Emotion Regulation and Mental Health. *Clinical Psychology Science and Practice*, Vol. 2(2), Summer 1995, pp. 151-164.
- Ickes 97:** Ickes, W., Ed. (1997). *Empathic Accuracy*. Guilford Press, New York.
- Isen 87:** A. M. Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, 52(6), pp.1122-1131,
- Izard 93:** Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, 100(1), pp. 68-90.
- Jameson 90:** Jameson, F. (1990). *Signatures of the Visible*. Routledge, Chapman & Hall, Inc., New York, NY.
- Jameson 91:** Jameson, F. (1991). *Postmodernism (or, The Cultural Logic of Late Capitalism)*. Duke University Press, Durham, NC.
- Jenks 93:** Jenks, M. (1993). Ten Easy Steps to Defusing Anger. *Customer Relationship Management, Vol. 12, No. 2*, (1993). p. 36.

- Kagan 94:** Kagan, J. (1994). *The Nature of the Child: Tenth anniversary edition*. New York: Basic Books.
- Kahneman 73:** Kahneman, D. (1973). Arousal and Attention. Ch. 3 of *Attention and effort*. Prentice Hall, Englewood Cliffs, NJ. pp. 28-49.
- Kirsner 98:** Kirsner, S. (1998) Moody furballs and the developers who love them. *Wired*, September 1998, pp. 138 ff.
- Kitayama 94:** Kitayama, S., and Niedenthal, P. M., Eds. (1994). *Heart's Eye: Emotional Influences in Perception and Attention*. Academic Press, New York.
- Lang 90:** Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1990). Emotion, Attention, and the Startle Reflex. *Psychology Review*, Vol. 97, No. 3 (1990). pp. 377-395.
- Lawson 65:** Lawson, R. (1965). *Frustration: The Development of a Scientific Concept*. MacMillan, New York.
- LeDoux 94:** LeDoux, J. E. (1994). Emotion, memory and the brain. *Scientific American*, June 1994, pp. 50-57.
- Lewis 89:** V. E. Lewis and R. N. Williams. (1989). Mood-congruent vs. mood-state-dependent learning: Implications for a view of emotion. In D. Kuiken (Ed.), *Mood and Memory: Theory, Research, and Applications*, Volume 4 of Special Issue of the *Journal of Social Behavior and Personality*, (2), pp. 157-171.
- Marquis 94:** S. Marquis and C. Elliott. (1994). Emotionally responsive poker playing agents. In *Notes for the Twelfth National Conference on Artificial Intelligence AAAI-94 Workshop on Artificial Intelligence, Artificial Life, and Entertainment*. AAAI, 1994.
- Mayer 95:** Mayer, J. D., and Salovey, P. (1995). Emotional Intelligence and the Construction and Regulation of Feelings. *Applied and Preventive Psychology*, Vol. 4(3), Sum. 1995, pp. 197-208.
- Mehrabian 95:** Mehrabian, A. (1995). Theory and evidence bearing on a scale of Trait Arousability. *Current Psychology*, 14(1) Spring 1995. pp. 4-28
- Mesquita 92:** Mesquita, B., and Frijda, N. H. (1992). Cultural Variations in Emotions: A Review. *Psychological Bulletin*, Vol. 112(2), Sept. 1992, pp. 179-204.
- Metalsky 93:** Metalsky, G. L., Joiner, T. E., Hardin, T. S., & Abramson, L. Y. (1993). Depressive reactions to failure in a naturalistic setting: A test of the the hopelessness and self-esteem theories of depression. *Journal of Abnormal Psychology*, 102, 101-109.
- Morkes 98:** Morkes, J., Kernal, H. & Nass, C. (1998). Effects of humor in computer-mediated communication and human-computer interaction. In *Proceedings of the Conference of Human Factors in Computer Systems (CHI98 Summary)*, CHI 98, Los Angeles.
- Myers 89:** Myers, D. G. (1989). *Psychology, Second Edition*. Worth Publishers, New York.
- Myers 93:** Myers, D. G. (1993). *The Pursuit of Happiness: Who Is Happy and Why*. Avon Books, New York.
- Norman 86:** Norman, D., and Draper, S., Eds. (1986). *User-Centred System Design*. Lawrence Erlbaum Associates, Hissdale, NJ.

- Norman 88:** Norman, D. (1988). *The Psychology of Everyday Things* (a.k.a. *The Design of Everyday Things*). Basic Books, New York.
- Nugent 95:** Nugent, W. R., and Halvorson, H. (1995) Testing the Effects of Active Listening. *Research on Social Work Practice* 5(2), 1995, pp. 152-175.
- Oatley 94:** Oatley, K., and Duncan, E. (1994). The Experience of Emotions in Everyday Life. *Cognition and Emotion*, Vol. 8 (4), July 1994, pp. 369-381.
- Picard 97:** Picard, R. W. (1997). *Affective Computing*. MIT press, Cambridge, MA.
- Preece 94:** Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994). *Human-Computer Interaction*. Addison-Wesley, New York.
- Reeves 96:** Reeves, B. and Nass, C. I. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.
- Riseberg 98:** Riseberg, J., Klein, J., Fernandez, R., and Picard, R. W. (1998). Frustrating the User on Purpose: Using Biosignals in a Pilot Study to Detect the User's Emotional State. In *Proceedings of the Conference of Human Factors in Computer Systems (CHI98 Summary)*, CHI 98, Los Angeles.
- Salovey 90:** Salovey, P. and Mayer, J. D. (1990). Emotional Intelligence. *Imagination, Cognition and Personality* 9 (1990), pp. 185-211.
- Scassellati 98:** Scassellati, B. (in press). Imitation and Mechanisms of Joint Attention: A Developmental Structure for Building Social Skills on a Humanoid Robot. To appear as part of a Springer-Verlag series.
- Shneiderman 97:** Shneiderman, B. (1997). Direct manipulation vs. agents: Paths to predictable, controllable, and comprehensible interfaces. In Bradshaw, J. M. (Ed.), *Software Agents*. AAAI/MIT Press, Cambridge, MA.
- Strauss 97:** Strauss, S. (1997). How to make a robot smile. *MIT Technology Review*, Vol. 100(7), Oct. 1997, p. 14.
- Starr 94:** Starr, P. (1994). Seductions of Sim: Policy as a simulation game. *The American Prospect*, Spring 1994, pp. 19-29.
- Strommen 98:** Strommen, E. (1998). When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney. In *Proceedings of the Conference of Human Factors in Computer Systems*, CHI 98, Los Angeles.
- Tannen 90:** Tannen, D. (1990). *You Just Don't Understand: Women and Men in Conversation*. Balantine Books, New York.
- Tosa 96:** Tosa, N, and Nakatsu, R. (1996). Life-Like Communication Agent—Emotion-Sensing Character “MIC” and Feeling Session Character “MUSE”. *Proceedings of the 1996 International Conference on Multimedia Computing and Systems (Multimedia '96)*.
- Tomkins 62:** Tomkins, S. S. (1962). *Affect, Imagery, Consciousness, volumes 1 & 2: The Positive Affects and The Negative Affects*. Springer, New York.
- Turkle 95:** Turkle, S. (1995). *Life on the Screen*. Simon & Schuster, New York.
- Vaillant 77:** Vaillant, G. E. (1977). *Adaptation to Life*. Little, Brown, Boston.

- Weizenbaum 66:** Weizenbaum, J. (1966). A Computer Program for the Study of Natural Language Communication Between Man and Machine. *Communications of the Association of Computing Machinery* 9 (1966), pp. 36-45.
- Winer 91:** Winer, B.J., Brown, D. R., Michels, K. M. (1991). Statistical principles in experimental design, 3rd Edition. McGraw-Hill, New York.
- Winnecott 71:** Winnecott, D. W. (1971). Transitional Objects and Transitional Phenomena. Chapter 1 of *Playing and Reality*. Basic Books, New York.
- Zajonc 98:** Zajonc, R. B. (1998). Emotions. In Gilbert, D. T., Fiske, S. T., and Lindzey, G. (Eds.) *The Handbook of Social Psychology, Vol. 1* (4th Edition). pp. 591-632.
- Zasloff 94:** Zasloff, R. L. (1994). Loneliness and pet ownership among single women. *Psychological Reports* 75(2), October 1994, pp. 747-752.
- Zisselman 96:** Zisselman, M. H., Rovner, B. W., Shmuely, Y., and Ferrie, P. (1996). A pet therapy intervention with geriatric psychiatry inpatients. *American Journal of Occupational Therapy* 50(1), January 1996, pp. 47-51.

## BIOGRAPHICAL NOTE

Jonathan Klein completed his Masters degree in Media Arts and Sciences working with Professor Rosalind W. Picard at the MIT Media Laboratory in Cambridge, MA. His two-year tenure at the Media Lab was spent working on Professor Picard's Affective Computing project, a multi-disciplinary research effort that aims to enable computational media to recognize, model, synthesize, and respond to human emotions. Mr. Klein's work involves developing and testing computational systems that account for, and support, the emotional experience of users.

Mr. Klein received a Bachelor of Science degree with highest honors and high distinction in Human-Computer Interaction and Computer Graphics, from the University of Michigan in Ann Arbor. There, he worked with Professor Elliot Soloway to develop interaction design solutions for a number of major research efforts in educational technology, collaborative research media, virtual world interfaces, digital library front-ends, and multi-user games for the Web. He also spent 12 years in industry as a professional art director, graphic designer and illustrator.

His research interests also include designing and building emotionally expressive capabilities into computer systems (see the Bruzard project, at [http://affect.www.media.mit.edu/projects/affect/AC\\_research/projects/bruzard.html](http://affect.www.media.mit.edu/projects/affect/AC_research/projects/bruzard.html)); using machine learning to build conversational speech interface agents that adapt to users' conversational styles (see Mr. Klein's TurnStyles project, at [http://affect.www.media.mit.edu/projects/affect/AC\\_research/projects/turnstyles.html](http://affect.www.media.mit.edu/projects/affect/AC_research/projects/turnstyles.html)), and extending the research and practice agenda of the HCI community to include support for a broad range of human needs.

Mr. Klein, who himself has been frustrating people for years, is married and has a 5-month-old daughter. He and his family currently live in Arlington, MA.

