Affective Learning Companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance

by

Winslow Burleson

B. A. Physics, Rice University (1995)
M. S. E. Mechanical Engineering, Product Design (1997)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 2006

Author..
Program in Media Arts and Sciences,
School of Architecture and Planning
August 9, 2006

Certified by ....
Rosalind W. Picard
Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by.......................................
Andrew B. Lippman
Chairman
Department Committee on Graduate Students

Affective Learning Companions:
strategies for empathetic agents with
real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance
by
Winslow Burleson

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on August 18, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

This thesis has developed an affective agent research platform that advances the architecture of relational agents and intelligent tutoring systems. The system realizes non-invasive multimodal real-time sensing of elements of user's affective state and couples this ability with an agent capable of supporting learners by engaging in real-time responsive expressivity. The agent mirrors several non-verbal behaviors believed to influence persuasion, liking, and social rapport, and responds to frustration with empathetic or task-support dialogue. Pilot studies involved 60 participants, ages 10-14 years-old, and led to an experiment involving 76 participants, ages 11-13 years-old, engaging in the Towers of Hanoi activity. The system (data collection, architecture, character interaction, and activity presentation) was iteratively tested and refined, and two "mirroring" conditions were developed: "sensor driven non-verbal interactions" and "pre-recorded non-verbal interactions". The development and training of the classifier algorithms showed the ability to predict frustration/help seeking behavior with 79% accuracy across a pilot group of 24 participants.

Informed by the theory of optimal experience (Flow) and a parallel theory of a state of non-optimal experience (Stuck), developed in this thesis, the effects of "affective support" and "task support" interventions, through agent dialogue and non-verbal interactions, were evaluated relative to their appropriateness for the learner's affective state. Outcomes were assessed with respect to measures of agent emotional intelligence, social bond, and persuasion, and with respect to learner frustration, perseverance, metacognitive and meta-affective ability, beliefs of one's ability to increase one's own intelligence, and goal-mastery-orientation. A new simple measure of departure dialogue was shown to have a significant relationship with the more lengthy and explicit social bond Working Alliance Inventory survey instrument; its validity was further supported through its use in assessing the social bond relationship with other measures. Over-estimation of the duration of the activity was associated with increased frustration. Gender differences were obtained with girls showing stronger outcomes when presented with affect-support interventions and boys with task-support interventions. Coordinating the character's mirroring with intervention type and learners' frustration was shown to be important.

Thesis Supervisor: Rosalind W. Picard
Title: Professor of Media Arts and Sciences

Affective Learning Companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance

Winslow Burleson

Thesis for the degree of Doctor of Philosophy
at the
Massachusetts Institute of Technology

September, 2006

Thesis Advisor — Rosalind Picard
Professor of Media Arts and Sciences
MIT Media Laboratory

Thesis Reader — Mitchel Resnick
Professor of Media Arts and Sciences
MIT Media Laboratory

Thesis Reader — Kenneth Forbus
Professor of Computer Science and Education
Professor of Education
Electrical Engineering and Computer Science
Department Northwestern University

Acknowledgements

I would like to thank my advisor, Rosalind Picard for her strong support, inspiring advice, and wonderful guidance. Thank you for being my advisor and for doing a fantastic job! Thanks to my committee members Mitchel Resnick and Kenneth Forbus for providing direction, clarity, and enthusiasm for my work. Thanks to my mentors Ted Selker for your friendship, guidance, and for cultivating my inventive abilities; thanks to Teresa Amabile for welcoming me to your research group and for providing exciting challenges to advance my creativity; thanks to Dan Ariely for nurturing in me critical analytical skills; thanks to Ken Perlin for your engaging collaboration in the development of the affective agent research platform, and thanks to Issac Colbert for your support throughout my time at MIT.

I would like to thank many contributors, supporters, and collaborators in my work, particularly Nalini Ambady, Jeremy Bailenson, Tim Bickmore, Carol Dweck, Mark Lepper, Maryln Welsch, the members of the Amabile Research Group, the researchers and innovators at Deutsche Telekom's T-Labs, and my supporters at TU Berlin. I would also like to thank the Things That Think consortium and NSF, Grant No. 0087768 and 0325428.

Not only could my work not have been done without the members of the Affective Computing Group and the Context Aware Computing Group, but they have also been a great team of friends and colleagues to work with. I thanks the many contributors to the system, the theory, the experiment and the dissertation, particularly Ashish Kapoor, John Rebula, David Merrill, Sumi Ariely, Rob Reilly, Barry Kort, Devin Neal, Louis-Philippe Morency, Carson Reynolds, Phil Davis, Karen Liu, Marc Strauss, Salene Mota, Dorothy Warner, Stephen Ornes, Brittany Shoot, Anna Pandolfo, Susan Spilecki and the MIT Writing Center and many UROPs.

I thank the educators, schools, centers, and the 100 plus parents that provided their children and resources to advance this research. I particularly thank Rob Reilly who hosted me and opened many school doors. I thank the Charlestown Computer Clubhouse and Danielle Martin, the Wellesley Middle School and John D'Auria, the Salvation Army in South Boston, the Lanesborough School, Richmond Consolidated School, and St Mark School.

I thank my many friends for their moral support, particularly Brian Little, Katherine Isbister, Andrea Lockerd, David Merrill, Tammy Chan, Elizabeth Churchill, Chris Verplaetse, Paul Nemirovsky, Elin Sjursen, Elisabeth Sylvan, Ernesto Arroyo, Magda Ismail, Carolin Young, and many, many others.

For their life long support, love, and inspiration, I thank my sister Betsy Burleson, my parents, Nan and David Burleson., and my fiancée Camilla Nørgaard Jensen and her family.

Table of Contents

List of Figures

List of Tables

1 Introduction

Imagine you are trying to solve a problem using a computer, and while you are deeply engaged, you are trying different possible solutions, curious about which will work. None of them have worked yet, but you are enjoying trying different things. The whole time there is a pedagogical agent quietly present in the background, who had introduced itself earlier as "Casey", and seems friendly, but has mostly left you alone, which is fine with you. After a while, however, you start to feel like you are out of ideas, and there is nothing left to try. You seem to be going backwards instead of forwards, and the little bits of failure and frustration you were ignoring earlier are starting to add up. You feel stuck. You look over to the agent, who now has a caring expression and you ask it for help. It is supportive, and soon you are eagerly engaged again, trying a different strategy that you hadn't tried before.

Today's agents tend to not know when to help and when to leave you alone – they may be "intelligent" about math or science, but they are largely ignorant when it comes to reading cues from you about whether it is a good time to interrupt you or to leave you alone. The research described in this thesis is aimed at building pedagogical agents that don't just see whether or not you are making mistakes, but also see how you seem to be doing while engaged (or not) in the learning experience. The research is enabling them to see the difference between when you are frustrated or stuck and may be ready to quit, or when you are curiously exploring and engaged in the task. A platform has been built that allows sensing many aspects of this information, and testing different strategies for responding to it.

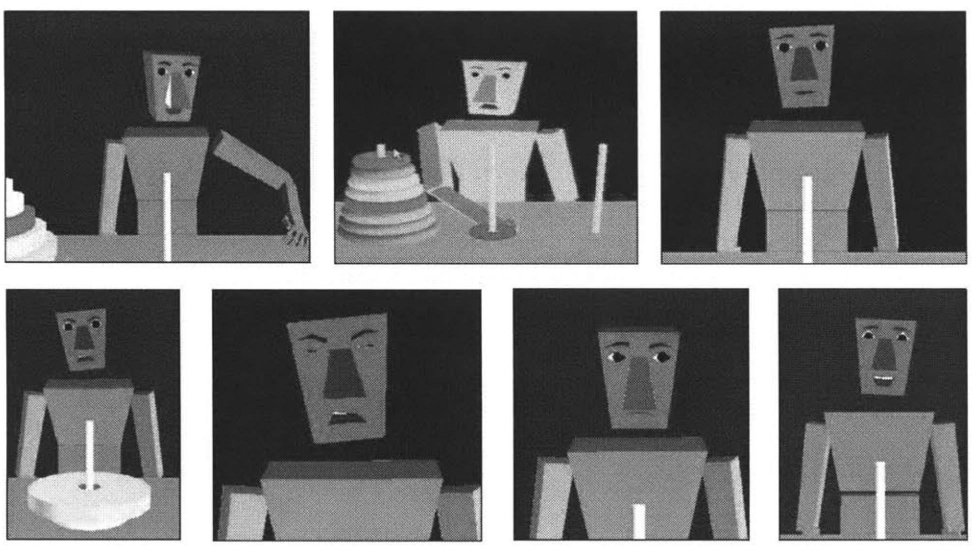


Figure 1.1 Affective Learning Companions are capable of a wide range expressions (The agent shown here was co-developed with Ken Perlin and Jon Lippincott at NYU.)

Through the sensing of learners' emotional states and the appropriate expression of empathy, the pedagogical agents are beginning to have the ability to create caring relationships with their learners. Caring and empathy are used by expert human tutors to keep their students engaged. It is hypothesized that caring and empathetic expressions (see Figure 1.1.) can be employed in the relationship between a pedagogical agent and student and that developing an agent with these abilities will improve students' learning.

This research focuses on the role of feelings in learning. It uses an affect-focused approach to supporting learners by sensing their level of frustration and assessing the appropriateness of task based vs. affect based interventions. Most of today's educational technologies have been developed for particular domains (math, reading, science, etc…). Ultimately it is expected that affective approaches to supporting learners, similar to those explored in this theses, will be used in conjunction with the existing domain approaches. It is also expected that the affect skills developed by learners will be transferable across domains, as people develop affective strategies that can be diversely applied.

Several research steps are needed to demonstrate the importance of the affect-focused approach. A fist step, at the core of this thesis, is to contrast affective strategies with task based strategies. Once progress on the first step has been achieved, a second step would be to demonstrate that the benefits of affective approaches are transferable across domains. This second step is not the focus of this thesis and remains as future work. The affective strategies implemented in this thesis are investigated by developing and demonstrating the affective abilities of pedagogical agents and assessing the impact of these agents on learners.

Pedagogical agents that have the ability to detect learners' emotional states can respond differently when learners are bored or interested. An agent that could recognize some common negative states could also teach learners how to use an awareness of these states to guide them in their learning. Such agents could, for example, help learners realize that a feeling of frustration doesn't mean that they are bad at this challenge, but rather that they can use this feeling as a signal that it may be time to try a different strategy: Frustration can thus be converted into a reminder to find a different route. Researchers have found that over half of the interactions that expert tutors have with their students are focused on emotion and engagement (Lepper, Woolverton et al. 1993).

Affect interacts with cognition and part of this research aims to illuminate more of this interaction. Consider cognitive beliefs about intelligence. What students believe about their own intelligence has been shown to have a dramatic impact on the way they approach challenging tasks (Dweck 1999). If they believe, as 40% of learners do, that their intelligence is a personal "trait", that is fixed and can not be increased, then they tend to quit at the first sign of difficulty. These students who hold "trait beliefs" prefer to do tasks that show others that they can perform well. If, however, students hold "incremental beliefs", if they believe they can increase their intelligence, as another 40% do, then they tend to meet challenges enthusiastically and to perceive them as learning opportunities. Fortunately for those students that quit prematurely, learners' beliefs about their intelligence can be changed.

Theories of metacognition were introduced by Flavell to describe how learners can use strategies and self-awareness to improve their thinking processes (Flavell 1976; Efklides and Vauras 1999). These theories describe metacognition as having three elements: metacognitive knowledge, metacognitve experience, and metacognitve skill. Metacognitive knowledge is what you know about how cognition works, e.g., "If I take a different perspective, than my current one, I may learn more about the problem." Metacognitive experience is an awareness of your thinking, e.g., "I am thinking in a muddled way compared to my usual clear thinking." Metacognitive skill is the ability to coordinate metacognitive knowledge and metacognitive experience, e.g., the ability to recognize that you don't understand something and to apply a metacognitive knowledge strategy to improve your understanding.

This thesis defines meta-affect as comprising three things, meta-affective knowledge, meta-affective experience, and meta-affective skill. First, meta-affective knowledge is knowledge about how affect works, e.g., "Frustration can arise when a goal that matters to you is impeded." and "Frustration has value for causing you to stop the current approach and consider alternative approaches". Second, meta-affective experience is an online awareness of your feelings and what action tendencies they are provoking in you, e.g., "I feel like quitting; more specifically, I feel frustrated and that is making me feel like quitting." Meta-affective experience is more than awareness of a current emotion, it is a conscious reflection on what that emotion is doing to you, or may do to you. Thus it is more than consciously feeling an emotion (emotional experience). Third, meta-affective skill is the ability to coordinate meta-affective knowledge and meta-affective experience, e.g., Instead of quitting because of feeling frustrated, calming oneself and thinking of alternate strategies, perhaps including seeking help, until the problem is resolved. This thesis describes the development and evaluation of an Intelligent Tutoring System that has as one of its primary goals facilitating learners' metacognitive and meta-affective skills.

The system used in this research is beginning to be able to detect elements of learners' feelings, such as when they are frustrated. Drawing upon technologies that have been developed by members of the Affective Computing Group over the past decade the system developed in this thesis can predict, with 79% accuracy, whether a learner will quit at a given time (Kapoor, Burleson et al.). While this prediction is limited to a specific activity, setting, duration of activity, and requires that several sensors function reliably (various sensors do not work well when users are wearing glasses or jewelry or when the day is hot or the individual is sweaty) it is adding to the understanding of learners and their affective states. A supportive interaction with a pedagogical agent might help learners, especially those that are predicted to quit, perceive this moment as an opportunity to apply their metacognitve knowledge and develop their meta-affective skill, all in the midst of a real learning experience.

***Flow*: Optimal Experience** *- Csikszentmihalyi*	***Stuck*: A State of Non-Optimal Experience** *- Burleson*
All encompassing	**All encompassing**
A feeling of being in control	**A feeling of being out of control**
Concentration and highly focused attention	**A lack of concentration and inability to maintain focused attention**
Mental enjoyment of the activity for its own sake	**Mental fatigue and distress caused by engagement with the activity**
A distorted sense of time	**A distorted sense of time** *Weybrew 1984; Czerwinski et al. 2001*
A match between the challenge at hand and one's skills	**A *perceived* mismatch between the challenge at hand and one's skill**
Frequently associated with positive affect	**Frequently associated with negative affect**

Table 1.1. Elements of Flow and Stuck juxtaposed

The development of the pedagogical agent's interactions with learners has been guided by the theory of Flow: Optimal Experience and has lead to the development of a complementary theory of Stuck: a state of Non-optimal experience, that describes elements of the feeling of frustration during learning activities; see Table 1.1. (Csikszentmihalyi 1990; Burleson and Picard 2004). Flow experiences occur when students are fully engaged in a challenge matched to their skill. These moments are not a good time for a pedagogical agent to interrupt the learner, as an intervention at this time might disrupt the flow experience and learning. Stuck experiences are times when learners are overwhelmed by the task at hand and perceive that they do not have the skills to pursue the challenge. These are moments when a supportive intervention might facilitate a change in the learner's perspective. An interaction that helps them to realize that this is a learning opportunity, rather than a time to quit, can help learners adopt metacognitive strategies. In effect, these interventions can help learners become better learners.

To sense learner's emotions and enable pedagogical agents to respond with appropriate behaviors the Affective Agent Research Platform has been developed (Figure 1.2.) (Burleson, Picard et al. 2004). The Affective Agent Research Platform consists of a real-time scriptable character agent capable of a wide range of expressive interactions together with the ability to sense affective information from the learner. To detect the learner's emotions and to inform this character's behavioral interactions, a set of sensors is used. These sensors have been developed over the past several years and validated in a variety of experiments by the Affective Computing Group (http://affect.media.mit.edu/publications.php). The pressure mouse detects the intensity of the user's grip on the mouse, which has been shown to correlate to frustration. The skin conductance sensor is a well established indicator of user arousal. Recent experiments with the seat posture chair show that it can be used to classify motivational states such as engagement, boredom, and break-taking. The facial-expression camera can measure head nod/shake, mouth fidgets, smiles, blink events, and pupil dilations. These sensors provide data to a system that comprises a data logger, system server (that coordinates the other elements), classifier, behavior engine, and character engine.

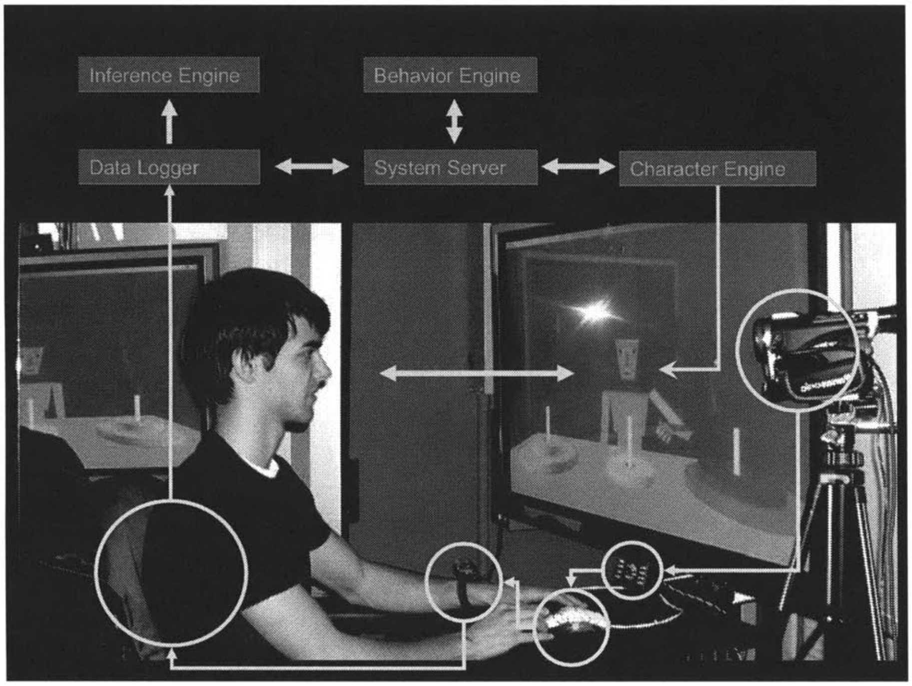


Figure 1.2. Affective Agent Research Platform

The system uses the sensor readings in real time to generate serendipitous interactions that are just-in-time responses to user's affect. When a student smiles the character can chose to smile back. This responsiveness is possible because of algorithms in the behavior engine, and because of dynamically scripted character attributes at multiple levels. The system also has sequences of fixed behaviors, a behavioral repertoire; incorporated in these sequences are variables which can be updated in real-time. This approach is particularly suited to affective expression.

Affective expression in human-to-human interactions can be enhanced by non-verbal social mirroring, a kind of imitation of another's nonverbal cues, where the imitation occurs after a 3-4 second delay so that it is not irritating. Non-verbal social mirroring has likewise been shown to be productive in creating agents that are more likable and persuasive (Bailenson, Beall. et al. 2005). The system sensors and dynamic scripts enable the pedagogical agent to engage in non-verbal social mirroring in more extensive ways than have been enabled before. This thesis investigates how the mirroring of the learner influences social bond and persuasiveness of communication.

In addition to the non-verbal social mirroring the sensor readings are used offline to train a classifier that predicts help seeking based on measuring several parameters believed to be related to frustration and other relevant affective states. This classifier can later be used in a real-time way with learners. New machine learning techniques have contributed to the interpretation of the affective states of learners: interest, boredom, break taking, arousal, frustration, and quitting. The classifier can be used to help determine when the character should provide explicit interventions. Figure 1.3. shows the timeline of a learning task from t=0 at the beginning until task completion. In determining when to provide an intervention, most technologies today base this decision only on the game state, which can include information such as the level of difficulty, number of moves, the timing of moves, and the amount of progress the student is making. The game state may provide some indication of when to intervene, but the timing might not be optimal, as learners could be making similar mistakes while in a flow state or in a stuck state. In this system the sensors allow us to detect information related to affect such as frustration, which can help distinguish flow from stuck. If the sensors detect frustration around the same time as the game state information indicates an intervention is needed then the game state algorithms are sufficient for these individuals. This is shown by the box in the figure below. However, individuals that hold "trait beliefs" frequently quit at the first sign of difficulty. Sensors in this system try to detect these individuals' frustration, and may be able to provide better timing for an intervention, possibly keeping these individuals from quitting. Alternatively, those that hold "incremental beliefs", who believe they can increase their intelligence, are eager to pursue challenges. A premature game-state interruption might deprive them of the opportunity for independent learning and flow experiences. If however this same individual then needed additional encouragement later, the affectively aware system would again be better positioned to facilitate effective responses. Through continuous awareness of a participant's learning patterns and affective states (Stuck or Flow) a pedagogical agent with an affective classifier should be better able to help a learner than one guided by pre-programmed game state responses.

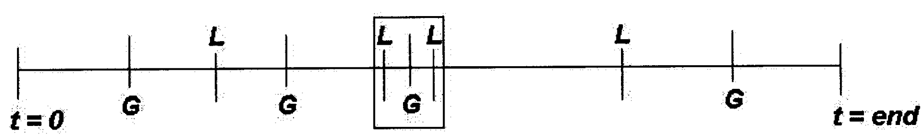


Figure 1.3. Using sensors the classifier algorithms can tell the system to intervene at times of user frustration, indicated by the letter "L" on the time line above. Without sensors, systems only have game-state information to determine when to intervene, indicated by the letter "G".

Summary

The Affective Agent Research Platform is capable of multi-sensor logging and generation of affective response by pedagogical agents to children's interactions in real-time. This allows the Affective Learning Companions to engage in non-verbal social mirroring, which has been shown in prior work to play a beneficial role in human-human interactions. This research has overcome many challenges: developing and testing a complex research platform, synchronizing information from multiple streams of sensors together with the learning task, calibrating task difficulty and training the classifier algorithms with data from 24 learners. The research has calibrated the task difficulty and obtained data from learners that indicate they are frustrated or need help and from those that persevere, so that the system predicts frustration and needing help vs. perseverance behavior with 79% accuracy (tested on 24 learners) (Kapoor, Burleson et al.). An experiment was then conducted with seventy six children aged 11-13 to test non-verbal social "mirroring" and two types of interventions (*affect support* and *task support*), for their impacts on learners' perseverance, frustration, social bonding, and motivation.

Affective sensing and machine learning are leading to the refinement of algorithms that can detect and classify an increasing range of learners' affective states. This platform is being used to explore several affective findings in the social, behavioral, and learning sciences. This research is applying these resources to help learners persevere through Stuck, while developing metacognitive and meta-affective skills to become better learners.

2 Background and Theory

This chapter will discuss background research as it relates to this thesis. The chapter will also articulate a new theory of Stuck: non-optimal experience, which has evolved over the course of this thesis. The included literature starts with a discussion of educational technology and educational psychology as these relate to learning and failure in educational activities. The discussion then turns to the distinct elements of metacognition and meta-affect and their roles in developing learners' self awareness and enabling them to regulate the learning process, using meta-affective skills. Research on self theories of intelligence will be discussed as it relates to learner's engagement in challenging learning opportunities. The match of challenge with skill will be related to the theory of Stuck, which parallels the theory of Flow: optimal experience, as a means for understanding learner's experience and the impact of that experience on learning. The discussion of Stuck and Flow will be put into context with an example of the challenges presented by the Towers of Hanoi activity. The role of affective interactions by human tutors and teachers with their students will be discussed as they relate to strategies for interaction and student learning outcomes. Some of the limitations of experiments in education and psychology will be discussed with respect to the ability of human experimenters to reliably observe their participants and to control their own affective expressions. Human-agent interactions, in terms of verbal, non-verbal, and affective sensing capabilities will be discussed as a technological means for overcoming some of these limitations. Human-agent affective interactions, such as the ones employed in this research, will be presented as a tool for contributing to learners' metacognitive and meta-affective skills (defined in chapter 1) to persevere through frustrating learning experiences.

2.1 Enabling better responses to failure

Failure is important to learning and instrumental to the development of multiple points of view required for deep understanding. In order to accommodate exploratory failures, Papert incorporated debugging - testing and fixing things that have gone wrong – as a central element of his constructionist method (Papert 1980). This iterative process may play a role in developing the multiple points of view that Minsky suggests are important to understanding. Minskey says you never understand anything until you understand it in more than one way (Kay 1991). Stanford's Department of Mechanical Engineering has saying that encourages learning from one's failures: "Spectacular failure is better than moderate success." (Faste 1996) This is not an overtly masochistic agenda; the message is that not striving for spectacular success will achieve little, and if moderate success is attained, not enough has been desired or accomplished. Kay's version of this sentiment explains: "Difficulty should be sought out, as a spur to delving more deeply into an interesting area. An education system that tries to make everything easy and pleasurable will prevent much important learning from happening." (Kay 1991) This idea parallels Csikszentmihalyi's notion of matching adequate challenge with skill in service of Flow: optimal experience (Csikszentmihalyi 1990), discussed later in this chapter.

In this vein, in, "Motivation and Failure in Educational Systems Design," Roger Schank and Adam Neaman describe the utility of simulated Learning By Doing environments in accelerating the pace of learning through frequent exposure to difficult circumstances that arise infrequently in real world situations. This exposure will inevitably accelerate the rate of failure and, if motivation is sustained, the rate of learning as, "novices are exposed to rare, but critical, experiences" (Schank and Neaman 2001). Schank and Neaman acknowledge that fear of failure is a significant barrier to learning and believe this can be addressed in several ways: minimizing discouragement by lessening humiliation; developing the understanding that consequences of failure will be minimal; and providing motivation that outweighs or distracts the unpleasant aspects of failure. They show that they have been able to sustain the motivation of learners, who care about what they are doing, by providing them access to experts at the time of failure. Through questions, stories, anecdotes and additional experiences, learners are given the opportunity to, "expend the effort to explain their failures". Learners are given the opportunity to achieve and become expert (Schank and Neaman 2001). Many have taken the approach of tailoring the task to the individual user in an effort to maintain motivation or increase the opportunity for Flow (Malone 1981; Monk 2000; Hill, Gratch et al. 2001). In their paper, "Toward the Holodeck," Hill et al describe a Holodeck-like setting that they constructed and discuss their evaluation of the merits in terms of its immersive, believable, and motivating qualities. This terminology is remarkably similar to descriptions of psychological Flow.

While this task manipulation approach seems appropriate in some circumstances, it runs the risk of missing an important opportunity for users to learn affective self-awareness and to engage in related metacognitive and meta-affective skill development. In other circumstances it is simply not possible. The task may not lend itself to modification or modularization, it may be unbounded, or its analysis and manipulation may be beyond the current capabilities. Proposed instead is an affective approach. Regardless of task state information or activity tailoring, it is possible to gather information about the user's affective state and to use this to guide affective interactions on behalf of the user. In many circumstances, tasks are harder than anyone would like them to be. And, these tasks may not always tailor themselves to facilitate motivation. Considering these conditions from a pedagogical perspective, it is important to take into account the phenomenon of "negative asymmetry;" that the staying power of negative affect tends to outweigh the more transient experience of positive affect (Giuseppe and Brass 2003). Unfortunately, for the purposes of sustaining learners' motivation, negative asymmetry means that negative affect experienced from failure and repeated failure is likely to persist disproportionately to the positive affect experienced from success. A separate and compounding factor is that it is often easier for novices to see their failures than to realize their successes. One approach is for educators and innovators to try harder to create motivating learning environments which celebrate achievement and provide sustaining inquiry opportunity at times of failure, such as those advocated in other studies (Monk 2000; Hill, Gratch et al. 2001; Schank and Neaman 2001). This approach does not address how to help learners become better at coping with and regulating negative feelings associated with failure. In contrast, the proposed affective approach is to provide tools and experiences that foster affective self-awareness, skills for dealing with failure, frustration, and other forms of negative affect, as well as skills for identifying and benefiting from incremental success. These would have the goal of developing within users a greater metacognitive and meta-affective skill to control their own motivational engagement with difficult tasks, despite recent failures and difficult tasks. Metacognition and meta-affective skill will be further discussed in section 2.2.

Kort, Reilly and Picard propose a model of constructive learning that relates learning and emotion in an evolving cycle of learner affective states. Their proposed cyclic trajectory begins with anticipation, expectation, and exploration, a stage where intervention is to be discouraged. If the learner progresses to disappointment or discouragement and stays there too long, then intervention may be productive. They argue that this cycle, including its negative states, is natural to the learning process, and that learners can develop skills to keep moving through it, propelling themselves out of the failure mode and into a more hopeful state conducive to continued exploration and learning (Kort, Reilly et al. 2001). If learners become aware of their own cyclic trajectories, as they encounter challenges and persevere through failure, they can more readily develop metacognitive and meta-affective skills to overcome failure.

2.2 Metacognition, Meta-Affective Skill, and Self Theories of Intelligence

Theories of metacognition were introduced by Flavell to describe how learners can use strategies and self-awareness to improve their thinking processes (Flavell 1976; Efklides and Vauras 1999). These theories describe metacognition as having three elements: metacognitive knowledge, metacognitve experience, and metacognitve skill. Metacognitive knowledge is what you know about how cognition works, e.g., "If I take a different perspective, than my current one, I may learn more about the problem." Metacognitive experience is an awareness of your thinking, e.g., "I am thinking in a muddled way compared to my usual clear thinking." Metacognitive skill is the ability to coordinate metacognitive knowledge and metacognitive experience, e.g., the ability to recognize that you don't understand something and to apply a metacognitive knowledge strategy to improve your understanding. Metacognitive skills make use of both metacognitive knowledge and metacognitive experience and allow the individual to self regulate their cognitive processes through planning, monitoring of progress or process, and effort allocation.

This thesis defines meta-affect as comprising three things, meta-affective knowledge, meta-affective experience, and meta-affective skill. First, meta-affective knowledge is knowledge about how affect works, e.g., "Frustration can arise when a goal that matters to you is impeded." and "Frustration has value for causing you to stop the current approach and consider alternative approaches". Second, meta-affective experience is an online awareness of your feelings and what action tendencies they are provoking in you, e.g., "I feel like quitting; more specifically, I feel frustrated and that is making me feel like quitting." Meta-affective experience is more than awareness of a current emotion, it is a conscious reflection on what that emotion is doing to you, or may do to you. Thus it is more than consciously feeling an emotion (emotional experience). Third, meta-affective skill is the ability to coordinate meta-affective knowledge and meta-affective experience, e.g., Instead of quitting because of feeling frustrated, calming oneself and thinking of alternate strategies, perhaps including seeking help, until the problem is resolved. This thesis describes the development and evaluation of an Intelligent Tutoring System that has as one of its primary goals facilitating learners' metacognitive and meta-affective skills to enable learner's, who experience frustration, to implement strategies that will help them persevere and continue to pursue learning opportunities.

This thesis seeks to develop an Intelligent Tutoring System that facilitates learners' development of metacognitive and meta-affective skills. Many researchers working on metacognition express the difficulties of conducting experiments in this domain (Flavell 1979; Efklides 2002). It is particularly difficult to access affective and cognitive aspects of individuals' internal thought processes. In spite of this, Carol Dweck's work on self theories of intelligence presents promising findings for understanding why learners fail and how to help them succeed. She has found that individuals' beliefs of their own intelligence profoundly affect their motivation, learning, and behavioral strategies, especially in response to their perception of failure (Dweck 1999). This research has identified two predominant groups of individuals: "incrementalists," who believe their own intelligence can be enhanced, and "trait learners," who believe their intelligence is largely fixed. She has found that when instrumentalists fail at a task, they tend to increase their intrinsic motivation for the task, believing that if they try harder, they will get better and smarter. When trait based individuals fail, they exhibit avoidance and decreased intrinsic motivation for the task, believing instead that their previous performance defines their ability. They act on their desires to avoid further confirmation of what they perceive to be their "trait based" inability. She has developed a simple strategy of metacognitive knowledge, a strategy for thinking about thinking: the strategy is to think of "the mind is like a muscle and through exercise and effort you can grow your intelligence." By adopting this simple strategy people can shift their self theories of intelligence and alter their goal/mastery orientation (increasing their mastery orientation).

2.3 Flow and Stuck

An individual's perception of the challenge at hand and how well their skills match it can determine whether they experience Flow or Stuck. Flow is the theory of optimal experience and Stuck is a new theory, proposed by this thesis research, of a state of non-optimal experience frequently encountered during frustrating learning activities. When a student is experiencing productive engagement with a task, it is likely to be a positive experience. In order to better understand when a student is experiencing unproductive engagement with a task, the elements of negative affective states have been directly compared with the elements of Flow. Flow is a feeling of being in control, a state of concentration and highly focused attention, mental enjoyment of the activity for its own sake, a distorted sense of time, and a match between the challenge at hand and one's skills. In direct contrast, elements of negative affect in learning include: feelings of being out of control; a lack of concentration and inability to maintain focused attention; mental fatigue and distress caused by engagement with the activity; and a negative distorted sense of time. Frustrating tasks seem to last longer then they actually do, and they consequently tax endurance. Several researchers have experimentally linked participants' over estimation of an activity's duration with self-reported levels of frustration (Zeigarnik 1967; Weybrew 1984; Czerwinski, Horvitz et al. 2001; Liu and Picard 2005). The non-optimal learning experience, described by the above listed attributes, is commonly called "stuck" (Table 2.1).

Awareness of one's affective state can influence a person's ability to alter that state. Conscious awareness of Flow is an interruption during which the meta-affective feelings of awareness tend to diminish happiness and the sense of optimal experience (Csikszentmihalyi 1990). The feelings of awareness serve as interruptions external to Flow. This relationship between the feelings of awareness, interruption, and Flow suggests a similar relationship may exist for Stuck. If this relationship exists then, feelings of awareness serve as interruptions external to Stuck. These interruptions may assist users in mitigating the detrimental effects of Stuck on their learning. The detrimental effects of Stuck are the feelings of frustration such as "aarrghh; this isn't working" and hopeless interrupting thoughts like "I can't do this." These interruptions are feelings of difficulty, internal to the Stuck experience. They are not the external feelings of awareness necessary to interrupt the Stuck experience. Developing in learners the ability to trigger the feelings of awareness external to Stuck when they are experiencing feelings of difficulty internal to Stuck can cultivate in these learners the ability to interrupt and persevere through Stuck. Just as learners can use their own feelings of awareness to interrupt Stuck, they can also benefit from interruptions from others (including intelligent tutoring systems). These interruptions of Stuck are productive and probably more welcome then external interruptions during Flow.

Since repeated failure is critical to deep understanding and developing expertise, mitigating the effects of Stuck is especially important in light of the previously discussed phenomenon of negative asymmetry. People who learn to persist through frustration and avoid Stuck are better able to deal with failure. By becoming better at failing, they become better at learning.

The experiences of Flow, Stuck, and phenomena of negative asymmetry, along with the benefits of affective self-awareness in mitigating Stuck, each have a fundamental impact on learning. Research on motivation shows that it is intimately intertwined with failure, frustration, affect, awareness, and emotional intelligence. The current understanding of these psychological effects informs the development of affective agents that behave with appropriate affective interactions with learners.

***Flow*: Optimal Experience** *- Csikszentmihalyi*	***Stuck*: A State of Non-Optimal Experience** *- Burleson*
All encompassing	**All encompassing**
A feeling of being in control	**A feeling of being out of control**
Concentration and highly focused attention	**A lack of concentration and inability to maintain focused attention**
Mental enjoyment of the activity for its own sake	**Mental fatigue and distress caused by engagement with the activity**
A distorted sense of time	**A distorted sense of time** *Weybrew 1984; Czerwinski et al. 2001*
A match between the challenge at hand and one's skills	**A *perceived* mismatch between the challenge at hand and one's skill**
Frequently associated with positive affect	**Frequently associated with negative affect**

Table 2.1. Elements of Stuck juxtaposed Flow

In the field of Intelligent Tutoring Systems there is a distinction between, on the one hand, adjusting the environment or task to facilitate Flow, and on the other, empowering the user through self-awareness to participate in self-regulated motivational strategies. Many Intelligent Tutoring Systems choose to adjust the challenge level to keep the learner engaged. If a learner is frustrated, these systems will make the activity easier. In contrast, through social interactions, we choose to help individuals tailor self-perceptions of their ability with respect to a challenge. If learners are able to alter their perception of failure and negative affect, then they may be able to mitigate the detriments of negative asymmetry. This may enable them to persevere and succeed at greater challenges. With an increased awareness of their negative affect, learners might be better equipped to reengage in challenging learning experiences in the future.

Towers of Hanoi

To study Flow and Stuck in individuals, it is productive to select a task that can elicit both challenging engagement and frustrating responses from participants. Also, given that this work emphasizes the affective aspects of learning it is productive to choose a domain where the cognitive aspects are already well studied. The large body of research literature on the Towers of Hanoi (see Figure 2.1) activity and its variants has informed the use of this activity in this study. The Towers of Hanoi activity consists of three poles with disks stacked on the first pole. The goal is to move the disks from the first pole to the third pole. There are two rules: first, a larger disk may not be placed on a smaller one; second, only one disk may be moved at a time. As a learning scenario, since Towers of Hanoi is recursive, it presents an important opportunity for repeated failure and recovery. It has also been the subject of considerable mathematical and psychological study. One focus of the research has been on child development; in particular, on an individual's executive functioning (i.e. decision making processes), inhibition, and problem solving strategies (Simon 1975; Welsh, Satterlee-Cartmell et al. 1999; Bishop, Aamodt-Leaper et al. 2001; Espy, Bull et al. 2004).

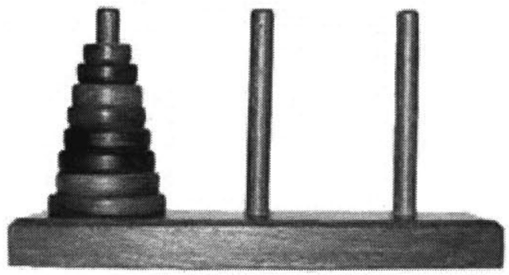


Figure 2.1. Tower of Hanoi

The extensive literature on this activity made it an appealing choice of a learning activity, but since my research is centered on the affect of individuals my approach goes beyond the scope of existing research. A frequent concern in developing educational experiences is the transferability of the skills learned from one domain to another. Skills linked to an individual's awareness of affect and to his or her response to affect, are a primary concern of this research. These skills are expected to be more generally transferable than domain based skills. Towers of Hanoi is an activity that is potentially challenging and/or frustrating. This attribute of the activity allows for the investigation of individuals affective responses and use of metacognitive strategies with respect to their Flow and Stuck experiences. It is possible to compute at any given instant exactly how far a learner is from the solution (in terms of the minimum number of moves left), giving an objective measure of "performance" which facilitates the emphasis on the affective experience (Rueda 1997).

2.4 Affective Elements of Tutoring

Social relationships have been shown to be influential in many learning situations. They can help learners develop responsibility and help children adopt mastery orientation as an approach to challenges (1981; Strain 1981; Webb 1987; Tudge and Caruso 1988). Caring relationships have also been shown to be related to academic performance (Wentzel 1997). The encoding of affect within human-human interactions is very powerful. In their research on "thin slices," Ambady and Rosenthal demonstrate that when participants in their studies are shown a short segment of video, as little as six seconds of a teacher's first interactions with their student, they can predict teacher effectiveness and student end of term grades (Ambady and Rosenthal 1993). Thin slices research provides evidence that some aspects of relational studies may not require longitudinal study since the formative elements of relationships can be very strong from the outset.

They also found that humans can not fully control their own expressions. When they applied the thin slices approach to the court room, they found that judges' beliefs about the guilt of a defendant are inadvertently transmitted affectively in their brief, "unbiased" instructions to the juries (Blanck, Rosenthal et al. 1990). The implication of this "emotional leakage" is that in the study of human-human interactions, human confederates may introduce a great deal of variance. These confederates have inherent limitations in their abilities to *reliably* observe their participants and to respond with their own affective expressions, as a *controlled* experimental treatment. This limitation becomes particularly daunting when experimenters desire to investigate multiple elements of complex human-human interactions. Learning interactions, which are strongly influenced by nonverbal communication, demand new experimental approaches to confront this limitation.

Knowing when to provide affective support and when to provide domain relevant knowledge is a tutoring skill that can facilitate learning (Higgins 2001; Baker, Corbett et al. 2004). Approximately 50% of expert tutors' interactions with their students are affective in nature (Lepper, Woolverton et al. 1993). For example, at times of Stuck, expert tutor interactions can be empathetic and remind a student of Dweck's strategy (the mind is like a muscle); at times of Flow, interactions should support but not interrupt the learner's experience. Dweck's work on self theories of intelligence uses interventions at a group level, and generally provides them prior participant's engagement in an activity. This research applies Dweck's metacognitive strategy to provide individually tailored interactions to help individuals improve their meta-affective skill i.e. to use their frustration as an indication of a learning opportunity. This approach uses an Affective Learning Companion to provide interactions based on a learner's affective states (Flow and Stuck). These affective human agent interactions can be *reliable* and *controlled* across experimental participants.

2.5 Affective Human Agent Interaction

One of the objectives of an Intelligent Tutoring System is to have its agents engage in affective interactions with users. Existing agent systems typically infer human affect by sensing and reasoning about the state of a game or an outcome related to an action taken by the user within the computing environment. Use of such an approach is illustrated by the pedagogical agent COSMO, who applauds enthusiastically and exclaims "Fabulous!" if the student takes an action that the agent infers as deserving of congratulations (Lester, Towns et al. 1999). There are learning situations in which this reaction would be warmly received and perhaps reciprocated with a smile by the user and situations where it would not.

While reasoning based on a user's direct input behaviors is important and useful, it is also limited. For example, COSMO has no ability to see how the user responds nonverbally to its enthusiasm. COSMO is unable to tell, for example, if the user beamed with pride or frowned and rolled her eyes, as if to say that COSMO's response was excessive or otherwise inappropriate. If the latter, it might be valuable for COSMO to acknowledge its gaffe, thus making it less likely the user will hate it or ignore it in the future. Thus, there is a need and a desire to advance agent capabilities to include perceptual sensing of nonverbal affective expressions together with the channels that are traditionally sensed in interactive agent systems.

Through the use of verbal, non-verbal and affective interactions, Intelligent Tutoring System agent technologies are getting better at realizing real-time multi-modal affective interactions between agents and learners. However, to date, there are no examples of agents that can sense natural (both verbal and non-verbal) human communication of emotion and respond in a way that rivals that of another person.

Verbal Dialogue

A simple way to support verbal interactions is through pre-generated text-to-speech dialogue. This allows users to respond to the agent by clicking on multiple-choice text responses. This type of asymmetric text and voice based interaction, while not as ideal as symmetric natural language processing system, has been shown to be effective for relationship building, in the context of a physical trainer agent (Bickmore 2003).

Characters that use text-to-speech voices sound mechanical and impersonal. In general, children's voices for text-to-speech are less natural sounding than adult voices. Researchers are working on making text-to-speech sound more natural and to incorporate affect into these speech models (Tao and Kang 2005). To date, the speech models that generate affective speech are cumbersome to use.

In terms of scripting affective dialogue to help learners deal with frustration, a lesson can be learned from Klein et al.'s "frustration-handling" agent. This agent was a text based agent that provided a series of interactions that first assessed and confirmed users' levels of frustrations, then empathized with them. This research showed that the active listening dialogue strategy used led to lower levels of frustration (Klein, Moon et al. 2000; Klein, Moon et al. 2002). More sophisticated models of general discourse have been developed in the BEAT and Collagen systems. Real time interaction has been exhibited in the BEAT System (Cassell, Vilhjálmsson et al. 2001), which focuses on dynamic generation of appropriate physical rapport (body language) synchronized with verbal discourse. The Collagen system's discourse models emphasize the importance of immediacy and responsiveness (Rich, Sidner et al. 2001). The control and dynamic generation of appropriate behavior and speech is an ongoing focus of the development of Intelligent Tutoring System agents (Lester, Callaway et al. 2001; Shaw, LaBore et al. 2004).

Non-Verbal Communication in Agent Systems

Most computational agents show expressive behaviors, often via facial movements or various gestures. Affective expressions are thought to be useful to help make agents "believable" (Bates, Loyall et al. 1992). Expressive behaviors have additionally been associated with useful outcomes such as making agents likeable (Koda and Maes 1996; Elliott, Rickel et al. 1999; Lester, Towns et al. 1999; Johnson, Rickel et al. 2000). In more recent systems, agent expressions have been responsive to human expressions, contributing to making agents "relational" (i.e. able to construct long-term social-emotional relationships with users). For example, the Laura agent, expanding on the empathetic "frustration-handling" agent of Klein et al (Klein, Moon et al. 2000; Klein, Moon et al. 2002), received verbal (text-only) expressions of a variety of affective states from the user. Laura responded with both verbal and nonverbal expressions of empathy based not only on the user's current expression, but also on previous expressions (Bickmore and Picard 2004). Thus, the agent could respond to the statement "I'm feeling down" by moving closer to the user, displaying a facial expression of concern, and speaking an appropriate verbal response such as "Sorry to hear that." If day after day a user continued to indicate these feelings, Laura's wording would change to acknowledge the ongoing problem. If things escalated, the agent would refer the user for medical help.

Using non-verbal communication to create and assess social bonds has been the focus of several research efforts. To assess users' impressions of the Laura agent interactions in terms of caring, trust, and likeability Bickmore used a survey instrument, the modified Working Alliance Inventory (see Appendix B for the modified Working Alliance Inventory used in this thesis). Bull found that he could measure and predict the quality of the relationship in human human interactions by studying social mirroring; i.e., the pattern of symmetrical movements of one individual in response to another individual (Bull 1983). Bailenson has used social mirroring in his work on Transformed Social Interactions. He has shown that an agent that employs mirroring of the user's head motions is perceived as more persuasive and likable than one that displays pre-recorded head motions (Bailenson and Yee 2005). He used a four second delay so that users would not disturbed by overt mirroring. Human human social mirroring goes beyond visual attributes. Marci and Gottman have found that measurements of skin conductance from couples in counseling sessions can be used to predict the strength of their relationships and divorce (Gottman and Levenson 1992; Marci and Riess 2005).

Recognizing and responding to affective information is a vital part of natural intelligent interaction. These two skills are widely recognized as components of so-called "emotional intelligence" (Salovey and Mayer 1990; Goleman 1995). If an agent cracks a joke, and Bobby smiles while Cynthia frowns, then it would probably be fine for it to flash a smile back at Bobby, while the same expression back at Cynthia might be perceived as mean. Depending on the agent's goals, one response could be much more intelligent than another. If an agent winks and does a cute little dance that irritates you, and if it repeats that little dance and you show increased irritation and perhaps visible anger at it, then it might be wise for it to be able to see your response and subsequently to act in a way that acknowledges its failure, that is if the goal includes wanting you to have a favorable impression of it. How someone chooses to respond to your emotion greatly colors your opinions of their competence, trustworthiness, likeability, and more.

Human-human interaction is not the same as human-computer interaction, nor is there a need to limit the development of systems to this objective. However, results from studies of human-human interaction can usefully inform the design of human-computer interaction (Reeves and Nass 1996; Moon 2000). Another example, in addition to social mirroring, that illustrates the importance of an appropriate affective response is provided by the work of Robinson and Smith-Lovin (Robinson and Smith-Lovin 1999), who described how if a person responds positively to something bad happening, then that person will be less liked. Alternatively, if a person responds in a way that is affectively congruent, then that person will be more liked. These findings seem to support the current approach in pedagogical agent research where the character smiles when you succeed and looks disappointed if you make a mistake or fail.

However, it is also known that pre-school subjects smile as much after failure as they do after success (Schneider and Josephs 1991). Human expression is not necessarily affectively congruent with the task, so it is unclear what the agent expression should be if it is to be perceived empathetically. In human-human interactions, it is important to realize that there are several different conditions for smiles, including nervousness, humor, and success. It is expected that such interactions will also hold true for human-agent interactions. However, to achieve deeper understanding of such interactions requires new advances, namely the development of technology that can recognize and respond in real time to affect.

2.6 Affect and Learning in Intelligent Tutoring Systems:

Embodied conversational agents are capable of developing trusting and beneficial relationships with humans and sharing combined physical and virtual space with children. Some examples of this include helping children develop literacy skills (Bickmore 2003; Ryokai, Vaucelle et al. 2003) and developing a personal physical trainer that gives users the impression that it cares about them about its users (Bickmore and Picard 2004). The ability to recognize and respond to affective information in an empathic way plays a role in these relationships. It has been argued that being attuned to the child's emotional state through affective sensing will be important to the development of Intelligent Tutoring Systems and learning companions (Picard 1997; Conati 2002; Adcock and Van Eck 2003).

Several researchers, including Tak-Wai Chan who coined the term "Learning Companion" with respect to Intelligent Tutoring Systems, promote the idea of presenting agents as peer-learning companions (Chan and Baskin 1990; Aimeur, Frasson et al. 2000). One rational for this is that peer-tutors might be effective role models because they are less likely to invoke anxiety in learners; learners may believe they can attain the same level of expertise as their tutors (they may not believe that they can attain an adult teacher's level of expertise).

In his work on the development of animated agents for learning, Lester demonstrated a phenomena he has termed the persona effect. This effect is the positive affect response and engagement shown by users when interacting with an animated agent (Lester, Converse et al. 1997). The persona effect is essentially a social presence effect.

The strategy of pedagogical adaptation that is social, in that it is mindful of the learners' affective state, has been pursued in the work of several researchers. In their work on interactive pedagogical drama for health interventions, Marsella and Johnson use affective elements of textual dialogue to inform and adapt their agents with the goal of altering user affective states through changes in their perspective rather than in the task (Marsella and Johnson 2003). Aimeur's provoking agents employ this strategy through cognitive dissidence: the agent companions serve as troublemakers to perturb the affective state of individuals. When a learner is comfortable with the status quo, the troublemakers introduce problems that are aimed at generating pedagogically productive insights by the learners (Aimeur, Frasson et al. 2000). Baylor has found motivational differences based on appearance, attire, gender, and race and that providing users a choice of agents impacts their motivation and learning (Baylor, Shen et al. 2003; Baylor, Shen et al. 2004).

It is widely acknowledged that creating an Intelligent Tutoring System is challenging and complex (Lester, Towns et al. 1999; Johnson, Kole et al. 2003). One common strategy for animated character development is to create a pre-defined behavioral space, a set or sequences of behavioral actions that can be elicited by the system at appropriate times (Johnson, Rickel et al. 2000; Lester, Callaway et al. 2001). To address the challenge of developing animated characters that have rich behaviors and expressions, Perlin has implemented a strategy he calls scriptable agents. His agents behave under the guidance of adaptive algorithms and can therefore exhibit a rich dynamic behavioral repertoire (Burleson, Picard et al. 2004). Perlin's work on facial expression has generated characters capable of nuanced facial expressions which display a continuous range of affect. These 3D face models have been used to help children with Autism to learn how to accurately recognize human facial affect. This work demonstrates that characters can display affective expressions that are recognizable by children (Brandt 2004).

Another central issue in the development of an agent system and pedagogical interaction is the parameters for agent interactions with learners. Lepper, du Boulay, Johnson, Luckin, have reviewed and suggested several strategies for a variety of sophisticated interaction guidelines based on human human interactions and human tutoring and teaching tactics (Lepper, Woolverton et al. 1993; du Boulay and Luckin 2001; Johnson, Kole et al. 2003). In personal conversations with Dweck and Lepper each supports the approach of developing an Intelligent Tutoring System that uses affective interactions in response to learner's affect to help learners apply the "the mind is like a muscle and through effort you can grow your intelligence" metacognitive strategy (Dweck 2004; Lepper 2004).

An emerging approach to Intelligent Tutoring System development is to assess and attend to learners' affective states with the assistance of sensors. Some agents are beginning to have interactions that are influenced by algorithms that couple data from affective sensors with embedded models of politeness and helping strategies. Prendinger has shown that characters that attend to cultural norms for politeness and that use skin conductance can provide calming interventions that lower user's arousal, Conati's systems have been focused on using skin conductance and task analysis to contribute to development of formal models of affect (Ortony, Clore et al. 1988; Conati and Maclaren 2004). There is a large gap between the information structured in the formal models and the information needed to guide an affective agent's interactions. An alternative approach that is complementary is to develop rich arrays of real time multi-modal affective sensors. One project that takes this approach is the AutoTutor project at the University of Memphis (D'Mello, Craig et al. 2005). This project incorporates a posture chair, facial expression camera, and conversational cue analysis to inform agent interactions with college students. Once a correlation is made between learners' affective states and the sensor values this correlation can inform agent human interactions; agents can become responsive to learners' affect in real time.

2.7 Sensing Affect

Affect can be expressed in many ways — not just through voice, facial expressions and gestures, but also through the adverbs of any aspect of the interaction. Affect modulates how a learner types and clicks, what words are chosen and how often they are spoken. It also impacts how a learner fidgets in his or her chair and how he or she moves his or her head and facial muscles. In the development of an affective Intelligent Tutoring System the most promising approach is to integrate many channels of information in order to better understand how affect is communicated.

Physiological and affective sensors (including sensors that measure heart rate, skin conductance, elements of respiration, blood oxygen levels, pressure exerted on a mouse, posture in a chair, gait analysis, brain oxygen levels, etc.) are emerging as new technologies for human agent interactions (Picard 1997; Allanson and Fairclough 2004). There are many challenges to the design and use of multi-modal affective sensors. For ease of use and natural interactions, it is desirable to have systems that are not intrusive to users and which do not require any or extensive training. Individual sensors also have signal-to-noise issues and robustness issues. Then there are the reactions of the users to the sensors, in terms of ethical issues, privacy, and comfort. Some researchers have found that children do not find a skin conductance sensor to be intrusive; in fact if kids are "deprived" of the opportunity to use the sensor they often feel that they have missed out on part of the experience of the interaction (Conati 2004).

In continual use over several days, multimodal affective sensing has been used to classify eight basic emotional categories in individuals with 81% accuracy (Healey 2000). Research incorporating the posture analysis seat (Mota and Picard 2003) and the Blue Eyes camera (Haro, Essa et al. 2000; Kapoor and Picard 2003) to conduct multimodal affect recognition has classified engagement, boredom and break taking behavior with over 86% accuracy (Kapoor and Picard 2005). While the development of classifier algorithms is an ongoing research topic, pattern recognition with multi-modal sensors has been shown to be an effective strategy in the development of affective sensing (Kapoor, Picard et al. 2004).

3 Design of an Affective Agent Research Platform

A new platform for affective agent research has been developed. The platform has a modular architecture that is facilitated by the System Server. The platform integrates an array of multi-modal affective sensors that send information to the Data Logger. A real-time Behavior Engine and Character Engine are used to present a 3d scriptable expressive humanoid agent within a graphical virtual environment. The platform also uses classifier algorithms to detect elements of user's affective experience. The research platform and architecture focus on the sensing and analysis of signals related to affect, and on the ability to interpret and respond to these, in real-time, with an expressive scriptable agent. The Behavior Engine and Character Engine include dynamically scripted character attributes at multiple levels. This approach is particularly suited to affective expression. This platform can be used to explore several affective findings in the social, behavioral, and learning sciences.

The user sits in front of a wide screen plasma display. On the display appears an agent and 3d environment. The user can interact with the agent and can attend to and manipulate objects and tasks in the environment. The chair that the user sits in is instrumented with a high-density pressure sensor array and the mouse detects applied pressure throughout its usage. The user also wears a wireless skin conductance sensor on a wristband with two adhesive electrode patches on their hand and forearm. Three cameras in the system, a video camera for offline coding, and the blue-eyes camera, record and sense additional elements of human behavior.

This multi-modal approach to recognizing affect uses more than one channel (e.g. facial expression alone) to sense a broad spectrum of information. This approach applies techniques from psychophysiology, emotion communication, signal processing, pattern recognition, and machine learning, to make a classification from this data. Since any given sensor will have various problems with noise and reliability, and will contain only limited information about affect, the use of multiple sensors should also improve robustness and accuracy of classification.

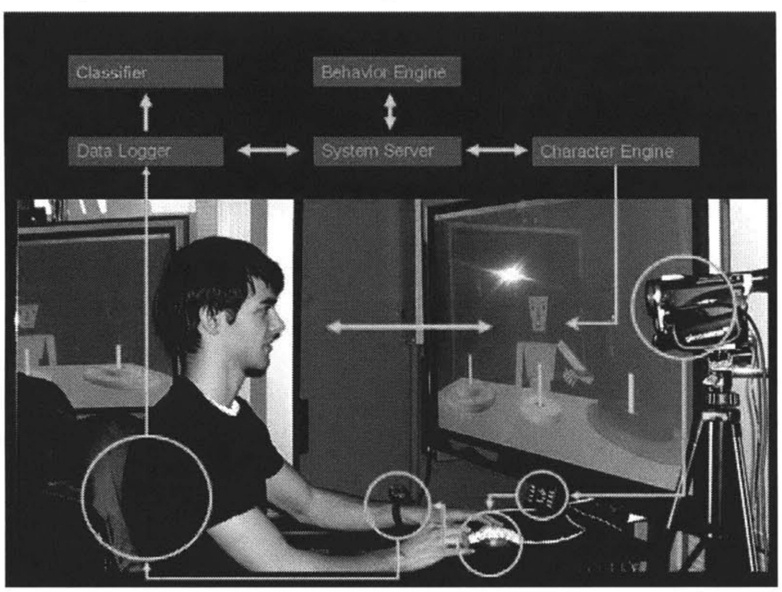


Figure 3.1. System Architecture with sensors listed from right to left: video camera, Blue Eyes camera, pressure mouse, skin conductance sensor, posture chair.

3.1 Sensors

The multi-modal sensor system consists of a Pressure Mouse, a Wireless BlueTooth skin conductance sensor (Strauss, Reynolds et al. 2005), a Posture Analysis Seat, a Facial Action Unit analysis using the Blue Eyes camera system, and Head Tracking. This system expands upon the earlier work (Kapoor, Mota et al. 2001), that used only facial and postural information. Through the combination of all these modalities, the agent system is provided with a better understanding of the affect and interactions of the user and is also able to determine the contribution of each of the sensors to the modeling of affect (Kapoor, Burleson et al.; Kapoor, Picard et al. 2004).

Pressure mouse

The Pressure Mouse has eight force-sensitive-resisters that capture the amount of pressure that is put on the mouse throughout the activity (Reynolds 1999). Users who have been administered a frustration inducing online application form have been shown to produce increasing amounts of pressure related to their level of frustration (Dennerlein, Becker et al. 2003).

Wireless BlueTooth skin conductance

In collaboration with Gary McDarby, at Media Lab Europe, Carson Reynolds and Marc Strauss, at the MIT Media Lab, developed a wireless version of an earlier "glove" that senses skin conductance. While the skin conductance signal does not explain anything about valence – how positive or negative the affective state is – it does tend to be correlated with arousal or how activated the person is. High levels of arousal tend to accompany significant and attention-getting events (Boucsein 1992).

Posture analysis seat

The Posture Analysis Seat utilizes the TekScan sensor pad system developed for medical and automotive applications (Tekscan 1997). The system uses pattern recognition techniques while watching natural behaviors to "learn" what behaviors tend to accompany states such as interest and boredom (Mota and Picard 2003). The system thus detects the surface-level behaviors (postures) and their mappings during a learning situation in an unobtrusive manner so as not to interfere with the natural learning process. Through the chair, significant detection of nine static postures and four temporal patterns associated with levels of learner interest has been demonstrated.

Blue Eyes camera system

Kapoor and Picard (Kapoor and Picard 2003) have been developing automatic tools for computer vision and machine learning that are capable of detecting facial movements and head gestures used as conversational cues and communications of emotion. The system currently detects some upper facial features such as eyes and eyebrows, as well as their motion and action: eyes squinting or widening, eyebrows being raised, and head nods and shakes. These techniques are being extended to include lower facial features like cheeks and the mouth, which express smiles, fidgets, and tension. The data logging includes full frame synchronized capture of the Blue-Eyes (Haro, Essa et al. 2000) camera images at 20 Hz giving the opportunity to code for additional facial action units as they are identified.

Head Tracking

The Head Tracking System (Morency, Rahimi et al. 2003; Morency, Sundberg et al. 2003) is built upon the Small Vision System developed by SRI International and the MEGA-DCS stereo camera (Videre Design 2004). This system also incorporates a real-time head nod and head shake algorithm (Kapoor, Mota et al. 2001) and provides information on the intersection of the user's gaze and the screen plane. This plane can be shifted to various reference depths within the environment to ascertain the virtual object that a user is directing their head toward. This type of sensing helps to facilitate shared attention behaviors.

This simulation used a wide plasma screen that provides greater spatial resolution between objects. This caused users to move their head to a greater extent than they would on a smaller screen and to attend to different objects and points of interest. This facilitates the use of the head tracker.

The head tracker employed proved to be unreliable. After less than 10 minutes of interaction it would fail to reacquire the position of the users head. Therefore it was not used with particiants in the studies.

Video capture

The video camera recorded the user and the onscreen activity. It was positioned so as to acquire both an image of the user and an image of the screen that is reflected in a mirror positioned behind the users head. This system was chosen so as not to miss any of the features of the user/character interaction and provide true (same image) synchronization. When the system is initialized, a datagram signal is sent to start the DirectX video capture and the time is noted in the log.

Game state

While game state is not a traditional sensor, it is used by the system as a source of data and is treated as a sensor channel in a manner similar to each of the other sensors. The system records the disk state after each move, checks if it is legal or illegal, increments the move count, calculates the optimal number of moves to the end of the game (Rueda 1997), and evaluates progress in terms of number and significance of regressions. This data can also be used to explore users' engagement and intent: understanding of the game, proceeding in a focused way, or becoming disengaged.

3.2 System Architecture

The system has several modules: a System Server, Sensors, Data Logger, Behavior Engine, Character Engine, and Classifier, see Figure 3.1. above. Each sensor sends its signal via UDP packets through a socket to the Data Logger. The use of UDP packets, because of their low overhead, facilitates the acquisition of multiple real-time packets and synchronized timestamps. Upon receipt of a UDP packet a time stamp is generated and the data is stored in a local text file. (This text file is currently submitted off-line to the Classifier.) Figure 3.1. shows the flow of aggregated UDP packets from the sensors to the Data Logger which are then forwarded to the

System Server. This forwarding uses TCP packets, to ensure their receipt. The System Server parses the data and forwards, via TCP packets, relevant data to the Behavior Engine. The Behavior Engine uses algorithms and thresholds to decide when and how to direct the character's actions. A bidirectional arrow in Figure 3.1. indicates that information from the Behavior Engine is also passed back via the TCP socket to the System Server. The Behavior Engine's decisions are then forwarded by the System Server to the Character Engine. The Character Engine controls the character actions, elements in an OpenGL virtual environment (disks and slideshow display). This virtual environment also enables the Character Engine to monitor the user's mouse clicks and disk movements. A Second bidirectional arrow in Figure 3.1. indicates that the Character Engine passes data from these monitored events to the System Server. The System Server in turn also forwards relevant mouse event and disk movement data to the Behavior Engine. All information that is received by the System Server is time stamped and logged in a text file for analysis (see Appendix D for further information on this process). As shown in the System Specification (Table 3.1.) the system comprises several machines and languages which exhibit differing data rates, and benefit from many contributors. Each module is connected with UDP or TCP sockets. The system architecture is modular allowing additional modules to be added in a similar manner.

Machines	Contributors	Platforms	Languages	Events/ Second
Seat Sensor	Tekscan/ SteelCase Inc.	Win 98	C++	10 packets sent
Blue Eyes	IBM Research/ Media Lab	Linux	C++	30 frames captured 8 packets sent
Skin Conductance	Media Lab/ Media Lab Europe	Win 2000	Python	11 packets sent
Pressure Mouse	Media Lab	Win 2000	Python	11 packets sent
Character Engine	Ken Perlin/ NYU	Win ME	Java/C++	30 updates
Data Logger/ System Server/ Behavior Engine	Media Lab	Win 2000	Java	~100 packets received
Video Capture	Media Lab	Win 2000	C++	30 frames captured

Table 3.1. System Specification, each packet contains 10-120 ASCII characters that represent the sensor values.

3.3 Behavior Engine and Character Engine

On a wide screen plasma display a 3D virtual environment is presented. This OpenGL Virtual Environment can display a character, virtual disks, selectable text and buttons, and images for a slide show presentation. The system can also play audio files. While direct user driven interaction with the environment occurs only through the pressure mouse, this information along with each sensor's information is used by the Behavior Engine to determine the behavior of the character and virtual environment (e.g. resetting the disks if the user makes an illegal move).

While the Behavior Engine determines the behavior, the animation of the character is managed by the Character Engine through the use of Character Behavior Scripts which contain two types of events, scripted events and serendipitous events. Scripted events are sequences of text that explicitly tell the character what actions to undertake. They are predetermined events which can be called upon by the system to elicit specific interactions. For example, in Table 3.2. the script on the left instructs the character to move its mouth for several seconds with a broad smile. Longer sequences can be scripted to expand the behavioral repertoire to include the introduction of the character to the user, a slide show presentation, and the delivery of precisely controlled *affect support* or *task interventions* (e.g. supportive comments when classification made from sensor data indicates they might be beneficial to users). Examples of *affective support* and *task support* dialogues appear later in this chapter; additional dialogues are found in Appendix C.

In contrast to the predetermined scripted events, serendipitous events are real-time interactions driven by sensors and algorithms. In Table 3.2. the script on the right instructs the character to move its mouth for several seconds. Initially the character displays a broad smile (smile .8). Midway through this script the smile value is updated. The value in the "behavior_smile_value" enables the character to tailor its smile expression to respond to the user. If the user is not smiling then the character uses the new "behavior_smile_value" and stops smiling, half way through the script. For example, if the Blue Eyes camera detects a user's smile at 80% confidence this information is transmitted to the Data Logger. The Behavior Engine will then aggregate 2 seconds of Blue Eyes data from the Data Logger and determine, according to its algorithms, whether it should or should not tell the character to smile at this time. A running average of the previous 2 seconds of data is calculated for each sensor mapped channel. If the average value crosses either of the preset thresholds for that channel then a randomly chosen 80% of these events are selected to invoke behavior changes in the character. These behavioral changes are implemented with a 4 second delay[1]. The Behavior Engine sends the appropriate value of the "behavior_smile_value" variable to the Character Engine. This value is then used whenever the variable is encountered in a Character Behavior Script. This enables a real-time serendipitous interaction that responds to the users detected expressions. This strategy can also be used in a loop to update the "behavior_smile_value", to respond serendipitously to the user, continuously.

The "Serendipitous Events" column of Table 3.2. demonstrates that the scripting language includes variables that monitor the state of the virtual world, such as the "behavior_smile_value". This is one type of layering between pre-scripted behaviors and serendipitous events that can occur. Another type of layering occurs when the Character Behavior Scripts run multiple sequences in parallel. The "introduction script" can call "sequence talk" to elicit mouth movements. Calling "sequence talk" will interleave the mouth movements with the actions already called for by the "introduction script". Calling "sequence_talk_variable_smile" instead of "sequence talk" would combine these two layering methods. Since the scripts can call actions and sequences based on traditional control structures, such as "if" conditionals and "while" loops, the scripts are quite flexible. In this way the system can be developed with a rich combination of pre-scripted and serendipitous behaviors.

By separating serendipitous events from scripted events this control-architecture allows the character behavior to combine information from the behavior repertoire with real-time affective information and run them in parallel. When the character delivers an intervention it can use the affective information from a serendipitous event to customize the delivery in real-time. This ability to layer behavior allows the character to adapt its expressivity to the users and enables it to repeat the same scripted events with differing affects.

---

[1] The 4 second delay is long enough so that the character's mirroring behavior is not consciously detected by users, yet short enough for the mirroring to have a social effect (Bailenson 2005).

Scripted Events	Serendipitous Events
sequence talk **smile .8** mouth .3 wait .4 then mouth .4 wait .2 then mouth .5 wait .4 then mouth .3 wait .6 then mouth .2 wait .7 then mouth .4 wait .2 then mouth .3 wait .1 then mouth .5 wait .1 then	sequence talk_variable_smile **smile .8** mouth .3 wait .4 then mouth .4 wait .2 then mouth .5 wait .4 then mouth .3 wait .6 then **smile behavior_smile_value** mouth .2 wait .7 then mouth .4 wait .2 then mouth .3 wait .1 then mouth .5 wait .1 then

Table 3.2. Character Behavior Scripts that present the two types of events that are supported by the Character Engine, Scripted Events and Serendipitous Events.

3.4 Character's Expression of Emotion

The emotional expression of the character is very rich. The Character Engine contains internal scalar variables or "knobs" that can be modified over time by the script. These knobs include posture (stooped versus erect), knees more bent or unbent, rate of eye blink, face coloration, sidling (side-stepping), energy level (snappy, quick movement versus slow movement), involvement (body follows gaze direction more or less), and jitteriness (for creating more or less nervous appearance). Face affect knobs constitute an integrated subsystem. (Perlin 1997). These include head turn, nod and tilt, eyebrows up/down, eye gaze direction, eyes open/closed, eyelid-centers up/down, mouth open/closed, mouth corners up/down, mouth narrow/wide, sneering. Each of these controls can either have the same value for the left and right sides of the face, or can be given left/right asymmetric values. The latter case is used for such gestures as winking and one-sided sneering or smiling (See Figure 3.2.).

Rather than provide only a high level emotional API, this system provides lower level physical affect knobs, which the script writer can combine to create the appearance of higher level or more subtle emotional affects. In particular, by providing lower level controls, such as mouth corners raised, rather than "smile", this enables script writers to create the appearance of a very rich set of emotional states, including even self-contradictory emotional states. For example, a character's mouth can be smiling while his eyelids can convey sad or neutral affect.

Other knobs that control physiological appearance are also being used for this project. The system allows the programmer to control, at run-time, physical attributes such as height, girth, knee angle, leaning forward and backward, and swaying speed and magnitude. In order to maximize the empathic effect of mirroring, in the current project a subset of these are being used to roughly match the individual user's physical characteristics (such as elements of posture, agitation, arousal, and facial expression) in real-time. These capabilities can prove to be useful in other contexts, as well.

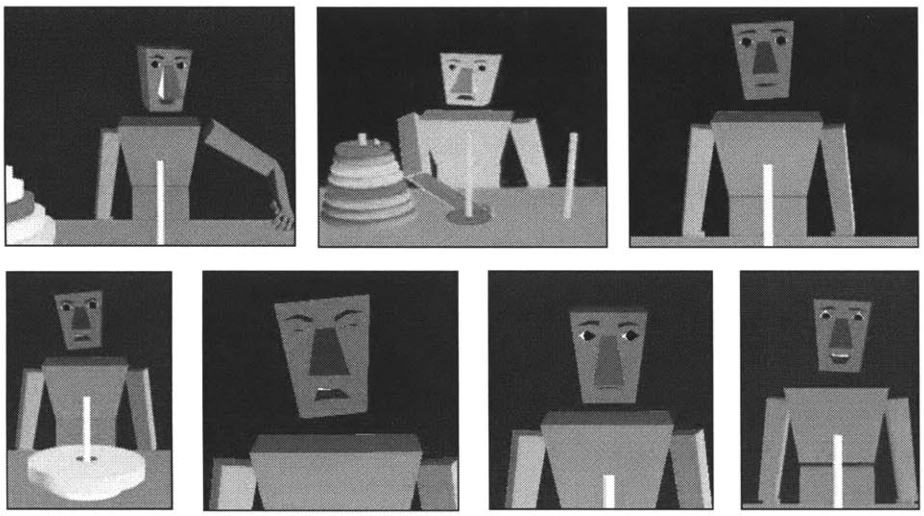


Figure 3.2. The agent is capable of a continuum of different expressions.

Crafting Character Expression

The architecture of scripted events, serendipitous events, and control knobs for affective expression is used for non-verbal communication. The goal for these interactions is to develop an empathetic relationship between the character and the user where the character "mirrors" the frustration and engagement of the user. Features of the user's emotional state are inferred by the Behavior Engine, and the Character Engine is instructed to direct the character to visually mirror aspects of the user's state. In this way, a virtual actor can appear to mirror users' emotional state without the virtual actor itself needing to have an extensive internal emotional model; the emotional state model is the built in the Classifier algorithms. Also, the "mirroring" can be conducted with variation that looks natural, so that it does not appear to be an exact duplication of what the user does.

Sensor:	**Character Behavior:**
Pressure Mouse	Magnitude of character swaying motion
Skin Conductance	Pigmentation of the skin tone (pale to flush)
Posture Analysis Seat	Leaning forward and leaning back
Blue Eyes Camera System	Smile, fidget, head movement, and head tilt

Table 3.3. Mapping of Sensors to Character Behavior

The mapping between the sensors and the character behaviors are shown in Table 3.3. The pressure exerted on the pressure mouse drives the apparent agitation of the character, which "sways" more at times of elevated pressure. The skin conductance sensor is mapped to changes in the color of the character; it becomes redder at times of greater arousal. The posture of the character, in terms of the interpersonal distance with the user, is controlled by the posture chair sensor and the associated Behavior Engine algorithms. The character's facial expressions and head movements are informed by data from the Blue Eyes camera system.

Developing the non-verbal mirroring conditions

The data collected by the Data Longer from pilot study participants was used to develop the two experimental conditions for non-verbal mirroring, *sensor-driven non-verbal mirroring* and *prerecorded non-verbal interactions;* the experiment methodology is described in chapter 4. As described later in this chapter, data from the pilot studies was processed to decompose each of the prerecorded sensor channels into several features and to calculate from the feature values the averages for 3 second time chunks for the duration of each participant's experience. While developing these chunks was integral to the process of training the Classifier algorithms it also allowed for comparisons to be made between participants in order to select the most "typical" participant data. These typical data files were the ones that were used to drive the character's behavior during the *pre-recorded non-verbal interaction* condition.

The *prerecorded non-verbal interactions* used in the experiment were recorded files from the data channels of participants that engaged in the third pilot study. These files were used by the System Server to provide pre-recorded data to the Behavior Engine. To determine which files to make available to the System Server, the mean values and the standard deviations of each of the behavior-mapped data channels and features (pressure, skin conductance, leaning forward or back, smile, fidget, and head tilt) for each participant were analyzed. Each participant's files were given a rank based on their proximity to the mean values relative to the other files, for each channel and feature. The ranks were summed and the files with the lowest overall ranking were investigated as candidates for driving the *prerecorded non-verbal interaction* condition. There were five files that had lower overall rankings than the others, these files also had no outlier rankings. These five files were then provided to the System Server, Behavior Engine and onward to the Character Engine and the resulting character behaviors were observed to determine the suitability of the interactions that they produced. The interactions were deemed suitable for all five files by two separate observers. Noting that Bailenson used multiple prerecorded files in his experiments (Bailenson 2005) so as not to bias the interactions in the prerecorded condition by the anomalies of any single prerecorded file, the five suitable files were all made available to the System Server. One file, selected at random, was used per participant in the *prerecorded non-verbal interaction* condition.

The *sensor-driven non-verbal mirroring* employs the sensor mapping described above in Table 3.3. If data for a channel was not available for a period greater than three seconds (e.g. when a participants pupils are not detected or if the skin conductance electrode as become detached) then in order to continue to display reasonable character behavior the individual channel would receive *prerecorded non-verbal interactions* from the randomly selected prerecorded file that was assigned to each participant at the start of the activity, until real-time data for that channel was available again.

Character Dialogue

In addition to non-verbal interactions the character interacts with the user through an asynchronous voice dialogue (Burleson, Picard et al. 2004). The character speaks using Microsoft's "Eddie" voice scripted with Text-Aloud, a text-to-speech application. When there are questions the words are presented in a text bubble, as well, for the user to read (Figure 3.3.). Users may respond by clicking on the available text responses. In the experiment described in chapter 4 and the pilot studies described in Appendix A, this asynchronous dialogue is used during the introduction, when the character presents the Towers of Hanoi activity and again during the intervention when the character provides tutorial support to participants. The character engine also supports the ability for the character to present a slide show; this feature is used to enable the character to present a persuasive message based on Dweck's treatments which have been shown to improve learners' self theories of intelligence and goal mastery orientation (Dweck 1999) (Figure 3.4.) (See Appendix C for additional character interaction images and dialogues).

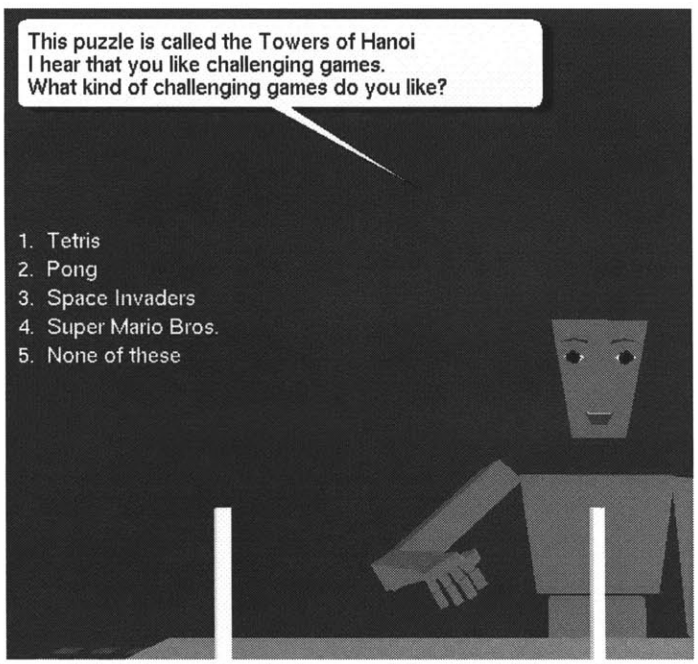


Figure 3.3. Character dialogue with selectable text response

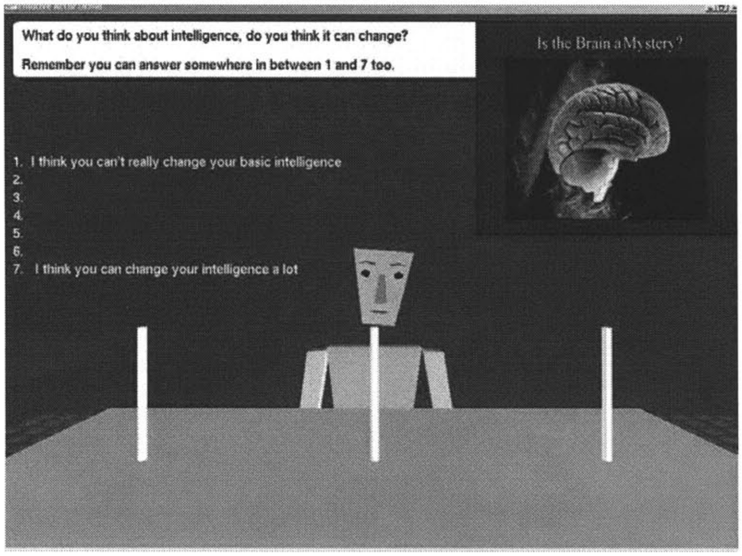


Figure 3.4. Character asking a question while presenting a slideshow

Dialogue Sequence for the Affective Support and the Task Support interventions

The following sequences present the tutorial support dialogue that the character provides as an intervention to participants in the experiment; see chapter 4 for the methodology of the experiment. It should be noted that while there are many elements of emotional intelligence, and many interesting methods for providing and studying affective support and task support there are practical limitations such as sample size and the time, effort, and expense, involved in running large numbers of participants that limit the number of conditions that can be studied. The interventions used in this thesis are informed by Klein and Bickmore's work on empathetic interventions (Klein, Moon et al. 2000; Klein, Moon et al. 2002; Bickmore and Picard 2004) and by Dweck's work on teaching metacognitive strategies that relate to meta-affective skill (Dweck 1999; Dweck 2004). Due to these multiple interests and due to the limitations mentioned above, the *affective support* intervention in this experiment has been designed to combine both empathetic interactions (with respect to the learners' levels of frustration) and reinforcement of meta-affective skills. This *affective support* intervention is juxtaposed by a *task support* intervention that is neither empathetic nor supportive of meta-affective skill. These two interventions were developed as stark contrasts (see table 3.4.).

	Empathetic (with respect to frustration)	Not empathetic (with respect to frustration)
Meta-affective skill support	The *affective support* used in this study includes Empathy and meta-affective skill support	Not empathetic, meta-affective skill
Other interventions (task, traditional metacognition, break taking, play, etc.)	Empathetic, other intervention (task, metacognitive, etc.)	The *task support* used in this study is not empathetic and does not support meta-affective skills

Table 3.4. *Affective Support* includes empathy and meta-affective skill support while *task support* excludes both of these elements of emotional intelligence. Other cells of the table represent other types and combinations of intervention strategies which are interesting areas for future investigation

Dialogue for Affective Support

I'm sorry I don't know more about this activity so I could help you through it. I do know that many people find it frustrating. On a scale from 1 to 7, how frustrated are you feeling right now?'

1. This is one of the most frustrating times I have ever felt while using a computer
2.
3.
4.
5.
6.
7. Absolutely not frustrated at all.

An adaptive response is given, for example, If the participant selects number 7 the following response is provided (see Appendix C for the response to other selections).

It sounds like you are extremely frustrated with this activity. Is that about right?"
Yes
No

If the answer from the participant is "No" then the following "repair" is stated:
Sorry about that, to clarify, how frustrated are you?

1. This is one of the most frustrating times I have ever felt while using a computer

.

.

7. Absolutely not frustrated at all.

The dialogue then continues with:
Wow, that must be really tough. I am really sorry doing this activity is making you feel that way.

How much effort do you feel you have been putting into this activity?"
1. Absolutely no effort at all

.

.

7. An enormous amount of effort

Another adaptive response is given, based on frustration and effort responses, for example if 7 was selected as the effort response the dialogue then continues with:
It is probably really aggravating to have to stick with this activity when you're already putting in a lot of effort and finding it so frustrating. Please remember that it is ok to be frustrated. It is great that you are aware of how you feel. Remember, frustration sometimes tells you to try things differently. It is like a navigation sign that says, "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work!

Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you WILL get better and smarter.

Do you think that you will be able to try these strategies?
1. Yes, I think I can
.
.
7 No, I do not think I can

Another adaptive response is provided, for example if seven is selected the response is:
It can be hard, but remember that's a sign that you are learning, stick with it and you will learn a lot.

I have to go now. Thank you for letting me watch you do this activity.
Watching you has helped me learn too. Sorry that I have to leave now.

How do you feel about continuing the activity?
1. I am very willing to stick with it.
.
.
7. I am not at all willing to stick with it.

One of the following two responses is provided based on the participant's answer:
Great, good luck! Please try as hard as you can. If you feel like you would like to stop there are a few buttons in the upper right hand corner that you can press.

I'm sorry that I have to ask you to continue anyway. Please just try as hard as you can. If you feel like you would like to stop there are a few buttons in the upper right hand corner that you can press.

Bye bye.

Participants are then presented with the there bye response which they can select.

"Ok, bye I was glad to have you here"
"Ok, bye"
"Ok glad you are finally going"

Dialogue for Task Support

The dialogue in the *task support* sequence does not adapt to the participant responses, so it is presented in its entirety in this section.

I'm sorry I don't know more about this activity so I could help you through it. I do know that many people find it frustrating. On a scale from 1 to 7, how frustrated are you feeling right now?'
1. This is one of the most frustrating times I have ever felt while using a computer
2.
3.
4.
5.
6.
7. Absolutely not frustrated at all.

How much effort do you feel you have been putting into this activity?

1. Absolutely no effort at all

.

.

7. An enormous amount of effort

Ok, well, here are some tips others have told me they think about while doing the activity. You can think about where you want the big disk to go and then try to move all of the other disks out of the way. These disks can go to another pole so that you can move the big disk to where you want it. Remember that if you get all of the disks that are in the way out of the way then you can move the disk that you want to move. Another way to think about this is to think about the small disks that are in the way. If you move these out of the way then you can move the disk that you want to move. Some people try to do this in as few moves as possible.

Do you think that you will be able to try these strategies?
1. Yes, I think I can

.

.

7 No, I do not think I can

Well, just give it your best.

I have to go now. Thank you for letting me watch you do this activity.
Watching you has helped me learn too. Sorry that I have to leave now.

How do you feel about continuing the activity?
1. I am very willing to stick with it.

.

.

7. I am not at all willing to stick with it.

Well try as hard as you can. If you feel that you need to stop there will be a few buttons in the upper right hand corner that you can press.

Bye, bye,

Participants are then presented with the there bye response which they can select.

"Ok, bye I was glad to have you here"
"Ok, bye"
"Ok glad you are finally going"

3.5 Testing the System

The system was tested in a series of pilot studies (see Appendix A) prior to proceeding with the experiment described in chapter 4. These pilot studies provided the opportunity to collect sensor data from pilot participants. Initial data collection (Figure 3.5.) showed that there were gaps in the frequency that data from each channel was collected. Reducing the number of sensors used improved the collection rate so it was determined that the Data Logger was getting overloaded. The Data Logger was redeveloped into a multi-threaded module so that it could both listen for new data and send its data on to the System Server at the same time. Testing the new Data

Logger showed significant improvement in the consistency of the frequency (Figure 3.6.). While there were still a few lapses in data collection it was decided that these occurred infrequently and did not warrant further attention, especially in light of the concurrent need to further develop the Behavior Engine and Character System.

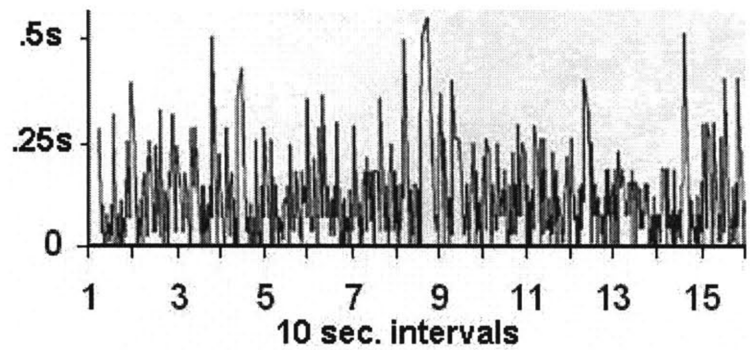


Figure 3.5. The plot shows the frequency of data collected by the Data Logger for a single sensor channel (Blue Eyes) with the initial Data Logger. The taller spikes indicate several 1/2 second gaps due to dropped packets during this 2 ½ minute time period.

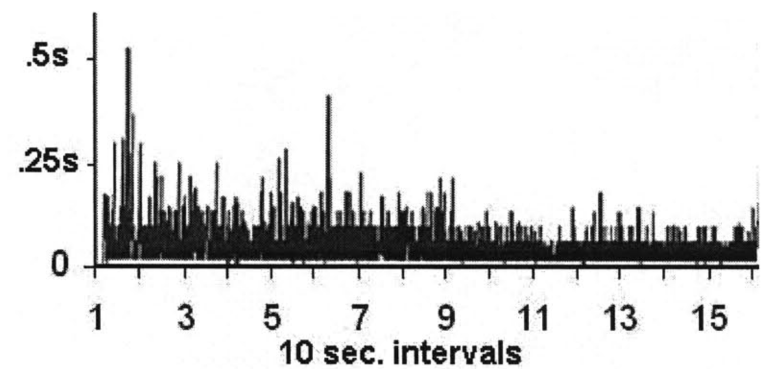


Figure 3.6. The plot shows the frequency of data collected by the multi-threaded Data Logger. There are substantially fewer spikes than in Figure 3.5., most data at a rate of .1 to .2 seconds.

Once the Data Logger was functioning properly further pilot studies collected additional data from participants. Figures 3.7., 3.8., 3.9., 3.10., 3.11., and 3.12. show the data collected from a single pilot participant in each of the various sensor channels. The data from these participants was then used to inform the design of the character's non-verbal mirroring behavior, described above, and to train the classifier algorithms, discussed in the next section).

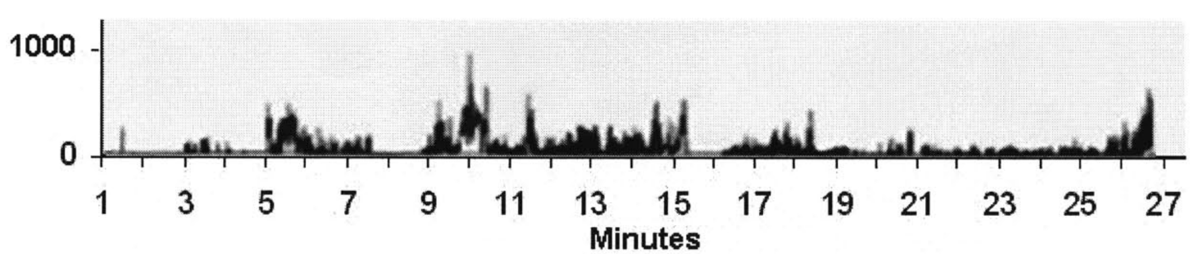


Figure 3.7. Pressure mouse data proportional to the force applied on the mouse surface

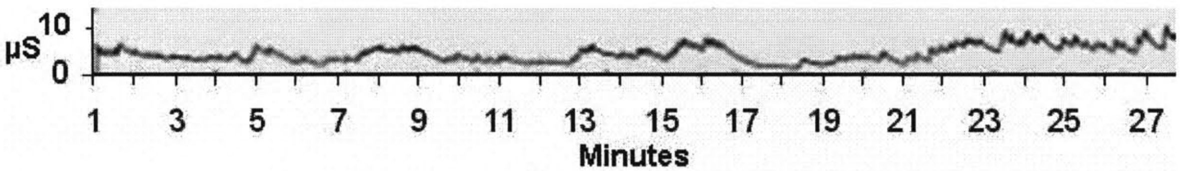


Figure 3.8. Wireless Blue Tooth skin conductance in microSiemens

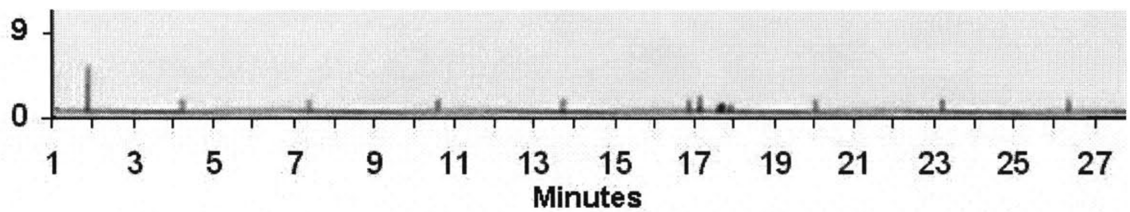


Figure 3.9. Posture analysis seat with 9 discrete posture states (this participant did not change their posture very much)

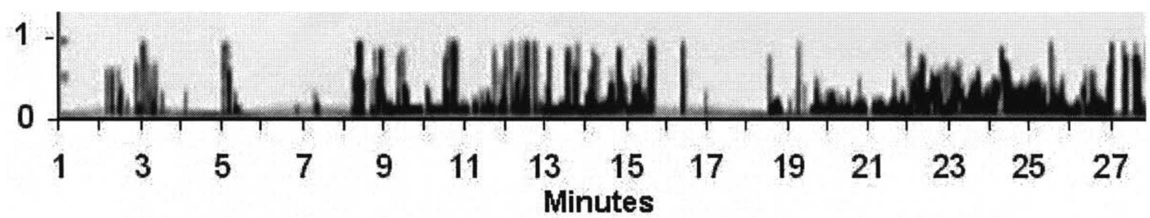


Figure 3.10. Blue Eyes camera classification in % confidence of a smile

Figure 3.11. Blue Eyes camera classification in % confidence of a fidget

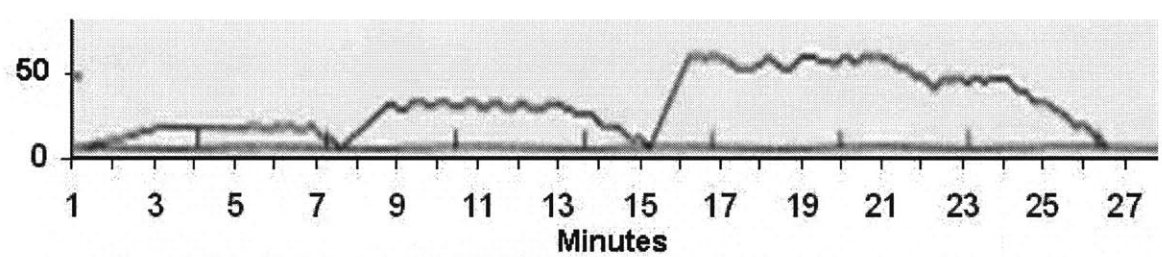


Figure 3.12. Game state representing the least number of moves, for an optimal solution from the current configuration of the disks

3.6 Affective Classifier

The simplest "affective classifier" for an empathetic agent is to recognize user movements and use these to move a character. This simple classifier enables the character to engage in mirroring or mimicking, such as leaning toward the user, smiling, nodding, and so forth. This can be recognized with the current pattern recognition tools that have been developed for each of the sensors (Kapoor and Picard 2005) and is currently implemented in the decisions and directions of the Behavior Engine.

A more advanced affective classifier has also been developed discerning elements of the user's affective states such "Stuck". By coupling data from all of the different sensor channels, this classifier achieves an improvement over any one channel (Kapoor, Burleson et al.; Kapoor, Picard et al. 2004). The current classifier can detect if an individual is likely to choose to keep engaging in a task, or whether the learner is frustrated or likely to seek outside help (quitting) with 79% accuracy (chance = 58%) (Kapoor, Burleson et al.). While most of the learning conducted by the classifier is done offline, in parallel with algorithm development, this system can also detect some kinds of user's affective states in real time. This real time ability was not necessary for this study since it was desirable to keep agent responses constant despite user differences. Therefore this more advanced classifier has not been used in the experimental methodology. Analysis of the user data was conducted offline, after the experiment, to determine the emotional state of the users at various times throughout their experience with the activity. This information is then assessed along with self-report measures in the analysis of the experiment. (As discussed in chapter 5.1, an investigation of the results of the classifier on data from the experiment indicates that the classifier did not detect individuals who were frustrated or seeking help; the classifier may need to be retrained and this retraining is a priority for future work.)

This analysis occurs in several steps. Labeling of the data comes from four sources: in situ self-labeling by the user's clicking on the quit button, self-report from the user surveys, the sensor feature data algorithms, and from human coding of the video data. The raw sensor data is processed for features, such as "startle events" in skin conductance, high pressure exerted on the pressure mouse, the ratio of leaning forward to leaning backward, and the confidence of detection of specific features (smiles and mouth fidgets) by the Blue Eyes facial expression camera. This information is chunked into 3 second time periods and averaged for each feature within those periods. A joint analysis, of all of the sensors, for the series of 3 second time sequences is conducted using unsupervised and semi-supervised clustering procedures: training with Hidden Markov Models, State Vector Machines, and Dynamic Bayesian Networks. The process that has been most promising is the use of the 5 most significant features for 150 seconds x 3 second frames. These five features include the: mean skin conductance, mean seat activity ratio of forward posture to backward posture seat forward/back ratio, mean confidence level of mouth fidget, and mean head tilt. Training on these and classifying with the "leave one out" method, yields 79% accuracy for an SVM with RBF kernel learning the parameters. Table 3.5. presents the results of this classification (Kapoor, Burleson et al.).

	Clicked a Button for frustration or help		Persevered on Task		Accuracy
	10 Samples		*14 Samples*		
	Correct	Misses	Correct	Misses	
Random (Control)	0	10	14	0	58.3%
1-Nearest Neighbor	6	4	10	4	66.67%
SVM (RBF Kernel)	6	4	11	3	70.83%
Gaussian Process	8	2	11	3	**79.17%**
SVM + Kernel of Gaussian Process	8	2	11	3	**79.17%**

Table 3.5. Classifier Results Showing 79.17 % accuracy.

3.7 Summary of the Affective Agent Research Platform

The Affective Agent Research Platform has been synthesized and iteratively tested. It is composed of several sensors that are relevant to the detection of affective states. It has a modular architecture with a System Server core that is flexible to facilitate additional modular improvements to the system capabilities. It has a character that exhibits behaviors that are broadly expressive and interactive through scripted behavioral repertoire and serendipitous events that use sensor events mapped to character behavior to provide *sensor-driven non-verbal mirroring*. Through pilot studies the system has been refined and an off-line Classifier capability has been developed that can detecting participant's likelihood of quitting with an accuracy of 79%.

4 Hypothesis and Method

This chapter describes the approach taken to utilize the affective agent research platform to develop and evaluate an emotionally intelligent affective learning companion that can create enhanced social bonds with learners through non-verbal mirroring and that can use effective intervention strategies to support learner's metacognitive and meta affective skill. The hypotheses are presented in terms of their relationship to the related work and theory presented in chapter two, and the system architecture discussed in chapter three. The methodology and measures are then presented along with prescriptive analytical guidelines for their evaluation, which is conducted in chapter five and discussed in chapter six.

4.1 Presentation of Hypothesis:

There are four main areas to which this work has planned contributions. First, it seeks to extend Bailenson's use of Transformed Social Interactions in four significant ways: providing a new domain (a learning platform), including a new meta affective skill based message, addressing a new audience (11-13 year old learners), and using a new set of less invasive sensors. The affective learning companion is expected to be more persuasive, and users will form a stronger social bond with the affective learning companion, when sensors inform the affective learning companion interactions (vs. when sensor information is not included).

Second, this research will create new applications of Dweck's strategies of intervention that facilitate shifts in learner's self-theories of intelligence. An affective learning companion that uses strategies of emotional intelligence and applies them in a reliable and controllable manner should have an ability to successfully introduce learners to Dweck's strategies. Adoption of these strategies is predicted to have behavioral effects. This should improve learners' awareness of and resilience to frustration, thereby increasing perseverance and intrinsic motivation. The hypothesis is that a learner's social bond with an affective learning companion and The level of persuasion a learner experiences from the affective learning companion's metacognitive and meta affective skill based message will have an impact on a learner's self-theories, frustration, and perseverance.

Third, this research will apply the theories of Flow and Stuck to the construction of an intervention. The design of interventions will also take into account strategies for empathetic and caring relationship development (Bickmore and Picard 2004) as well as Klein's work on "frustration handling" (Klein, Moon et al. 2002). An affective learning companion that exhibits emotional intelligence (active listening, appropriate interventions and social mirroring) is expected to increase learner's intrinsic motivation.

Finally, the research will evaluate Flavell's strategy for metacognitive skill development and investigate a new meta affective skill construct as it applies to an intelligent tutoring system with affective sensing ability (Flavell 1987). Meta affective skill addresses a learner's awareness of feelings during an activity. An affectively aware Learning Companion might facilitate a learner's awareness of their feelings. The affective classifier, informed by multi-modal affective sensing, allows the system to predict individual learner's affective state (Stuck and Not-Stuck). The effect of affective learning companion interventions on learners can then be evaluated. It is expected that metacognitive and meta affective skills will be exhibited at higher levels when learners interact with emotionally intelligent affective learning companion and will positively correlate with perseverance, willingness to continue, and intrinsic motivation.

In summary, the specific hypotheses are the following:

- H1: The affective learning companion is expected to be more persuasive, and users will form a stronger social bond with the affective learning companion, when sensor-driven non-verbal mirroring informs the affective learning companion's interactions.

- H2.A: A learner's social bond with an affective learning companion will positively correlate with his or her perseverance and self-theories – adoption of internal beliefs that he or she can increase his or her own intelligence and the adoption of mastery orientation.

  H2.B: The level of persuasion learners experience from the affective learning companion's metacognitive and meta-affective skills based message will positively correlate with social bond, with perseverance, and will negatively correlate with frustration.

- H3: An affective learning companion that exhibits emotional intelligence (active listening, appropriate interventions, and sensor-driven non-verbal mirroring) will increase learners' intrinsic motivation and reduce frustration.

- H4: Metacognitive and meta affective skill will be exhibited at higher levels when learners interact with emotionally intelligent agents and will positively correlate with perseverance, willingness to continue, and intrinsic motivation.

The measures used in these hypotheses will be discussed below.

4.2 Methodology

There were four experimental contrasts derived from a 2 x 2 design: interactions between affective learning companion and child were either *sensor-driven non-verbal mirroring* or *prerecorded non-verbal interactions* acquired from a prior participant; interventions by the affective learning companion were either an *affect support intervention,* designed to attend to the emotional state of the learner or *task support intervention,* designed to provide the learner with constructive information about the activity (see chapter 3 for examples of the *affect support* and *task support* dialogues, and see Appendix C for transcripts of the character dialogues from the introduction and *affect support*). The two *mirroring* conditions are the *sensor-driven non-verbal mirroring* and the *prerecorded non-verbal interaction,* described in chapter 3.

	Sensor-driven non-verbal mirroring	*Prerecorded non-verbal interaction*
Affect support intervention	*Affect support, non-verbal mirroring*	*Affect support, prerecorded non-verbal interaction*
Task support intervention	*Task support, non-verbal mirroring*	*Task support,, prerecorded non-verbal interaction*

Table 4.1. Four experimental contrasts in a 2 x 2 Design comparing *intervention* x *mirroring* conditions

A pre-test was administered to determine children's self theories of intelligence and their goal mastery orientation (Dweck 1999). Participants were randomly assigned to one of the four *non-verbal interactions x intervention* conditions. The affective learning companion was presented; it then engaged in *NVM* or *prerecorded* interactions throughout the time of its presence. The affective learning companion presented a slide show during which it asked the learner several questions. The slide show was based on a script used by Dweck that has been shown to shift children's beliefs about their own intelligence toward incremental self theories (Dweck 1999). The affective learning companion then presents the Towers of Hanoi activity and explains that it may have to leave before the learner completes the activity. The affective learning companion instructs the learner to, "Click on a disk to start whenever you want, I'll just watch and help if I can." The learner is given four minutes to engage with the activity before the character intervenes with either an *affect support* or *task support* based intervention. During the intervention self-report measures are obtained through the interaction with the affective learning companion when it asks face to face questions of the learner, e.g." On a scale from 1 to 7, how frustrated are you feeling right now?" . At the end of the intervention the learner is encouraged to continue with the activity and the affective learning companion departs, saying "bye bye." (see chapter 3 for the intervention dialogues). Participants have the opportunity to respond by pressing one of three *bye.buttons*: "Ok, bye", "Ok, bye I was glad to have you here", or "Ok glad you are finally going"; presented in different random order for each participant, to control for presentation order effects, see Figure 4.1. After they click one of the three bye buttons, or after 20 seconds elapses (when the bye buttons disappear/time out so as not to distract the participant), the three quit buttons that the character previously discussed appear (Figure 4.2.) in the upper right corner of the screen offering the opportunity for the learner to end the activity: "I want to stop because I'm too frustrated to continue", "I've put in all the effort that I can and want to stop", and "I want to stop for some other reason".

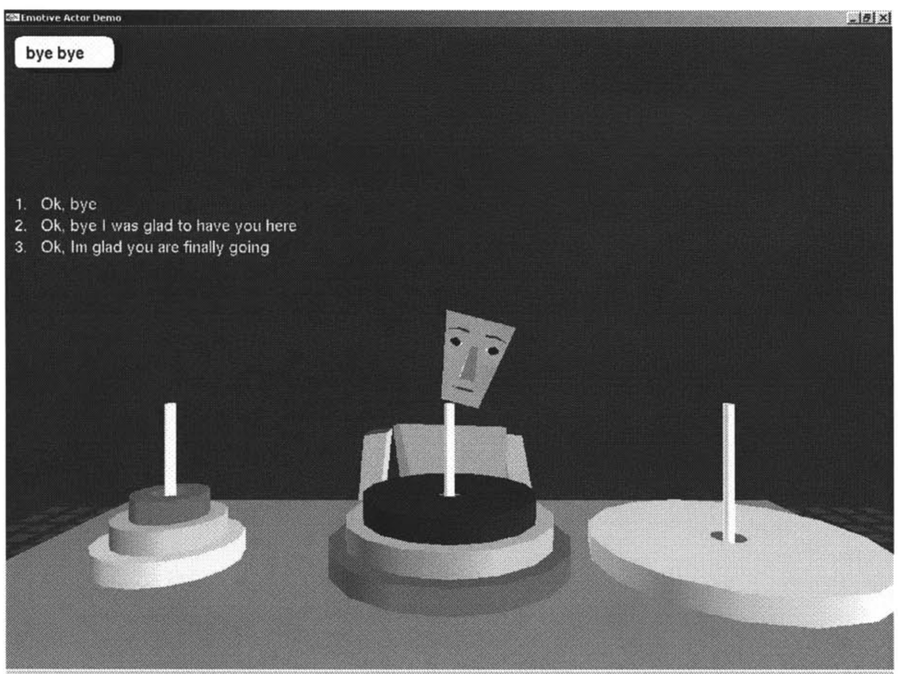


Figure 4.1. Character saying good bye with randomly ordered *bye.buttons* responses

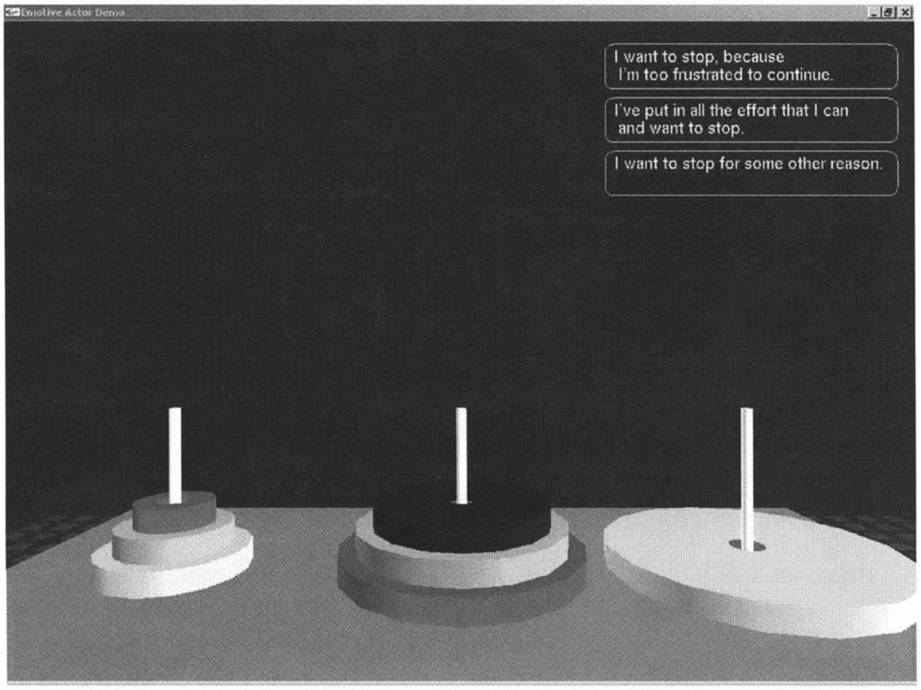


Figure 4.2. Three buttons offering the participant the opportunity to quit

Protocol Events for Subjects in all Four Conditions	Duration in minutes
Assent and consent forms	~3
Initial Survey Questions and Pre-Test (including Self Theories of Intelligence and Goal Mastery Orientation)	~10
Character introduces activity and shows Slide Show (based on Dweck's message)	~10
Participant engages in Towers of Hanoi activity	4
Character provides *affect support* or *task support intervention*	~2
Participants persist in Towers of Hanoi task with "quit" buttons present	up to 15 minutes from the start of the activity
Post-activity survey of experience	~3
Neutral affect inducement video	1.5
Post-Test (including Self Theories of Intelligence and Goal Mastery Orientation)	~10
Modified Working Alliance Inventory	~2
Opportunity to reengage with Towers of Hanoi	2

Table 4.2. Experiment protocol with durations in minutes; the approximate values indicate that these events have participant interactions and therefore some variation in duration

At the time the learner clicks a button, or 15 minutes after the start of the activity, whichever happens first, the learner is presented with post-activity questions about the experience e.g. "How many minutes would you say this activity took from the time you first moved a disk until now?", "Mark how much of the time you were frustrated", and others, (see Table 4.4.). After these questions the learner is presented with a 1.5 minute video clip of a seascape, as a neutral affect inducement (Rottenberg, Ray et al. in press) to help alleviate frustration that may bias answers to subsequent questions. The learner is then presented with post-test questions on self theories of intelligence and goal mastery orientation, followed by a modified Working Alliance Inventory (Bickmore and Picard 2004) to gauge his or her impression of the character. Finally the learner is again presented with the Towers of Hanoi activity, still in its previous state, along with instructions indicating that, "It will be a couple of minutes before the next activity is ready. You can do whatever you want now, just stay seated here please." After two minutes another message appears and says: "Thanks for waiting." The experimenter informs them they are done, and conducts a debriefing. This final two minute period allows a learner to reengage in the activity, as an indication of intrinsic motivation.

4.2.1 Participants

The participants were 11-13 year old children from three semi-rural schools in western Massachusetts (Lanesborough School, in Lanesborough; Richmond Consolidated School, in Richmond; and St Mark School, in Pittsfield). Seventy six participants were randomly assigned to the conditions as shown in Table 4.3. Attrition eliminated 10 students due to a variety of factors: some participants needed to leave prior to completing the activity, due to unexpected all-school meetings or changes in transportation schedules, power failures due to storms, some equipment failures occurred, a few participants were unresponsive to the character interactions (e.g. did not answer several questions even after the experimenter instructed them to do so, so the timing of the introduction and the intervention were inconsistent with respect to other participants), and one participant who was identified by her teacher as a student with special needs (her rapid response times to the self report questions also indicated that she did not take time to read the questions.) Beyond the 10 that were not used due to attrition there were four participants who were not included because they finished solving the Towers of Hanoi puzzle on their own: one of these four indicated that they knew the game, a fifth participant that indicated that they knew the game well was also excluded. Totaling these 5 with the 10 excluded due to attrition, 15 of the 76 participants were excluded, resulting in 61 participants who were included. In addition to the participants in this experiment there were several pilot studies which collectively had over 60 participants (see Appendix A, for a discussion of the pilot studies).

	Sensor-driven non-verbal mirroring	*Prerecorded non-verbal interaction*
Affect support intervention	16 valid out of 20 assigned 8 girls valid out of 10 8 boys valid out of 10	14 valid out of 19 assigned 5 girls valid out of 8 9 boys valid out of 11
task support intervention	15 valid out of 18 assigned 7 girls valid out of 8 8 boys valid out of 10	16 valid out of 19 assigned 9 valid out of 11 7 boys valid out of 8

Table 4.3. Valid participants assigned to the four experimental contrasts of the 2 x 2 design and included in the analysis of the hypotheses.

4.2.2 Measures:

This section will present the measures used in this experiment. The presentation sequence is guided by the order in which they are obtained from the participants and is consistent with the protocol presented in Table 4.2., Table 4.4. shows the measures, method used to obtain the measure, and the source of the method. While it repeats the sequence described in the methodology, this is done intentionally to provide clarity. Survey instruments and questions for this experiment and for the pilot studies can be found in Appendix B.

The initial survey questions asked participants their *age, hours of computer use* on an average day and number of *hours playing computer games*; these questions can be found in Appendix B. Their *gender* and *school* were recorded by the experimenter.

The pre-test survey instruments included the *pre-test self theory of intelligence* and *pre-test goal mastery orientation* measures (Dweck 1999). These instruments were also administered after the activity as *post-test self theory of intelligence* and *post-test goal mastery orientation.* These each get scored to obtain a single value for *self theory of intelligence* and another single value for *self theory of goal-mastery orientation.* (The *self theories* survey instruments can be found in Appendix B).

After the pre-test surveys the participants were asked to click on start and the character appeared. The character introduced itself, the activity, and the slide show. The character asked several questions and participants responded to the character using an on-screen 7 point Likert scale. These questions asked participants if they knew the game, if they thought people have the ability to change their intelligence, if they remember a time when they were frustrated, how hard it is for them to do new things when they are frustrated, if they thought the slideshow and its message that the mind is like a muscle would help them, how hard they thought the activity would be, and if they would be able to try new strategies as they do the activity.

The participants then preceded with the Towers of Hanoi activity with seven disks on the left most pole and the goal of moving them to the right most pole, moving only one disk at a time, and without putting a bigger disk on top of a smaller one. After 4 minutes, during which the participants pursued the activity, the character intervened with either *affect support* or *task support*. The sensor data, from the 150 seconds period prior to the character intervention, was used by the affective classifier, off-line to provide a *likelihood of frustration/help-seeking* rating (Kapoor, Burleson et al.) (also see chapter 3 for more discussion on the affective classifier).

The affective classifier *likelihood of frustration/help-seeking* rating was also used to obtain a *congruence* measure. This measure was the degree to which the intervention (*affective support* or *task support*) matched the prediction of the affective classifier. *Affective support* is expected to be more productive for learners that the affective classifier identifies as being frustrated/help-seeking. The *affective support* intervention is therefore said to be congruent for these participants. *Task support* interventions are expected to be more productive for those learners whom the affective classifier predicts will not quit/seek-help and is therefore also said to be congruent. The inverse matching would be rated as incongruent.

During the character intervention a series of questions are asked by the character and participants are again able to answer with a 1-7 response (see Appendix C for screen shots that include each of these questions). These questions asked participants how frustrated they were feeling right now, how much effort they had been putting into this activity, if they would be able to use the strategies presented in the intervention, and if they are willing to stick with the activity. The response to how frustrated they are can also be used to develop a second congruence measure, e.g. using this measure in place of the affective classifier's rating, with the same assumptions for congruence as described above.

At the end of the intervention with the character, the character said "bye bye" and participants were given the opportunity to click on one of three buttons to say good bye. The buttons were presented in a randomized order to each individual to avoid order effects, they read: "Ok, bye", "Ok, bye I was glad to have you here", and "Ok glad you are finally going". This response was coded as a categorical *social bond behavior* metric that will be analyzed in conjunction with the self-report *social bond* measure.

When a participant clicks on one of the three quit buttons or if they have persisted through the activity time period to the end, they are presented with a post activity questions (see Table 4.4.). The *perseverance* behavior of the participant is assessed as the duration of the activity, from the start of moving the disks until the participant clicks on a button to end the activity or 15 minutes, whichever happens first. Also, if a participant does not click on a button, but rather continues until the end of the 15 minute time period this is recorded as separate measure indicating that the participant has shown the maximum perseverance.

The first question, in the post activity survey, asked participants how many minutes they would say this activity took, *perceived duration*. This question was followed by a series of questions that are used as a measure of the participant's *frustration*. These questions ask participants how much of the time they were frustrated, to rate their maximum level of frustration, and if they clicked on a button how frustrated they were at that time.

The survey proceeded to ask if they would like to *try the activity again* and separately if the *activity was too hard* (see Appendix B for exact format of questions). A series of *metacognitive* and *meta affective skill* based questions were then presented asking about the helpfulness of thinking of the mind as a muscle, the amount of time different strategies were used and the perceived helpfulness of using different strategies, how many strategies were used, how aware of their frustration they were, how much control they felt during the activity, what their ability to concentrate on the activity was, and if they felt like they had the skills necessary to do this activity. Since these questions were directly about the activity they were asked prior to the neutral affect inducement video to minimize the time between the participants experience of the activity and these questions.

The neutral affective inducement video was then presented (Rottenberg, Ray, Gross, in press) to mitigate the effects of frustration on the participants answers for the *post-test self theory of intelligence* and *post-test goal mastery orientation* survey instruments (the post-tests are unaltered re-administrations of the pre-tests).

The final survey was the *social bond* with the character obtained through a self-report survey, based on a modified Working Alliance Inventory with a few additional questions (Appendix B). In addition to standard Working Alliance Inventory questions, about their direct perceptions of the character, this survey also asked if they thought that the character sensed their frustration, if they were more aware of how hard the activity was, how frustrated they were with the character present, and if being more aware of their frustration helped them keep going. They were also asked if the character was easy to understand and how old they thought the character was. These measures were obtained after the *post-test self theory of intelligence* and *post-test goal mastery orientation* surveys to prevent the affective states, which may have been induced in participants by recalling negative or positive impressions of the character, from affecting participants' responses to those measures. It should be noted that any induced affective states will however have an increased impact on the measure of *intrinsic* motivation.

Finally *intrinsic motivation* was measured by assessing reengagement with the activity during the final two minutes of the participant experience. *Intrinsic motivation* will be analyzed through the use of two variables. The first is a binary variable showing either no further engagement in the activity or further engagement measured as a one or more disks being moved. The second is a continuous measure for those that showed engagement indicating how many moves they made.

A composite measure, *Motivation*, will be investigated by analyzing correlations between the two binary measures for *intrinsic motivation* and two self-report measures, the *willingness to continue* and the expressed desire for *next task difficulty.*

Teachers familiar with each of the individual students were asked how long they think a student will spend on a task that they present to him or her, after the teacher has left the room. The rating was conducted on a 7 point Likert-scale from "not long at all" to "very long". There was only one rating per student and each teacher had worked with the student within the past year. This rating was labeled as *teacher rated perseverance* (see Appendix B for the presentation of the *teacher rated perseverance* question).

Measure	**METHOD**	**SOURCE**
Initial Survey Questions	self report	developed
Age		
hours of computer use		
hours playing computer games		
gender		
school		
Pre-Test	self report	Dweck
pre-test self theory of intelligence		
pre-test goal mastery orientation		
Introduction questions (asked by character)	responses to character's questions	developed
Do you know the rules? Have you seen this game before?		
What do you think about intelligence, do you think it can change?		
Can you think of a time when you practiced something and felt like you got better at it?		
Have you been stuck in a line whishing you didn't have to wait?		
On a scale from 1 to 7 , how hard do you think it'll be for you to do new things if you get frustrated?		
What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort?		
If things get hard for you in this activity, do you think you will be able to try different ways of doing the activity?		
On a scale from 1 to 7 , how hard do you think this activity will be for you?		
Affective Classifier	Affective Classifier	trained - Kapoor, Burleson, Picard
likelihood of quitting/help-seeking		
Congruence	Affective Classifier/ self report/random assignment	trained/developed

Intervention question (asked by character)	self report	developed
On a scale from 1 to 7, how frustrated are you feeling right now?		
It sounds like you feel somewhat frustrated with this activity. Is that about right?		
How much effort have you been putting into this activity?		
Do you think that you will be able to try these strategies?		
How do you feel about continuing the activity?		
Social bond behavior (Bye.button response)	quasi behavior	developed
"Ok, bye I was glad to have you here"		
"Ok, bye"		
"Ok glad you are finally going"		
Perseverance (in minutes)	behavior	standard
Perceived duration (percent over-estimation of actual time)	self report (drop down box, 0 -- 100)	Weybrew, Czerwinski; Zeigarnik ; Picard, Liu
How many minutes would you say this activity took from the time you first moved a disk until now?		
Frustration - Post activity questions	self report	developed
Mark how much of the time you were frustrated.		
Mark how frustrated you were at the most frustrated time.		
If you clicked on one of the buttons, mark how frustrated you were at that time.		
Post activity questions	self report	developed
Mark how excited you are now that you are done with this activity.		
I would like to try this activity again.		
This activity was too hard.		
Metacognitive and Meta Affective Skills questions	self report	developed
How helpful was it to think of your mind as a muscle?		
How much of the time did you try to use different strategies to do this activity?		
If you used different strategies how helpful did you find them?		
If you used different strategies, how many did you use? (drop down box, 0 -- 100)		
While you were doing this activity how many times where you aware of your frustration?		
How much control did you feel you had during this activity?		
What would you say about your ability to concentrate on this activity?		
Do you think you have the skills needed to do this activity?		
Neutral Affect Inducement	treatment (video)	Rottenberg, Ray, Gross
Post-Test Self Theories of Intelligence		
post-test self theory of intelligence	self report	Dweck
post-test goal mastery orientation		
Social bond questions	self report	developed
Modified Working Alliance	self report	Bickmore,

		Picard
Casey seamed to sense when I got frustrated, even though he couldn't help me.	self report	developed
With Casey there I was more aware of how hard or frustrating the activity was.	self report	developed
Being more aware of how hard or frustrating it was helped me to keep going.	self report	developed
How old is Casey?	self report	developed
It was easy for me to understand what Casey was saying.	self report	developed
Intrinsic motivation	behavior	Dweck
Motivation scale (possible composite)	analysis	developed
intrinsic motivation (behavior)		Dweck
How do you feel about continuing the activity?	self report	developed
I would like to try this activity again	self report	developed
Teacher rated perseverance	self report by participants' teachers	developed

Table 4.4. The measures, the methods used to obtain the measure, and the source of the method

4.3 Strategy for Hypothesis Analysis:

In the process of investigating each of the hypotheses, the effects of fixed factors: age, school, and gender; hours of computer use; and hours playing computer games, will also be assessed. These factors will be studied for use as covariates, to control for the unintended variances.

- H1: The affective learning companion is expected to be more persuasive, and users will form a stronger social bond with the affective learning companion, when sensor-driven non-verbal mirroring informs the affective learning companion's interactions.

To investigate H1 the participants who experienced *prerecorded non-verbal interactions* will be compared to those who experienced *sensor-driven non-verbal mirroring* with respect to *persuasion* and *social bonding*. There are several opportunities to measure *persuasion*. First the analysis will use several of the questions that the characters asked participants during the introduction. Then it will use the questions asked during the intervention: "Do you think that you will be able to try these strategies?" and" How do you feel about continuing the activity?"; controlling for the intervention condition received and for their answer to how frustrated they are at that time, as covariates. A third opportunity is to see if participants report that *the character was easy to understand* in the *sensor-driven non-verbal interaction* group. The rational for this is that *understanding* in a general sense might relate to *social bond* and therefore be influenced by *mirroring*. A fourth means of assessing *persuasion* is to test for differences in *perseverance* as a behavioral result of the persuasive message; this is possible since the character's message attempts to persuade, and provide strategies to help, participants to persevere. A fifth method of investigating *persuasion* is by comparing the participant's responses to the *metacognitive* and *meta affective skill* based message and the differences in participant's answers to Dweck's *pre-test and post-test measures*, as these measures deal directly with the topic that the agent is presenting.

There are, likewise, several opportunities to measure the *social bond*. There is the opportunity to investigate a quasi-behavioral measure of the social bond by comparing the "good bye" responses of the participants. This quasi-behavior measure along with the social bond questions in the modified Working Alliance Inventory, individually and in aggregate through the construction of a scale will be investigated. There is the possibility to control for the effects of the following two self report measures, *if the character was easy to understand* and *how old they thought the character was*. It should be noted that using these two measures as control variables is complicated by the fact that each one of them measures elements of participants' experience of the character, as such they can not really be said to be independent of the social bond. In the spirit of discovery and learning, the investigation will be conducted and the results discussed.

The correlation between the several measures of *persuasion* and *social bond* will be assessed to see if participants that show evidence of a stronger *social bond* also show evidence of greater levels of persuasion. Scales of these two constructs, composed from multiple measures, will also be investigated.

- H2.A: Learners' social bonds with an affective learning companion will positively correlate with their perseverance and self-theories – adoption of internal beliefs that they can increase their own intelligence and the adoption of mastery orientation.

  H2.B: The level of persuasion learners experience from the affective learning companion's metacognitive and meta affective skills based message will positively correlate with social bond, with perseverance, and will negatively correlate with frustration.

The investigation of H2 will use the *social bond* and *persuasion* variables and scales, developed to analyze H1. These scales will be used to predict learner's *self-theories*, *frustration*, and *perseverance* using regression. Participants' *pre-test self theories* will be used as covariates, to control for ceiling effects (participant that already hold strong positive *self theories* that help them *persevere* through *frustration* are unlikely to exhibit additional changes due to *persuasion* and *social bonding*. Different investigations for the various measures of *frustration* will be conducted. In these investigations *teacher rated perseverance* will be used as a control along with the *type of intervention*.

- H3: An affective learning companion that exhibits emotional intelligence (active listening, appropriate interventions, and sensor-driven non-verbal mirroring) will increase learners' intrinsic motivation and reduce *frustration*.

The investigation of H3 will proceed with the analysis of differences in participant's *intrinsic motivation* with respect to the separate and collective measures of *intervention type, congruence,* and *sensor driven vs. prerecorded non-verbal interaction*. These three measures describe the ways in which this experiment varied the emotional intelligence of the character. A *composite motivation* scale will also be investigated that will include changes in *goal-mastery-orientation* (differences between the pre-test and the post-test) as an indicator of the "type" of motivation. This test is supported by the close theoretical association between *goal-mastery-orientation* and *intrinsic motivation.*

- H4: Metacognitive and meta affective skill will be exhibited at higher levels when learners interact with emotionally intelligent agents and will positively correlate with perseverance, willingness to continue, and intrinsic motivation.

The analysis of H4 will build from the separate and collective measures of emotional intelligence used in H3. In this analysis these measures will be used to assess their impact on *metacognitive* and *meta affective skill*. The construct of *meta affective skill* will be developed from the *metacognitive* self-report questions, from relevant interactions with the character regarding *willingness to stick with the activity* and *if they would like to try the activity again* and secondarily by investigating the changes in *self-theories*.

Metacognitive and *meta-affective skill* will also be investigated for correlations with the *perseverance* measure and *teacher rated persistence* will be used as a covariate.

Chapter 5

The analytical approach outlined in chapter 4 will be implemented and the results presented in sections 5.1, 5.2, 5.3, and 5.4 corresponding to each of the four hypotheses, H1, H2, H3, and H4. As this analysis is conducted "significance" will be determined at $p <= 0.05$. As is frequently the case in related literature, analysis that finds $0.05 < p <= 0.10$ will be discussed as having "a trend toward significance." Correspondingly analysis that finds $p>0.10$ will be cited as having "no significance". Prior to the direct analysis of H1, H2, H3, and H4 a few general aspects of the data will be discussed.

It should be noted that the values of "negative" variables (e.g. frustration, annoyance, etc.) have been coded so that larger values describe situations that are "more negative" (e.g. greater frustration, more annoyance, etc.).

The duration of the introductory measure varied do to the speed of participant responses. Of the 61 participants included in the analysis, some took up to 1 min longer than the mean of 10.6 min. so this means they were less than 10% longer. This was not considered a reason to exclude any participants.

Of the 61 participants 16 quit by pressing one of the three quit buttons, 10 quit by pressing "I want to stop because I'm too frustrated to continue", 5 quit by pressing "I've put in all the effort that I can and want to stop", and 1 quit by pressing "I want to stop for some other reason"; 12 of the boys quit and 4 of the girls quit.

A 2x2 ANCOVA for *mirroring and intervention* groups shows that the number of participants that *quit* did not differ significantly between the four experimental conditions, nor were the differences significance between the *mirroring* or *intervention* conditions.

5.1: H1 Results:

- H1: The affective learning companion is expected to be more persuasive, and users will form a stronger social bond with the affective learning companion, when sensor-driven non-verbal mirroring informs the affective learning companion's interactions.

This hypothesis deals with differences in *persuasion* and *social bond* due to *mirroring.* The relevant variables, measures, and scales for each of these constructs will now be discussed. The analysis of *persuasion* will be presented in section 5.1.1 and the analysis of *social bond* will be presented in section 5.1.2.

5.1.1. Persuasion Measures and Non-Verbal Interactions:

To assess the effects of *persuasion* the various opportunities and their sub-measures will be discussed as they relate to persuasion. The assessment of an overall scale of *persuasion* is not appropriate for the existing measures, since individually these measures do not attempt to measure the same construct (some measure frustration while others measure participant's beliefs regarding efficacy and their expectation for the use of strategies with respect to proceeding with the activity). Therefore assessing their reliability, i.e. their internal consistency -- collective ability to correlate to a hypothetical "true" measure of *persuasion,* is not a valid analytical approach. For example, none of them measure the level of "convincing" that has taken place relative to some prior viewpoint. Instead of developing a scale, the approach here is to examine multiple persuasion-related measures, such as those employed in the pioneering work of Bailenson that showed mirroring effects on persuasion (Bailenson and Yee 2005). While Bailenson used only post-test survey instruments to measure participant's agreement with the agent's message the approach here includes that approach (e.g. through the *post test of self-theories* instruments, which are scales in their own respect) but goes beyond that approach by also using [extending a richer multi-facetted approach to including] several measures of response and behavior throughout the actively. These opportunities and their sub-measures are assessed individually and the collective differences, with respect to *mirroring*, are discussed. In contrast to assessing inter-measure reliability this approach will assess the effects of the treatment on a variety of different measures that are thought to be influenced by the *persuasive* qualities of the *mirroring* treatment.

1. The first opportunity to assess *persuasion* is to use several of the questions that the characters asked participants throughout the introduction, before the task (Table 5.1.1) and to assess these for differences with respect to the *mirroring* condition. Differences in *persuasion* by *gender* are also assessed. All of the topics in these questions were addressed in the initial presentation. The first question was asked near the beginning of the slideshow so there was little chance for the *persuasive* effects of *mirroring* to have had much effect at that time. The second was asked toward the middle of the slide show, so approximately 5 minutes of *mirroring* interactions had taken place. Questions 3, 4, and 5 were asked after the slide show, before the activity, so a total of approximately 8 to 10 minutes of *mirroring* interactions had taken place. The overall quality, in terms of the types of interactions, frequency and magnitude of the *mirroring* interactions, was similar between the two groups; the primary difference was that the *sensor driven non-verbal interaction* condition received real-time interactions, while the *pre-recorded non-verbal interaction* condition received interactions driven by a data file that had been selected to be typical of the interactions of most participants in pilot study 3 (see chapter 3 for a discussion of the selection of the files for the *pre-recorded non-verbal interaction* condition).

1. What do you think about intelligence, do you think it can change? 1 - I think you can't really change your basic intelligence 7 - I think you can change your intelligence a lot
2. On a scale from 1 to 7, how hard do you think it'll be for you to do new things if you get frustrated? 1 - The hardest thing in my life 7 - The easiest thing in my life
3. What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort? 1- No, I do not think it helps 2- It would probably not help very much 3- It would probably help some 4- Yes, I think it helps
4. If things get hard for you in this activity, do you think you will be able to try different ways of doing the activity? 1 - No, I think it will be very hard 7 - Yes, I think it will be easy for me
5. On a scale from 1 to 7, how hard do you think this activity will be for you? 1 – Not very hard at all 7 – Extremely hard

Table 5.1.1. Persuasion related questions asked by the character during the introduction, responses questions 3, 4, and 5 are reverse coded from their onscreen presentation

There are no significant differences for questions 1, 2, 4, or 5 between the *mirroring* or *gender* groups. Regarding question 3, while the differences between *mirroring* groups for this question 3 are not significant the *gender* differences are significant ($p = 0.03$, $F = 4.9$, mean for *boys* = *3.3,* mean for *girls* = 3.7). This indicates that *boys* were less likely to agree that knowing that the mind is like a muscle will be helpful to them. Each of the above questions was reanalyzed to add *pre-test self theory of intelligence* as a covariate. The general results for significance were unchanged. Likewise including *pre-test goal mastery orientation* did not change the results.

2. The second opportunity to assess *persuasion* is to investigate differences between *mirroring* and *gender* groups for the questions asked during the intervention by the character (Table 5.1.2.), controlling for the *intervention* received, for their *frustration* response, and their *effort* response, as covariates. The intervention occurred four minutes after the start of the activity. Since it is known

that affect can bias self response questions, controlling for *frustration* and *effort* is an attempt to separate the variance due to this bias from the variance due to *persuasion.*

Covariate. On a scale from 1 to 7, how frustrated are you feeling right now? (*frustration* response) 1= Absolutely not frustrated at all 7= This is one of the most frustrating times I have ever felt while using a computer
Covariate. How much effort have you been putting into this activity? (*effort* response) 1= Absolutely no effort at all 7= An enormous amount of effort
1. Do you think that you will be able to try these strategies? (*strategies*) 1 = No, I do not think I can 7 = Yes, I think I can *These strategies refer to the strategies presented in the intervention. Task based strategies were presented in the task support intervention and affect based strategies were presented in the affect support intervention.*
2. How do you feel about continuing the activity? (*stick with it*) 1 = I am not at all willing to stick with it 7 = I am very willing to stick with it

Table 5.1.2. Questions asked by the character during the intervention in both conditions.

There is no significant difference in responses to the *strategies* question between the *mirroring* groups. However, there is a significant difference between *gender* groups (p=0.003, F =9.4, mean of *boys* = 4.8, mean of *girls* = 6.1). The reverse-coded anchor points of the 7 point Likert scale are "1. No, I do not think I can" and "7. Yes, I think I can". *Girls* indicated they would be able to use the strategies presented in the intervention to a greater extent than boys.

There is no significant difference in responses to the *stick with it* question for the *mirroring* or *gender* groups.

3. The third opportunity to assess *persuasion* tests for differences in *the character was easy to understand* across *mirroring* groups. No significance was found.

4. The fourth opportunity to assess *persuasion* tests for differences in *perseverance* as a behavioral result of the persuasive message, i.e. how long did the students actually engage in the task begun after the introductory message. This is possible since the character's message attempts to persuade, and provide strategies to help, participants to persevere. This means of assessing persuasion through *perseverance* does not show significant differences for *mirroring.* There are significant differences in *perseverance* across *gender* groups (p=0.016, F = 6, mean for *girls*= 14.3 minutes, mean for *boys* = 12.2 minutes). There were a four individuals (discussed in chapter 4) who were excluded from the analysis because they finished or mistakenly thought they had finished (i.e. moved the tower to the middle tower or interpreted the large white disk as a plate rather than a disk, and in these cases were not interested in continuing after the experimenter corrected their impressions). Two were *boys* and two were *girls.*

5. A fifth opportunity to assess *persuasion* tests for differences in responses to the *metacognitive /meta-affective* questions (Table 5.1.3) and differences in the *change in self theories of intelligence* between the *mirroring* and *gender* groups.

Two *metacognitve/meta-affective* scales were developed, as discussed in section 5.1.2. There are no significant differences in the levels of two *metacognitive/meta-affective* scales with respect to *mirroring* or *gender* (p>0.4). There are no significant differences in either of the *change in self theories* measures with respect to *mirroring* and *gender* (p>0.4).

Overall, considering five different measures of persuasion, the hypothesis that mirroring influenced persuasion was not confirmed, additional discussion of these measures can be found in chapter 6.1.

5.1.2. Metacognitive/Meta-affective Scales:

1	How helpful was it to think of your mind as a muscle?	Neither scale 1 or scale 2, and therefore not used.
2	How much of the time did you try to use different strategies to do this activity?	Scale 1
3	If you used different strategies how helpful did you find them?	Scale 1
4	If you used different strategies, how many did you use? *Answer: I used about [drop down box] different strategies.*	Scale 1
5	While you were doing this activity how many times where you aware of your frustration?	Scale 1
6	How much control did you feel you had during this activity?	Scale 2
7	What would you say about your ability to concentrate on this activity?	Scale 2
8	Do you think you have the skills needed to do this activity?	Scale 2

Table 5.1.3. Metacognitive/meta-affective questions, scale 1 is meta-affective skill, scale 2 is Flow/Stuck

Two scales were constructed, *metacognitive/meta-affective scale 1* (alpha = .70), and *metacognitive/meta-affective scale 2* (alpha = .75). While the second scale was designed from measures that were excluded from the first scale it is still interesting to see that the correlation is not significant (p = .665, r= -.059). Since question 4 is a drop down box, allowing answers across a broader range than the 1-7 Likert scale (i.e. 0 to 100) this poses an issue for how to incorporate it in the scale, i.e. how to ensure that its contribution does not swamp the contribution of the other questions in the scale. In this experiment two participants answered that they used zero strategies, and a third who answered 10 was the only participant to answer greater than six. Given this distribution the responses of these three individuals was recoded to 1 and 7 respectively. Once these transformations were made the *metacognitive/meta-affective scale 1* still had a reliability of alpha = 0.7.

5.1.2. Social Bond Measures and Non-Verbal Interactions:

The second component of H1 is *social bond.* As is the case for *persuasion* there are several opportunities to measure the *social bond* and to compare differences across *mirroring* conditions.

There is the opportunity to analyze a quasi-behavioral measure of the *social bond* by comparing the *bye.button* response of the participants. This quasi-behavior measure along with the *social bond* questions are analyzed through the construction of scales. Two scales *character positive* and *character negative* were constructed from the modified Working Alliance Inventory questions (Table 5.1.4.).

The *bye.button* response was coded as an ordinal measure with higher values indicating more positive responses. Six participants pressed "Ok glad you are finally going", 15 pressed "Ok, bye", and 31 pressed "Ok, bye I was glad to have you here". Tables 5.1.5. and Table 5.1.6.show the frequencies of the *bye.button* responses for *mirroring* and *gender* conditions, respectively. Of the 61 participants 47 pressed the bye buttons. Some participants did not select a button. The buttons and the character disappeared from the screen either when a participant presses a button or 20 seconds after the buttons appears, so that they do not further distract participants.

Character Positive **(Alpha = .87)**
I like Casey.
Casey is intelligent.
Casey seemed to understand how I felt.
Casey is helpful.
If I did this again I would want Casey present.
Casey seamed to sense when I got frustrated, even though he couldn't help me.
Character Negative **(Alpha = .80)**
Casey was a distraction to me.
Casey annoyed me.
Casey is dumb.

Table 5.1.4. Questions in the Positive Character and Negative Character scales, all responses were on a 7 point Likert scale with 1= strongly disagree and 7= strongly agree

sensor driven = 1, pre-recorded = 0

bye.val -- goodbye value			Frequency	Percent	Valid Percent	Cumulative Percent
.	Valid	.00	8	57.1	57.1	57.1
		1.00	6	42.9	42.9	100.0
		Total	14	100.0	100.0	
.00	Valid	.00	3	60.0	60.0	60.0
		1.00	2	40.0	40.0	100.0
		Total	5	100.0	100.0	
1.00	Valid	.00	7	53.8	53.8	53.8
		1.00	6	46.2	46.2	100.0
		Total	13	100.0	100.0	
2.00	Valid	.00	12	41.4	41.4	41.4
		1.00	17	58.6	58.6	100.0
		Total	29	100.0	100.0	

Table 5.1.5. Showing the number of participants in the *mirroring* conditions that pressed each of the *bye.button* responses (. = no response, 0 = negative response, 1 = neutral response, 2 = positive response).

m=0, f=1

bye.val -- goodbye value			Frequency	Percent	Valid Percent	Cumulative Percent
.	Valid	0	6	42.9	42.9	42.9
		1	8	57.1	57.1	100.0
		Total	14	100.0	100.0	
.00	Valid	0	3	60.0	60.0	60.0
		1	2	40.0	40.0	100.0
		Total	5	100.0	100.0	
1.00	Valid	0	10	76.9	76.9	76.9
		1	3	23.1	23.1	100.0
		Total	13	100.0	100.0	
2.00	Valid	0	13	44.8	44.8	44.8
		1	16	55.2	55.2	100.0
		Total	29	100.0	100.0	

Table 5.1.6. Showing the number of participants by *gender* (m/f – *boys/*girls) conditions that pressed each of the *bye.button* responses (“.” = no response, 0 = negative response, 1 = neutral response, 2 = positive response).

5.1.2.1. The relationship between *bye.button* response and *social bond scales:*

Character positive, character negative, and *bye.button* response all correlated significantly with each other (see Table 5.1.7), with the lowest R Squared value = .22. The responses to these measures do not have sufficient inter-measure reliability to be used as a scale.

		qc.pos - character positive	qc.neg - character negative	bye.val -- goodbye value
qc.pos - character positive	Pearson Correlation	1.000	-.720**	.687**
	Sig. (2-tailed)	.	.000	.000
	N	59	59	46
qc.neg - character negative	Pearson Correlation	-.720**	1.000	-.472**
	Sig. (2-tailed)	.000	.	.001
	N	59	60	47
bye.val -- goodbye value	Pearson Correlation	.687**	-.472**	1.000
	Sig. (2-tailed)	.000	.001	.
	N	46	47	47

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5.1.7. Correlations of *Character positive, character negative*, and *bye.button* response

ANOVA analysis indicates that levels of *character positive* are significantly different ($p<0.001$, $F=27.2$, Eta Squared = .56) for participants grouped using their *Bye.button* selection. Eta Squared = .56 indicates that it can reasonably be expect that 56% of the variance in the *character positive* measure in the generalized population should be explained by the *bye.button* response. Games-Howell post-hoc analysis, which is particularly suitable for analysis with different numbers of participants per cell, indicates that participants with negative *bye.button* response differed significantly ($p<0.001$) from those with neutral and positive responses, while those with neutral and positive responses did not differ significantly from one another (mean negative = 17.4, mean neutral = 30.2, mean positive = 33.2; these values are the sum of the 6 responses in the *character positive* scale in Table 5.1.4., since they are Likert scale 1-7 responses possible values range from 6 to 42).

ANOVA analysis indicates that levels of *character negative were* significantly different ($p<0.001$, $F=0.2$, Eta Squared= .30) for participants grouped using their *Bye.button* selection. Eta-squared of .30 indicates that it can reasonably be expect that 30% of the variance in the *character negative* measure in the generalized population should be explained by the *bye.button* response. Games-Howell post-hoc analysis indicates that participants with negative *bye.button* response differed significantly ($p<0.001$) from those with neutral and positive responses, while those with neutral and positive responses did not differ significantly from one another (mean negative = 15, mean neutral = 8.9, mean positive = 8.0; these values are the sum of the 3 responses in the *character negative* scale in Table 5.1.4.: since they are Likert scale 1-7 responses, the possible values range from 3 to 21.

Age, bye.button response and *social bond:*

Bivariate correlation shows that age correlates significantly with *character positive* $r= -.34$, *character negative* $r=.30$, and *bye.button* $r=-.29$. These correlations indicate that older participants have a less positive impression and less positive quasi-behavioral *bye.button* response to the character than younger participants.

The relationship between *social bond, mirroring* and *intervention*

The three measures of *social bond*: *bye.button* response, *character positive,* and *character negative* were analyzed for differences with respect to *mirroring.*

Bye.button response:

There were no main effects for difference in the mean value of *bye.button* response for the 2 x 2 x 2 ANCOVA (*mirroring x intervention x gender).* The interaction of *intervention x gender* was significant (p= 0.005, F= 10.7). The mean response for *boys* with *task support* was between "Ok, bye" and "Ok, bye I was glad to have you here" while the mean for *boys* with *affect support* was the less positive, "Ok, bye" response. *Girls* with *task support* were only slightly lower than *boys* in this condition, somewhere between "Ok, bye" and "Ok, bye I was glad to have you here", but in contrast to the *boys* in the *affect support* condition the *girls* response in the *affect support* was the more positive, "Ok, bye I was glad to have you here," response.

To follow up on these differences, the mean *bye.button* responses show a trend toward significant difference (p = 0.08) for *gender.* Boys responded slightly more negatively than girls, when analyzed across both *mirroring* conditions.

Repeating the above analysis to include the covariates of *age* and *school* showed significant differences for *gender* with *girls* being more positive (p = 0.046, F=4.3, Eta(2) = .096). The interaction for *intervention x gender* was still significant (p = 0.01, F=7.1, Eta(2) = .151).

Character Positive:

There were no main effects for difference in the mean value of *character positive* for the 2 x 2 x 2 ANCOVA (*mirroring x intervention x gender)* with *age* and *school* as covariates. It was found that only the interaction of *intervention x gender* was significant (p= 0.024, F = 5.4, Eta(2) =.1). Without the covariates this interaction was more significant (p=0.007, F = 7.9, Eta(2)= .14). *Boys* have more positive impressions of the character that provides *task support* than the character that provides *affective support*, while *girls* have the opposite response.

Character Negative:

There were no main effects nor interactions for difference in the mean value of *character negative* for the 2 x 2 x 2 ANCOVA (*mirroring x intervention x gender)* with and without *age* and *school* as covariates.

With respect to the H1 hypothesis, *mirroring* did not significantly effect the measured aspects of *social bond.* It was, however, found that *intervention* had a different impact on boys than girls with respect to the development of *social bond.* The participants of this study were 11-13 years old. Ways in which the age and gender of the participants might have effected the results will be further discussed in 6.1 and 6.5.

The relationship between *character's emotional intelligence* and *social bond*:

The *character's emotional intelligence* is quantified to some degree using the *mirroring, intervention,* and *congruence* measures. *Social bond* has already been analyzed with respect to *intervention* and *mirroring* in the previous section. This leaves *congruence* and an overall measure of the *character's emotional intelligence* to be developed and analyzed. Multiple ways of measuring congruence are described below.

Congruence based on an Affective Classifier:

The values provided by the classifier trained on pilot data were investigated and it was found that none of the participants were categorized as being frustrated or likely to seek help. There was also very little distribution in the range of values provided (SD = 0.07). The use of the classifier measures was therefore abandoned in this analysis. The likely explanation for the problem is that the data was collected under different conditions, and contained varying ranges of noise. These problems could be addressed with future work that cleans the data and retrains the classifier algorithms under the new experimental conditions.

Congruence based on self-report of frustration at the time of intervention:

A measure of *congruence* was calculated by scaling an interim variable for the *frustration* response, that participants provided to the character at the time of the intervention, so that -1= "Absolutely not frustrated at all" 1= "This is one of the most frustrating times I have ever felt while using a computer.. A second interim variable was developed recoding the *intervention* condition so that *task support* = -1 and *affect support* = 1. [1] The two interim variables are multiplied by each other as summarized in Table 5.1.8. The new variable ranges from -1 = low congruence to 1 = high *congruence.* Recoding this variable into a binary variable by applying a mean-split (mean = -.14) provides the *congruence* variable, with low *congruence* = 0 and high *congruence* = 1, that is used in this analysis.

frustration (high =1)	affect support = 1	product = 1 (high congruence)
frustration (high =1)	task support = -1	product = -1 (low congruence)
frustration (low =-1)	affect support = 1	product = -1 (low congruence)*
frustration (low =-1)	task support = -1	product = 1 (high congruence)

Table 5.1.8. Calculating *congruence* through the multiplication of the interim *frustration* and interim *intervention,* the product of which is the interim *congruence* variable.

Differences in *social bond* due to *congruence* were analyzed. No significant difference for *bye.button* response was found between *congruence* levels; this was analyzed with and without the covariates of *age* and *school.* There was also no significance (i.e. $p > .1$) in the relationship between *congruence* and *character positive* or *character negative.*

Further exploration of character's emotional intelligence:
The three measures of *character's emotional intelligence (intervention, congruence,* and *mirroring)* were regressed in separate linear regression (without covariate variables) to determine if they predict the levels the three separate elements of *social bond (bye.buttons, character positive,* and *character negative).* There were no significant predictors of the *social bond.*

Overall the preceding analysis, considering three *social bond* measures --two scales (*character positive* and *character negative*) and one quasi-behavioral measure (*bye.button* response), does not support the hypothesis that mirroring influenced social bond. Additional discussion of these measures can be found in chapter 6.1.

5.2. H2 Results

- H2.A: A learner's social bond with an affective learning companion will positively correlate with his or her perseverance and self-theories – adoption of internal beliefs that he or she can increase his or her own intelligence and the adoption of mastery orientation.

- H2.B: The level of persuasion a learner experiences from the affective learning companion's metacognitive/meta-affective message will positively correlate with the social bond, with perseverance, and will negatively correlate with frustration.

This section will have two subsections to test the two parts of H2.

In section 5.2.1 the evaluation of H2.A will be presented, testing two sets of correlations:
1) social bond will positively correlate with perseverance
2) social bond will positively correlate with self-theories

[1] While *affect support* adapted to the level of frustration or non-frustration expressed by the learner, it was still considered "low congruence" in the case when the learner is absolutely-not-frustrated-at-all and was interrupted with a message saying "I am glad this activity is not making you feel that way…." (see Appendix C for the complete *affect support* character dialogue).

In section 5.2.2 the evaluation of H2.B will be presented, testing three sets of evaluations:

1) persuasion will positively correlate with social bond
2) persuasion will positively correlate with perseverance
3) persuasion will negatively correlate with frustration.

5.2.1. H2 A Results

5.2.1.1. Results for social bond will positively correlate with perseverance

Using partial correlation to account for the covariates of *age* and *school,* there was no significance found for the correlation of *social bond* measures with *perseverance.* Likewise a bivariate correlation for the *social bond* measures with *perseverance* showed no significance. The *social bond* measures were (*character positive, character negative,* and *bye.button* response). *Perseverance* was the time from the beginning of the activity until they pressed a *stop button* or until they had *persevered* for 15 minutes. The range in perseverance was from immediately after the character said good-bye at to the full 15 minutes.

5.2.1.2. Results for social bond will positively correlate with self-theories

The *self-theories* measures were obtained in the pre-test and post-test on-screen surveys. (The character was not visible during these surveys.) The partial correlation (controlling for *age, school, intervention, mirroring,* and *gender* as covariates*)* of *changes in goal mastery orientation* and *changes in self theory of intelligence* with *character positive* and *character negative* scales was not significant. Nor was the correlation of the *changes in self theories* measures with the *bye.button* response, controlling for the same variables. There was a trend toward significance in the *changes in self theories of intelligence* with the *bye.button* response ($p = .09$, $r=.21$). This trend indicates that participants that had more positive *bye.button* responses also had positive *changes in self theories of intelligence*. Splitting the participants by *gender* to further investigate the *bye.button* response correlation with *changes in self theories of intelligence* indicated no significance for *girls* ($p=.294$, $r=.15$) and a trend toward significance for *boys* ($p = .089$, $r=.24$) indicating that boys that responded more positively to the character also had larger *changes in self theories of intelligence* than boys that responded more negatively. Without the covariates of *age, school, intervention, mirroring* and *gender*, correlating the measures for social bond (*bye.button, character positive,* and *character negative*) with the *change in self theories* measures shows no significant correlation for *social bond* with *self-theories.* The trend toward significance in the *changes in self theories of intelligence* with the *bye.button* response increased [with] ($p = 0.07$, $r = .22$).

To summarize the results of the H2 A analysis, as presented above in 5.2.1.2, there is no significant correlation between *perseverance* and *social bond.* As presented in 5.2.1.1, there is no significant correlation between *changes in self theories* and *social bond.* There was a trend toward significance ($p = 0.07$, $r = .22$) in the *changes in self theories of intelligence* with the *bye.button* response.

5.2.2. H2 B Results

5.2.2.1. Results for persuasion will positively correlate with social bond

Since the fourth and fifth opportunities for assessing *persuasion,* described in 5.1, include the two measures used in H2A, *perseverance* and *changes in self theories,* this part of the analysis of H2B is partially redundant with the analysis of H2A found in section 5.2.1. The redundancy results from the interest of this thesis research in both *social bond* and *persuasion* and from the diverse opportunities that have been structured to analyze these two constructs. At times these different opportunities use the same measures. Since there is no overall scale for *persuasion* nor *social bond,* per section 5.1.1 and 5.1.2 respectively, a partial correlation of the *persuasion* measures with the three *social bond* measures is conducted. Before presenting the new analysis the results from 5.2.1.2 and 5.2.1.1 for *perseverance* and *changes in self theories* are summarized.

The first opportunity to measure *persuasion* is to correlate the responses to five of the questions asked during the introduction (see Table 5.1.1. for theses questions) with social bond.

The following results in this section use partial correlations to control for *age,* and *school.*

1. Responses to the question, "What do you think about intelligence, do you think it can change?", asked during the intro, did not significantly correlate with *character positive* or *character negative.* Correlation with *bye.button* response showed a trend toward significance (p=0.084, r = .20), with greater belief in change accompanying more positive choice of *bye.button* response

2. Responses to the question, "On a scale from 1 to 7 , how hard do you think it'll be for you to do new things if you get frustrated?", asked during the intro, showed no significant correlations with any of the three *social bond* measures.

3. Responses to the question, "What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort?", asked during the intro, show a trend toward significant correlation with *bye.button* response (p = .054, r = .24), and correlate significantly with *character positive* (p=.001, r= .42), and *character negative* (p=.004, r= -.34).

4. Responses to the question, "If things get hard for you in this activity, do you think you will be able to try different ways of doing the activity?", asked during the intro, correlate with *bye.button* (p=.018, r=-.33), but not with *character positive*, or *character negative*.

5. Responses to the question, "On a scale from 1 to 7, how hard do you think this activity will be for you?", asked during the introduction, correlates significantly with *character negative* (p=.025, r=.27), but not with *character positive* or *bye.button* response. The harder they thought it would be, up front, the more they reported disliking the character.

The second opportunity to measure *persuasion* is to correlate the responses to questions asked during the intervention (see Table 5.2.1. below) for these questions with social bond
Controlling for the *frustration* and *effort* responses, the responses to the question, "Do you think that you will be able to try these strategies?", asked during the *intervention* (different *strategies* were presented in each *intervention*) were significantly correlated to *character positive* (p=.031, r=.25) and show a trend toward significant correlation with *character negative* (p=.084, r= -.19). Correlation with *bye.button* response was significant (p=.038, r= .27). Bivariate correlations are also significant for *character positive* (p=.01, r=.30), *character negative* (p = .029, r=-.25), *and bye.button* response (p = .048, r = .25), respectively.

Controlling for frustration *and* effort *responses, the responses to the question, "How do you feel about continuing the activity?", asked during the* intervention *showed no significance for any of the measures of* social bond. *.Bivariate correlations also show no significance.*

The third opportunity to measure *persuasion* is to correlate if the *character was easy to understand* with social bond. Responses to the question, "It was easy for me to understand what Casey was saying." were significantly correlated with *character positive* (p=.003, r= .41) and *character negative* (p<.001, r= -.45), and not significantly correlated with *bye.button* response.

To summarize these findings there is minimal to mixed support across the correlations of the measures of *social bond* with the opportunities to measure *persuasion.*

5.2.2.2. Results for persuasion will positively correlate with perseverance

The discussion of the five opportunities to measure *persuasion,* found in 5.1.1, includes assessing *perseverance* as the fourth. That fourth measure (of how long they engaged in the task the first time) was not found to have a significant correlation with persuasion as a condition of mirroring, but it was found that girls persevered more than boys. The analysis in this section correlates the remaining measures of *persuasion* with *perseverance,* where perseverance measured the number of minutes from the time the activity started until a *stop button* was pressed, or until they persevered for 15 minutes.

1. What do you think about intelligence, do you think it can change?	(p > .1)
2. On a scale from 1 to 7, how hard do you think it'll be for you to do new things if you get frustrated? Participants that thought it would be hard for them to do new things *tend to* also persevered less.	(p=.097, r= -.17)*
3. What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort? Participants that agree tend to also *persevere* more.	(p=.025, r=.25)**
4. If things get hard for you in this activity, do you think you will be able to try different ways of doing the activity?	(p>.1)
5. On a scale from 1 to 7, how hard do you think this activity will be for you? Participants that thought the activity would be hard tend to also *persevere* less	(p=.037, r=-.24)*

Table 5.2.1. Results of the first opportunity to measure *persuasion* (i.e. to correlate the responses to five of the questions asked during the introduction to *perseverance*) (* indicates a trend toward significance, p<.1; ** indicates significance p<.05)

The second opportunity to measure *persuasion* is to correlate the responses to questions asked during the intervention (see Table 5.1.1. for these questions) to *perseverance*

Controlling for the *frustration* and *effort* responses, the responses to the question, "Do you think that you will be able to try these strategies?" asked during the *intervention* (different *strategies* were presented in each *intervention*) show significant correlation with *perseverance* (p=.002, r=.38).

Controlling for the *frustration* and *effort* responses, the responses to the question, "How do you feel about continuing the activity?" showed no significant correlation with *perseverance..*

The third opportunity to measure *persuasion* is to correlate if the *character was easy to understand* with *perseverance.* Responses to the question, "It was easy for me to understand what Casey was saying." were not significantly correlated with *perseverance.*

The fourth opportunity to measure *persuasion* is to assess *perseverance.* As noted in the introduction to this section, the current focus of the analysis is *perseverance,* and as such *perseverance* is implicitly self-correlated.

The fifth opportunity to measure *persuasion* is to correlate the *metacognitive/meta-affective* scales and *change in self theories* to *perseverance.*
Perseverance does not correlate significantly with either of the *metacognitive/meta-affective* scales nor with *changes in self theories*. There is a trend toward significance in the correlation with *metacognitive/meta-affective scale 1* (p=.092, r=.18).

To summarize these findings there is minimal to mixed support across the correlations of the measures of *perseverance* with the opportunities to measure *persuasion.*

5.2.2.3. Results for persuasion will negatively correlate with frustration.

Frustration Scale
There are several measures that can be used to assess frustration, these include: the *frustration* response at the time of *intervention,* the pressing of one of three flavors of *quit button*, self-report questions from the post activity questions (see Table 4.4.), *perseverance*, and *perceived duration* (recoded as *overestimation*).

The reliability (alpha = .59) across all of these measures is low. Removing *perseverance* from the scale yields a scale with sufficient reliability (alpha = .75). While the reliability could be improved by eliminating additional measures, doing so would reduce the number of non-self-report items. For this reason the analysis precedes using this scale, *frustration scale 1.*

To compute the scale requires that two of the measures be multiplied by constants, so that their relative contributions to the final scale will be equivalent. This is allowable since multiplying the individual measures of a scale by a constant does not alter the standardized reliability of the scale. The binary *quit button* measure is multiplied by the constant 7.0 so that the new values are either 0 or 7. The range in *overestimation* is from 0.33 to 2.87; multiplying this measure by the constant 3.0 creates a new range of .99 to 8.61. This constant is considered reasonable since it leaves only one participant below 1.0 and only two participants that exceed 7.0. Due to the similarity in the post activity self report questions (alpha = .82, see Table 5.2.2.), and the desire not to overwhelm the other measures in the overall frustration scale, the contribution of each of these question was divided by the number of questions (4 in this case) so that the group of questions, taken together, would have an equivalent influence on the overall scale as each of the other measures in the scale.

Mark how much of the time you were frustrated.
Mark how frustrated you were at the most frustrated time.
If you clicked on one of the buttons, mark how frustrated you were at that time.
This activity was too hard.

Table 5.2.2. Post activity self-report question that relate to *frustration* (alpha = .82)

In order to conduct analysis of H3 related to the *intervention* and the *congruence* components of *emotional intelligence* a second scale, *frustration scale 2* was created (alpha = .72) that excludes *frustration* response at the time of intervention from the *frustration scale 1.* The reason that the analysis of *congruence* uses the second measure is that the *congruence* measure was created based, in part, on the *frustration* response measure, which is part of *frustration scale 1*; *Frustration scale 2* was created by removing this measure from *Frustration scale 1.* Finally, a new binary variable, *frustration quartiles,* coded participants into the highest and lowest quartiles of the *frustration scale 1.*

frustration response (at the time of intervention)	scale 1
Mark how much of the time you were frustrated.	scale 1 and 2 (1/4 weight)
Mark how frustrated you were at the most frustrated time.	scale 1 and 2 (1/4 weight)
If you clicked on one of the buttons, mark how frustrated you were at that time.	scale 1 and 2 (1/4 weight)
This activity was too hard.	scale 1 and 2 (1/4 weight)
quit.button (if the button was pressed or not)	scale 1 and 2
overestimation (perceived duration/actual duration)	scale 1 and 2
Perseverance (time from the start of the activity until the quit button or until the 15 minutes elapses)	scale 1 and 2

Table 5.2.3. Measures included in frustration scale 1 (alpha = .75) and frustration scale 2 (alpha = .72)

Checking for random assignment with respect to frustration

Since the intervention does not differ prior to the *frustration* response question the response should not be different between the *intervention* groups. This difference was tested and the result is of some concern since there is a trend toward a significant difference (p=.083, F=3.1) with covariates of *age* and *school* it is also significant (p = 0.062, F=3.7, mean for *affect* = *2.5,* mean for *task* = 3.3). Similar tests for differences in *frustration* at the time of *intervention* were also conducted to test for effect of *mirroring* and *gender* at this time and no significance was found.

Similar analysis was conducted for *effort* response (*mirror x intervention x gender*). There was a main effect for *gender* (p=.042, F=4.3, Eta-squared = .076, *boys* mean *effort* response = .56 vs. *girls* mean effort *response* = .49). *Boys* self-report using more *effort* than *girls.* Running the same analysis with *mirroring* as a covariate (to the extent mirroring has an effect, it should have had some of its effect by the time the *effort* response question is asked, and should therefore be controlled for), is still significant (p=0.05, F=4.0, Eta-squared = .067).

The relationship between self reported frustration and overestimation

Overestimation was assessed for its ability to predict frustration (see chapter 2 for a discussion of related research). The overestimation measure is a percentage obtained by dividing the self reported perceived duration by the participant's persistence (the actual duration form the start of the activity until they either hit a quit button or until the 15 minute time period has elapsed). The minimum value was 33% and the maximum was 287% with a mean value of 112% and a SD = 53%. Controlling for *frustration* response at the time of intervention, there were significant differences found (p=.029, F=5.353) between participants in the lowest and highest *frustration quartiles* (mean low frustration = 73%, mean high frustration = 145%). Repeating the analysis without *frustration* response is still significant (p= 0.037, F = 4.8).

Presentation of the results for persuasion will negatively correlate with frustration.

The first opportunity to measure *persuasion* is to correlate the responses to five of the questions asked during the introduction (see Table 5.1.1. for theses questions) to *frustration.*

The following results in this section use partial colorations to control for *age,* and *school.*

1. Responses to the question, "What do you think about intelligence, do you think it can change?", asked during the intro, did not significantly correlate with *frustration.* No significant correlation with *frustration quartiles* was found.

2. Responses to the question, "On a scale from 1 to 7, how hard do you think it'll be for you to do new things if you get frustrated?", asked during the intro, showed no significant correlations with *frustration scale 1.* No significant correlation with *frustration quartiles* was found.

3. Responses to the question, "What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort?", asked during the intro, show a trend toward significant correlation with *frustration scale 1* (p=.087, r= -.19). This indicates that individuals who found this *metacognitve/meta-affective* message helpful also experienced lower levels of *frustration* throughout their experience. No significant correlation with *frustration quartiles* was found.

4. Responses to the question, "If things get hard for you in this activity, do you think you will be able to try different ways of doing the activity?", asked during the intro, showed no significant correlation with *frustration scale 1.* No significant correlation with *frustration quartiles* was found.

5. Responses to the question, "On a scale from 1 to 7 , how hard do you think this activity will be for you?", asked during the intro, correlates significantly with *frustration scale 1* (p=.021, r=.28). This

question showed significant correlation with *frustration quartiles* (p=.045, r=.33). Participants that expected the activity to be harder also found it to be more frustrating.

The second opportunity to measure *persuasion* is to correlate the responses to questions asked during the intervention (see Table 5.1.2. for these questions) to *frustration*.

Controlling for the *frustration* and *effort* responses, the responses to the question, "Do you think that you will be able to try these strategies?" asked during the *intervention* (different *strategies* were presented in each *intervention*) show significant correlation with *frustration scale 1* (p=.005, r= -.37). Without the covariates of *frustration* and *effort* responses there was still significance (p=.017, r=-.29). *Frustration quartiles* also showed significant correlation (p=0.034, r=-.37) with *strategies*. Participants that were less *frustrated* were more likely to indicate that they felt they could use the strategies.

Controlling for the *frustration* and *effort* responses, the responses to the question, "How do you feel about continuing the activity?" showed no significant correlation with *frustration scale 1*. Without the covariates of *frustration* and *effort* responses there was still no significance. *Frustration quartiles* showed no significant correlation.

The third opportunity to measure *persuasion* is to assess if the *character was easy to understand*. Responses to the question, "It was easy for me to understand what Casey was saying." were not significantly correlated with *frustration scale 1* or *frustration quartiles*.

The fourth opportunity to measure *persuasion* is to correlate *perseverance* with *frustration*. *Perseverance* correlates significantly with *frustration scale 1* (p<.001, r= -.71) and with *frustration quartiles* (p<.001, r=.89).

The fifth opportunity to measure *persuasion* is to correlate the *metacognitive/meta-affective* scales and *change in self theories* to *frustration*. *Frustration scale 1* and *frustration quartiles* do not correlate significantly with either of the *metacognitive/meta-affective* scales or with *changes in self theories*.

To summarize these findings there is minimal to mixed support of a negative correlation of the measures of *frustration* with the opportunities to measure *persuasion*.

5.3. H3 Results:

- H3: An affective learning companion that exhibits emotional intelligence (active listening, appropriate interventions, and sensor-driven non-verbal mirroring) will increase learner's intrinsic-motivation and reduce frustration.

This section makes use of the *emotional intelligence* measures discussed in 5.1.2 and the *frustration* measures discussed in 5.2.2.3. After discussing the measures of *motivation in* 5.3.1 this section will assess the following two questions in sections 5.3.2 and 5.3.3 respectively:

1) Does *emotional intelligence (intervention, congruence,* and *mirroring*) predict *motivation?*
2) Does *emotional intelligence* (*intervention, congruence,* and *mirroring*) predict *frustration?*

5.3.1. Measures of Motivation:

There are several measures for motivation and their preparation and development will be discussed here. In this chapter and the next the term *motivation* will be used to describe the collective measures that relate to motivation and which are therefore used to investigate a motivation scale. The term *intrinsic-motivation,* will be hyphenated, and will be used specifically to describe the measure that encodes the participant's reengagement in the activity.

The *intrinsic-motivation* variable was coded as a binary variable to indicate if any moves or no moves were made by participants (moves = 1, no moves = 0) during the two minute waiting period designed to see if participants would reengage in the activity. For this measure out of the 61 participants included there were 57 valid recordings in the data set. For a few participants there was a failure in the system after the activity but prior to the waiting period. There were 27 participants that did not reengage with the activity and 30 participants that moved at least one disk. The mean number of moves for those with *intrinsic-motivation=1*was 7.57 moves with SD = 4.43. A new variable was coded based on this mean split to assess differences for participants that exhibit *low intrinsic-motivation,* 1 to 7 moves, versus those that exhibit *high intrinsic-motivation,* 8 to 18 moves.

A composite scale of *motivation* was investigated. In addition to *intrinsic-motivation* the *pre and post test goal-mastery-orientation* measures were used to obtain the *change in goal mastery orientation* as an additional measure of *motivation.* This measure is an indicator of the "type" of motivation, either goal oriented e.g. the performance goal of finishing the task, or mastery oriented e.g. to learn and master the skills or activity.

Self report variables for motivation include: *strategy, stick with it,* and *like to try again,* as presented in Table 5.3.1. The reliability between these variables (alpha = .46) is insufficient to support the use of these measures as a scale; the measures will therefore be used individually.

Do you think that you will be able to try these strategies? 1 = No, I do not think I can 7 = Yes, I think I can	*strategy* asked during intervention
How do you feel about continuing the activity? 1 = I am not at all willing to stick with it 7 = I am very willing to stick with it	*stick with it* asked during intervention
I would like to try this activity again. 1 = Strongly disagree 7 = Strongly agree	*like to try again* asked during post activity questions

Table 5.3.1. Three self-report questions and the corresponding variable names used to measure motivation

Another composite *motivation* scale was also investigated using: *intrinsic-motivation, change in goal-mastery-orientation, strategy, stick with it,* and *like to try again.* No scale was developed as the reliability (alpha = .33) is insufficient.

5.3.2 Does *emotional intelligence (intervention, congruence,* and *mirroring*) predict *motivation?*

This question is assessed using the previously discussed (see section 5.3.1) motivation measures (*intrinsic-motivation, change in goal-mastery-orientation, strategy, stick with it,* and *like to try again*). The covariates of *age, school,* and *pre-test self theories* are used.

Intervention: there is no significant difference in any of the *motivation* measures between the two *intervention* groups. When *intervention* was assessed with *gender* most of the measures were not significant, however there was a main effect significance for differences in beliefs about *strategies* (p=.041, F=10, mean *girls* = 6.0, mean *boys* = 5.1. This indicates that *girls* report that they believe they will be able to use the *strategies* provided in the intervention, more than *boys* believe this. (The difference between this result and the similar analysis in 5.1.1., which also found significant gender differences with respect to *strategies,* is in the covariates that are used – *frustration* response and *effort* response in 5.1.1. and *age, school,* and *pre-test self theories* here in 5.3.)

Congruence: there is no significant difference in any of the *motivation* measures between the mean-split *congruence* groups.

Mirroring: there is no significant difference in any of the *motivation* measures between the *mirroring* groups. There is a trend toward significance for the *stick with it* measure (p=.07, F=4.0). Girls with *mirroring* reported (mean = 6.3) slightly higher willingness to *stick with it* than *girls* without mirroring (mean = 6.0). In contrast, boys with mirroring were quite a bit more willing to *stick with it* (mean = 6.4) than *boys* without mirroring (mean = 5.2).

Following up this trend with an ANCOVA (using *age, school, pretest self theories of* intelligence as covariates) on each of the *intrinsic-motivation measures* showed no significant differences.

To summarize these findings there is minimal for *emotional intelligence* (*intervention, congruence,* and *mirroring*) predicting *motivation.*

5.3.2. 2) Does *emotional intelligence* (*intervention, congruence,* and *mirroring*) predict *frustration?*

Intervention:

The following analysis will use *frustration* response, the self report measure of frustration provided by participants at the beginning of intervention, as a covariate. The importance of using this covariate is when subsequent measures of frustration (such as *frustration scale* 2) are analyzed the assessments will be of the "changes in frustration" from the time of intervention. This enables the analyses to attribute the differences to the effects that the *intervention* has on changing participant's levels of frustration.

Using *frustration* response as a covariate, *Frustration scale 2,* the scale that excludes *frustration* at the time of *intervention*, shows no significant differences across *intervention.* Further, comparing levels of *frustration scale 2* for *mirroring, intervention*, and *gender* and any interactions between these shows that the only differences that are significant are for *gender* (p=.010, F= 7.0, Eta(2) = .13, *boys* mean = 10.5, *girls* mean = 7.9). This indicates that even controlling for the level of frustration at the time of intervention, *girls* are less frustrated than *boys* at the end of the activity. No other significance or trend was found. Including *pre-test self theories of intelligence* as a covariate did not change the results.

Congruence

Since *congruence* is a measure that is created with respect to differences in *frustration* with respect to *intervention* condition it is inappropriate to test for significant differences in *frustration* in the *congruence* measure at the time of *intervention.* However, analyzing subsequent levels of *frustration* based on *congruence* is appropriate and is the focus of this analysis.

Frustration scale 2 the scale that excludes *frustration* at the time of *intervention*, shows no significant differences for *mean-split congruence.*

Analysis of differences in the *frustration scale 2* between *congruence* and *gender* conditions, controlling for *age, school, mirroring,* and *frustration* at the time of *intervention* shows the main effect significance for *gender*, discussed above in the *intervention* section. There was a significant interaction for *congruence* x *gender* (p= 0.043, F=4.327, Eta(2)= .086. *Boys* and *girls* that received interventions that were not *congruent* had similar levels of *frustration scale 2* mean = 9.4; *boys* that received *congruent interventions* had a mean = 11.5, indicating more *frustration*, while *girls* that received *congruent interventions* had a substantially lower mean = 6.9, indicating less frustration.

Mirroring

Both *frustration* scales 1 and 2 and the highest and lowest *frustration quartiles of scale 1* were tested, using *frustration* response, the first question of the intervention, as a covariate so that the "change in frustration" is what is being evaluated. No significant differences by *mirroring* groups were found.

To summarize these findings there is minimal for *emotional intelligence* (*intervention, congruence,* and *mirroring*) predicting *frustration.*

5.4. H4 Results:

- H4: Metacognitive/meta-affective skill will be exhibited at higher levels when learners interact with emotionally intelligent agents and will positively correlate with perseverance, willingness to continue, and intrinsic motivation.

This section will use the measures of *emotional intelligence,* and *motivation* to analyze their relationship with *Metacognitive/meta-affective* skill. Two *metacognitive/meta-affective* scales have been developed, as discussed in section 5.1.1. Measures for *emotional intelligence* have also been developed and discussed in 5.1.2. Measures of *motivation* have likewise been developed and discussed in 5.3.1. There will be two parts to this section:

1) Metacognitive/meta-affective skill will positively correlate with emotional intelligence.

2) Metacognitive/meta-affective skill will positively correlate with perseverance.
Metacognitive/meta-affective skill will positively correlate with willingness to continue.
Metacognitive/meta-affective skill will positively correlate with intrinsic motivations

5.4.1. Results for metacognitive/meta-affective skill will positively correlate with emotional intelligence

The two scales of *metacognition,* developed in 5.1.2, are analyzed for correlations with the measures of *emotional intelligence* (*mirroring, intervention,* and *congruence*). Controlling for *pretest self theories* as covariates, there was significant correlation between *gender* and the *metacognitive/meta-affective scale 2 (p=.017, r=-.29).* This indicates that overall *girls* experience less flow and more stuck than boys. It is possible that girls aged 11-13 are not as inherently engaged in this task as are the boys, but this remains a topic for future investigation. The *metacognitive/meta-affective scale 2* also tends toward significant correlation with *congruence* (p=0.076, r= -.20) indicating that participants that had more *congruent* interventions also exhibit lower *metacognitive/meta-affective scale 2* (i.e. lower levels of Flow / higher levels of Stuck).

And the *metacognitive/meta-affective scale 2* tends toward significant correlation with *intervention* (p=0.065, r= .21) indicating that participants, of all ages and both genders, regardless of level of frustration, who received *affect support* also self-reported higher values for *metacognitive/meta-affective scale 2*, indicating higher levels of Flow and lower levels of Stuck. In a bivariate (without covariates) correlation of the same measures none of these significances or trends is found.

To summarize these findings there is minimal support across the correlations of the measures of metacognitive/meta-affective skill with the elements of the character's emotional intelligence.

5.4.2 Results for metacognitive/meta-affective skill correlations with perseverance, willingness to continue, and motivation

The measures for assessing *motivation* are described in 5.3.1.

Controlling for *mirroring* and *intervention,* through partial correlation shows that there is no significance in the correlation between the *metacognitive/meta-affective* scales and *perseverance.*

Controlling for *mirroring* and *intervention,* through partial correlation shows that there is a trend toward significance with a negative correlation between the *metacognitive/meta-affective scale 1* and *stick with it* (p=.083, r=-.199). This may sound counterintuitive at first but there is a reasonable explanation discussed in chapter 6.4.

Controlling for *mirroring* and *intervention,* through partial correlation for the remaining measures of *intrinsic motivation* (*strategies, changes in goal-mastery orientation, like to try the activity again,* and binary *intrinsic motivation*) the only measure that shows a trend toward significance in correlation is *strategies* with both *metacognitive/meta-affective scales 1 and 2* (p=.061, r=.19) and (p=.053, r=.23), respectively.

There is significance in the correlation between *metacognitive/meta-affective scale 2* and *like to try the activity again* (p=.044, r=.24).

Since the major *intrinsic motivation* measure is the binary measure assessing if individuals reengaged in the activity this suggests that an ANCOVA should be conducted rather than a correlation. Using *age, school,* and *self theories* as covariates there is no significance for *metacognitive/meta-affective scale 1* (p=.92). In a similar test for differences in *metacognitive/meta-affective scale 2* there is a trend toward significance (p =.062, F=3.6).

Analysis beyond the hypothesis did find that there is a trend in significance for the difference in the levels of *metacognitive/meta-affective scale 1* between those that *quit* and those that did *not quit* (p=.084, F-3.1) when *age, school,* and *self theories* were used as covariates. In a similar test there was no significance in differences in participants reported levels for *metacognitive/meta-affective scale 2* (p=.86).

To summarize these findings there is minimal support across the correlations of *metacognitive/meta-affective skill* with the measures of *perseverance*, willingness to continue (*stick with it*), and *motivation.*

5.5 Gender Results

With an interest in explaining the general lack of support for the primary hypotheses H1, H2, H3, and H4 and to further explore the initial gender findings throughout chapter 5, this section will present the results of an exploratory analysis that was conducted to better understand gender differences and gender effects.

- *H1: the affective learning companion is expected to be more persuasive, and users will form a stronger social bond with the affective learning companion, when sensor-driven non-verbal mirroring informs the affective learning companion's interactions.*

While there were some significant differences, presented in section 5.1, between the *girls* and *boys* in the sampled population, conducting an exploratory analysis of H1 for the separate *gender* groups did not show support for H1 for either group. Here is a summary of the statistically significant gender differences found between *girls* and *boys* in section 5.1. *Boys* were less likely to agree that knowing that the mind is like a muscle will be helpful to them (p = 0.03, F = 4.9). *Girls* indicated they would be able to use the strategies presented in the intervention to a greater extent than boys (p=0.003, F =9.4). *Girls* persevere longer than *boys* (p=0.016, F = 6).

The *intervention* had opposing effects for *boys* and *girls* with respect to the *bye.button* response, with boys responding more positively in the *task support* condition than *boys* in the *affect support* condition and *girls* having the opposite relationship with respect to these two conditions. *Boys* also had more positive impressions of the character that provided *task support* than the character that provided *affective support*, while *girls* had the opposite response.

- *H2.A: A learner's social bond with an affective learning companion will positively correlate with his or her perseverance and self-theories – adoption of internal beliefs that he or she can increase his or her own intelligence and the adoption of mastery orientation.*

- *H2.B: The level of persuasion a learner experiences from the affective learning companion's metacognitive message will positively correlate with the social bond, with perseverance, and will negatively correlate with frustration.*

There were just a few significant differences, presented in section 5.2, between the *girls* and *boys* of the sampled population. Conducting an exploratory analysis of H2 for the separate *gender* groups did not show support for H2 for either group. Analysis in 5.2 shows that *boys* self-report using more *effort* than *girls*.

- *H3: An affective learning companion that exhibits emotional intelligence (active listening, appropriate interventions, and sensor-driven non-verbal mirroring) will increase learner's intrinsic-motivation and reduce frustration.*

Further analysis did not find H3 to be generally supported for either *girls* or *boys;* however several interesting findings were made. H3 was assessed for *girls* and *boys* separately. In the assessment of the relationship between the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*) and *girl* participants' *motivation* the only significant finding was an interaction between *intervention x congruence* ($p = .02$, $F = 6.288$). *Girls* who received *affect support* and had lower levels of *congruence* (i.e. *girls* that received *affect support* and were less frustrated) did not have as much *intrinsic-motivation* as those who had higher levels of *congruence* (i.e. *girls* who received *affect support* and were more frustrated). On the other hand, *girls* who received *task support* and had lower levels of *congruence* (i.e. *girls* that received *task supports were* more frustrated) had more *intrinsic-motivation* than those that had higher levels of congruence (i.e. *girls* who received *task support* and were less frustrated). For girls that were frustrated either intervention increased their *intrinsic motivation* over those that were less frustrated (at the time of intervention) (this will be specifically discussed in 6.5). For boys the only measure of *motivation* that was effected by the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*) was their willingness to *stick with it* which showed a trend toward significance ($p=.065$, $F=3.8$) indicating that boys were more willing to *stick with it* when they received *sensor driven non-verbal interactions* than when they received *pre-recorded non-verbal interactions*.

The relationship between the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*) and participants' *frustration* was also assessed separately for *girls* and *boys,* using covariates of *age, school, pre-test self theories*, and *frustration* response at the time of intervention. In 5.3 *girls* were reported to be less *frustrated* than *boys* at the end of the activity. Also presented in 5.3 is the interaction of *congruence* x *gender* ($p= 0.043$, $F=4.327$.) *Boys* and *girls* that received interventions that were not *congruent* had similar levels of *frustration scale 2* mean = 9.4; *boys* that received *congruent interventions* had a mean = 11.5, indicating more *frustration*, while *girls* that received *congruent interventions* had a substantially lower mean = 6.9, indicating less frustration.

In the follow up analysis, with groups separated by *gender*, there were differences in participants' levels of *frustration* with respect to *congruence* and the *congruence x intervention* interactions. Some of these differences are expected, as discussed in 6.3, due to the "encoding" of *frustration* in the *congruence* measure. In addition to those differences, for *girls* the interaction between *intervention x mirroring* ($p = 0.001$, $F=16.3$) was highly significant. This interaction indicates that *girls* who received *mirroring* and *affect support* had lower *frustration scale 2* (the post activity measure of frustration) than *girls* who received *mirroring* with *task support*. *Girls* without *mirroring* had the opposite relationship with the interventions -- *girls* who received *affect support* had higher levels of *frustration scale 2* than *girls* with out *mirroring* who received *task support*. *Girls* showed no main effect differences in *frustration scale 2* with respect to the type of *intervention* (*affect support* vs. *task support*) that they received (p>.8).

In addition to the "encoded" differences, for *boys,* there were significant (p= .009, F= 8.4) differences for *intervention* with boys showing twice as much post frustration if they received the *affective support* than if they received *task support.* The higher levels of *congruence* were also detrimental for *boys;* they showed almost twice as much frustration for high levels of *congruence* when compared to *boys* with low levels of *congruence.* There was also a trend toward significance (p= .061, F= 4.0) for *mirroring:* boys that received *mirroring* reported a third less frustration than *boys* that did not receive mirroring.

There was a significant interaction between *congruence x mirroring* (p= .047, F= 4.6). *Boys* that had low levels of congruence (with and without *mirroring*), and high levels of congruence with mirroring have approximately equal levels of *frustration scale 2.* These levels of frustration were approximately half the level of frustration experienced by *boys* without *mirroring* that had more *congruent* intervention (i.e. the fourth condition in the *congruence x mirroring* interaction).

- H4: Metacognitive skill will be exhibited at higher levels when learners interact with emotionally intelligent agents and will positively correlate with perseverance, willingness to continue, and intrinsic motivation.

Here too, further analysis did not find H4 to be generally supported for either *girls* or *boys*; however several additional interesting findings were made. For *girls* the *affective support* intervention was positively correlated to the *meta-affective skill* scale (p=.040, r =.37) and the (more *Flow/less -Stuck)* scale (p=.006, r=.52). Neither of these correlations was significant for *boys.*

(more Flow/less Stuck) is very significantly negatively correlated with *congruence* (p=.002, r = -.57). This indicates that if *girls* are not *frustrated* and receive *affect support* (low congruence) then they are also more likely to show high *Flow*/and *less Stuck.* If the *girls* were frustrated and received *task support,* which was also encoded as low *congruence*[2], then *girls* were also likely to show higher *Flow*/and lower *Stuck.* Girls with higher levels of congruence (*girls* that were *frustrated* and received *affective support* or *girls* that were not frustrated and received *task support*) experienced less Flow/more Stuck.

For *girls* there was a significant negative correlation between *Flow/Stuck* and *congruence* (p=.002, r = -.5712). This was not significant for *boys.* For *boys* the *meta-affective skill* scale shows a trend toward significant correlation with *mirroring* (p=.075, r= .29).

In contrast to the 5.4 result that no significant correlation between *meta-affective skill* and *Flow/Stuck* was present when assessed across both genders, the assessment with only *girls* shows significant correlation between *meta-affective skill* and *Flow/Stuck* (p = .010, r=.49). The assessment of only *boys,* for these same measures also shows a significant correlation (p=.021 r=-.40), but for boys this is a negative correlation. This is a clear instance where the grouping of the genders clearly mixes different gender effects, yielding no significance when assessed together.

For *boys* controlling for *age, school, self theories, mirroring* and *intervention,* through partial correlation shows that there is significance (p = .048, r=.34) for *meta-affective skill* correlating with *perseverance,* while there is no significance for *girls.* With these covariates, neither *gender* shows significance for *Flow/Stuck* correlating with *perseverance.*

[2] It is expected that if a participant was not frustrated at the time of intervention, then they may be experiencing Flow. If this were the case then providing "no intervention" might be a more appropriate or more "congruent" response then the *task support* intervention that these individuals received. However, for the purposes of this experiment and its interest in contrasting *affect support* with *task support* this condition was encoded as "congruent". Future work could further investigate these considerations.

The *meta-affective skill* and *Flow/Stuck* scales were investigated with the other measures of *motivation* (*stick with it, strategies, post test goal mastery orientation*, "I would like to try this activity again", and *intrinsic-motivation*) using the same covariates. For *girls* there was no significance for *stick with it*; significance was found for both scales with respect to *strategies*, (p= .027, r=.4766) for *meta-affective skill*, and (p= .009 r=.5691) for *Flow/Stuck*. For *girls* there was also significant correlation to changes in *goal mastery orientation* for the *meta-affective skill* scale (p=.008, r=.57) indicating that *girls* that report higher levels of *meta-affective skill* also report higher levels of *mastery orientation*. There was "almost" a trend toward significance for the correlation of the "I would like to try this activity again" question with *Flow/Stuck* (p=.105, r=.32).

Controlling for the same variables (*age, school, self theories, mirroring and intervention*) there was no significant difference in *meta-affective skill* when the *girls* that showed *intrinsic-motivation* (through reengagement in the task after the post-test surveys) were compared with those that did not reengage. There was a trend toward significant differences between these two groups of *girls* in terms of their *Flow/Stuck* (p = .067, F=3.8). Those that reengaged also had slightly higher levels of *Flow*/slightly lower levels of *Stuck*; both groups were fairly high on this scale, so there may also have been a ceiling effect – i.e. the differences may have been greater. In similar tests *boys* showed no significant differences across these groups and measures.

Boys show no significance for any of the remaining measures. There is "almost" a trend toward significance in the negative correlation of *meta-affect skill* with *intrinsic-motivation*; indicating that *meta-affect skill* for boys may hinder their *reengagement.*

Since this analysis was conducted as exploratory analysis, no further summary or conclusion will be presented in this section. Rather, section 6.5 will discuss these findings and propose several interpretations and a few conclusions.

Chapter 6 Discussion

It must first be stated that the results presented in chapter 5 hold for the sampled population of 11-13 year-olds from three schools in western Massachusetts and that the generalization of these results to a broader population should be verified through additional experimentation before the results can be considered to be generally valid. With this perspective in mind chapter 6 will proceed to discuss and at times speculate on some of the implications that these results might have if they were to hold true for a generalized population of 11 to 13 year-olds using a similar system (the system is discussed in chapter 3) with similar experimental conditions (as those discussed in chapter 3 and 4).

The first five sections of Chapter 6 contain sections that have a parallel structure to chapter 5's sections (i.e. section 6.1 discusses the results of section 5.1 which deals with H1, 6.2 relates to 5.2 and H2, etc.).

6.1 Discussion of H1 Results

The investigation of differences in *persuasion* and *social bond* across *mirroring* conditions, discussed in 5.1, took a multi-pronged approach to assessing several relevant measures.

In the assessment of *persuasion* there were five opportunities, some with multiple measures, that were evaluated, with none of these opportunities showing significant differences between the *mirroring* conditions. However, in the process of evaluating these opportunities several gender differences were found. It was found that *girls* responded slightly more positively than *boys* to the question, "What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort?" This difference, although small, was significant. It was found that *girls* indicate that they would be able to use the strategies presented in the intervention to a significantly greater extent than boys. There are significant differences in *perseverance* across *gender* groups, with the mean value for *girl's* perseverance 2.1 minutes longer than for *boys,* representing 17% greater perseverance. To get some perspective as to weather these effects might have been due to differences in *gender* population prior to the activity, the *pre-test of self theories* were analyzed and no significant differences were found across *gender* for either test.

Two *metacognitive/meta-affective* scales were developed. An examination of the questions that contribute to *metacognitive/meta-affective scale 1* reveals that this scale describes an individual's awareness of affect and their use of strategies to persevere through frustration (see chapter 2 regarding metacognitive and meta-affective skill). Thus a contribution of this thesis is the initial development of an instrument that for this population is a reliable measure of *meta-affective skill.* An examination of the three questions that contribute to *metacognitive/meta-affective scale 2* reveals that this scale describes an individual's reflection on their experience of Flow and Stuck during the activity, in the remainder of this discussion it will be considered a measure of *Flow/Stuck.* Higher levels of *Flow/Stuck* indicate higher levels of Flow/lower levels of Stuck. Future research might productively develop a scale of Stuck in contrast to existing scales of Flow (question 5 in Table 6.1 might be particularly useful in this effort).

A small, but important note is that the future use of these questions should consider revisions of or analytical strategies for dealing with the open ended drop down box which leads to one of the eight questions having a different range than the others, and how to incorporate responses to this question into the computation of a single value for the scale.

1	How helpful was it to think of your mind as a muscle?	Neither scale 1 or scale 2
2	How much of the time did you try to use different strategies to do this activity?	Scale 1
3	If you used different strategies how helpful did you find them?	Scale 1
4	If you used different strategies, how many did you use? *Answer: I used about [drop down box] different strategies.*	Scale 1
5	While you were doing this activity how many times where you aware of your frustration?	Scale 1
6	How much control did you feel you had during this activity?	Scale 2
7	What would you say about your ability to concentrate on this activity?	Scale 2
8	Do you think you have the skills needed to do this activity?	Scale 2

Table 6.1. Metacognitive/meta-affective questions, scale 1 is meta-affective skill, scale 2 is Flow/Stuck

In the assessment of *social bond* there were also several opportunities to develop different measures, with three new ones created and used in this thesis. Two scales were developed, *character positive* and *character negative*, and the quasi-behavioral measure of *bye.button* response was developed. While these measures showed strong correlation when the reliability of these three measures was jointly assessed it was shown to be insufficient to combine them as a scale. Participants that had negative *bye.button* responses also had significant differences in their *character positive* and *character negative* responses when compared with other participants. These differences and the strong correlations to the self-report measure indicate that the quasi-behavior *bye.button* measure is a useful and implicit (rather than explicit) measure of social bond. It should be noted that there was a significant negative correlation in social bond with *age*. An additional assessment of the reliability of a scale that would combine the three measures of social bond was conducted for each age 11, 12, and 13 years old and insufficient reliability was found. It had been anticipated that the younger participants might show less discrepancy between explicit and implicit measures, while older ones might show more inconsistency among these; this could be a topic for future investigation. There was also a significant difference across *gender*; *boys* had quasi-behavioral *bye.button* responses that were slightly more negative than *girls*' responses. While this *gender* difference was not found for the *character positive* nor *character negative* measures there was an interesting interaction, *intervention x gender*, that will be discussed in the coming paragraphs.

Once the *social bond* measures were established, the evaluation of social bond proceeded with respect to *mirroring* and *intervention.* There was no main effect found in the differences in the *social bond* across the *mirroring* or *intervention* conditions. In this study the participants were much younger (11-13 years old vs. college students) than the participants in the previous related studies by Bailenson (Bailenson, Beall. et al. 2005), which found liking due to head movement mirroring, and Bickmore (Bickmore and Picard 2004) which found that daily interactions with an agent led to increased social bond. Bickmore's study included no mirroring but did include occasional empathy as well as other relational non-verbal moves (frame shifts). Other studies of human gross-movement mirroring such as leg-crossing or hair-flipping were conducted with adults, and the context of these studies were more typical social interactions such as interviews, waiting room conversations, or dating, rather than learning a frustrating task (LaFrance 1982; Bull 1983). The frustrating learning task presented to users in this experiment may also have diminished some of the mirroring effects, as it might have drawn user's attention away from the character. For the population in this study it may not be enough to just mirror/give empathetic interventions to increase their social bond with the character, especially within the context of a difficult task.

While there was no main effect found in the differences in the *social bond* across the *mirroring* or *intervention* conditions, here too there were interesting *gender* effects. The interaction of *intervention x gender* was significant for *bye.response* and *character positive* but not for *character negative.* For both the *bye.button* response and *character positive* measures, *boys* had more positive impressions of the character that provided *task support* than the character that provided *affective support*. *Girls* had the opposite pattern in their response. Early adolescence is a time of rapid emotional development for girls, and boys tend to lag behind girls, so this finding may indicate one of the idiosyncrasies of working with this age group. Earlier work that examined a computer's empathetic responses in frustrating and stressful contexts and found the responses beneficial was conducted only with adults (Klein, Moon et al. 2002; Liu and Picard 2005), where gender differences in effectiveness of the empathetic frustration and stress responses were not found. Bickmore and Picard's experiment with a software agent (Bickmore and Picard 2004) also used only adults and had no gender effects for social bond across its relational vs. non-relational conditions.

To assess the relationship between *character's emotional intelligence* and *social bond* a process to quantify the *character's emotional intelligence* was developed based upon the *mirroring, intervention* and *congruence* measures. As stated above, *social bond* showed no main effect significance with respect to *mirroring* or intervention. A measure of high and low *congruence* was developed based on *frustration response*. *Social bond* also showed no significance with respect to this measure of *congruence.*

The Inference Engine classification had insufficient variance for it to be used in the development of *congruence.* Future work should include retraining the Inference Engine for this data set. There are several reasons that may have contributed to the need to retrain, including changes in the protocol with respect to the third pilot study, on which the Inference Engine was trained, and changes in the population.

Summary

There was no support for H1: the affective learning companion was not found to be more persuasive nor did it create stronger social bonds based on *mirroring.* The general finding was that no significant differences were found for the type of mirroring implemented by this character. Recall that the control condition also had movements that "mirrored" for another person and were presumed to be uncorrelated with the movements of all the other subjects who were in this control.

The process using the analytical strategy for multiple measures was shown to have mixed success. The successes were that several scales with statistical validity, *character positive, character negative, metacognitive/meta-affective scale 1 and 2* were produced. While the scales have shown to be reliable in this experiment, future experiments, inter-experiment comparisons, and correlations to generally accepted scales would need to be conducted before they can be considered as generally validated. The strong correlations between the quasi-behavioral *bye.button* response and *social bond* were established and are promising for future investigations of *social bond.* Further consideration of how the quasi-behavioral measure can be made more robust, e.g. how to ensure that all participants provide *bye.button* responses would be productive as would investigations to further assess if its negative correlation with *age* is a specific to this character and the 11 to 13 year-old *age* range.

Where scales could not be developed the analysis became cumbersome. For example, no scale could be developed for *persuasion* because the measures that are used to assess persuasion do not, individually, measure a common construct (i.e. they are not internally consistent). Instead, *persuasion* was assessed as a construct that has multiple opportunities with multiple sub-measures, so each sub-measure was assessed individually.

While H1 was not supported, several interesting findings for *gender* effects emerged from the analysis. The main results for *gender* was that *girls*, more than boys, responded that the slide show and its message would help them, that they would be able to use the strategies presented in the intervention, *persevered*, and responded positively in the *bye.button* response. While *boys* had more positive impressions of the character that provided *task support*, *girls* had more positive impression of the character that provided *affect support.* Gender findings will be discussed further in 6.5.

6.2 Discussion of H 2 Results

The investigation of correlations of *social bond* with *perseverance* and *self theories* and the investigation of correlations of *persuasion* with *social bond, perseverance,* and *frustration* are discussed in this section.

H2A: social bond would positively correlate with perseverance or with self-theories

There was no support for the first component of H2, the hypothesis H2.A that *social bond* would positively correlate with *perseverance* or with *self-theories.* The only significant or trend toward significant finding from the investigation of the three measures of social bond (*character positive, character negative,* and *bye.button* response) was that there was a trend toward significance that indicates that participants that responded with more positive *bye.button* responses also had positive *changes in self theories of intelligence.*

Summary

To summarize, this part of H2 is not supported by the results. It is possible that any social bond with the character may not transfer to motivation and perseverance in the task – unless the character were to emphasize the importance of the task to that character (like experimenters do to get subjects to stay in long-term studies "please do your best, this is really really important for my thesis…" while building rapport with the subject.)

There were four individuals (discussed in chapter 4) who were excluded from the analysis because they finished or mistakenly thought they had finished (i.e. moved the tower to the middle tower or interpreted the large white disk as a plate rather than a disk, and in these cases were not interested in continuing after the experimenter corrected their impressions). Two were *boys* and two were *girls*; one received *affect support* without *mirroring,* two received *affect support* with *mirroring*, and one received *task support* with mirroring. Their *character positive* impressions had a similar spread to the included participants' responses, while their *character negative* impressions were slightly less negative than the included participants' responses. Two of these participants said bye to the character with the neutral *bye.button* response and one used the positive *bye.button* (a distribution slightly less positive than that of the included participants); one did not have a *bye.button* respond (this response was an optional interaction). Related *gender* findings will be discussed in section 6.5.

H2B part 1: persuasion will positively correlate with social bond

Proceeding to H2B, the first part of the three parts of H2B states that *persuasion* will positively correlate with *social bond.* Social bond was assessed correlating the three measures to each of the five opportunities to assess *persuasion.* Only the significant correlations and trends will be mentioned here; all unmentioned correlations were not significant.

The first opportunity to measure *persuasion,* assessing responses to questions asked during the introduction, showed a trend toward significant correlation with *bye.button* response. The second opportunity, assessing responses to the questions asked during the *intervention* (*strategies* and *stick with it* using *frustration* response and *effort* response as covariates), showed no significant correlations. The third opportunity, assessing after the task if the *character was easy to understand,* correlated significantly with *character positive, character negative,* and showed a trend toward significant correlation with *bye.button* response. The fourth and fifth opportunities to measure *persuasion* showed a trend toward significance in the correlation of *changes in self theories of intelligence* (measured as the difference between pre test and post test scores) with *bye.button* response.

Summary

To summarize these findings there is minimal to mixed support across the correlations of the measures of *social bond* with the opportunities to measure *persuasion.* It is interesting to note that there were strong correlations between participant's responses to "It was easy for me to understand what Casey was saying." and the two survey based measures of *social bond, character positive* and *character negative*, while there was no significance in the correlation of the responses to this question and the quasi-behavioral *social bond* measure, i.e. the *bye.button* responses. This means that when asked a series of questions about the character, the participants perception of their ability to understand the character was correlated to their impressions of the character; the more they thought it was easy for them to understand what the character was saying, the more positive their impressions of the character. However their perceptions of their ability to understand what the character was saying did not influence the way they said bye to the character, i.e. their *bye.button* response. This discrepancy should be considered and investigated as the use of this quasi-behavioral *social bond* measure is developed and understood and used in future experiments. One aspect to consider in this process is the participant's interpretation of the question about their understanding. This question could be consider a question related to understanding the vocal qualities, volume, speech pattern, etc. of the character's voice, or it could be interpreted more generally to include the meaning and an understanding of the character's message. To disambiguate these multiple interpretations further studies could be more explicit in separating the distinctions of the various types of understanding that are relevant to participant's experience of social engagement.

H2B part 2: persuasion will positively correlate with perseverance

The rational for this hypothesis is that the introduction dialogue (see Appendix C for a transcript) presents ways to improve *meta-affective skill*, based on Dweck's research and her experimentally validated treatments that improve self theories of intelligence and goal mastery orientation. The introduction was explicitly designed to encourage learners to persevere. The perseverance measure was the measure of the amount of time individuals continued with the activity. The first opportunity to measure persuasion, assessing responses to questions asked during the introduction, showed significant correlation with responses for "What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort?" and "On a scale from 1 to 7, how hard do you think this activity will be for you?" and a trend toward significant correlations for responses to "On a scale from 1 to 7, how hard do you think it'll be for you to do new things if you get frustrated?"

The second opportunity to measure *persuasion* was to assess responses to two questions asked during the intervention. This showed significance for *perseverance* correlated with participant's belief about their ability to use the *strategies* presented in the *intervention* while no significance was found between *perseverance* and their willingness to *stick with it.* The third opportunity to measure *persuasion* and its relationship to *perseverance* was to assess if the *character was easy to understand,* this correlation was not significant. The fourth opportunity to assess persuasion (as outlined in 5.1) was to assess *perseverance*. Assessing the *perseverance* as a measure of *persuasion* with respect to *perseverance* is implicitly self-correlated. The fifth opportunity to measure *persuasion* with respect to *perseverance* was to asses its correlation to the *metacognitive/meta-affective scales* and *changes in self theories.* This showed no correlation with either of the *metacognitive/meta-affective scales* nor with *changes in self theories;* only a trend toward significance correlation between *perseverance* and the *metacognitive/meta-affective scale 1.*

To summarize these findings there is minimal to mixed support of the correlations of *perseverance* with the five opportunities used to measure *persuasion.*

In assessing the second opportunity, it is interesting that *strategies* showed significance in its correlation with *perseverance* while desire to continue does not – that a participant's belief in their meta-affective strategies, or task strategies in the task support intervention condition, trump the participant's indication of his or her willingness to stick *with it* with respect to the correlation with their *perseverance*.

Summary

To summarize these findings there is minimal to mixed support of the correlations of *perseverance* with the opportunities to measure *persuasion*. As possible explanations for this outcome, it is possible that participants' abilities to *persevere* are more stable (i.e. less malleable, or less subject to influence from the treatments of this experiment) than their response to the various *persuasion* measures. Either of these measures may have had a ceiling effect (i.e. may not have been malleable to a sufficient extent or may have already reached a maximum either for the population or with respect to the measures used in this experiment). It is also possible that these measures of *persuasion* simply do not have enough impact on the behavioral metric *perseverance;* behaviors are frequently harder to alter than beliefs. Another explanation is that there may be too many other influences on *perseverance*, such as a participant's affinity or disposition for this type of activity, which make it difficult to see the effects of *persuasion* on *perseverance*.

H2B part 3: persuasion will negatively correlate with frustration

An initial check was conducted to ensure that there were no differences across *intervention* type with respect to the level of *frustration*.response at the time of intervention. This test showed a trend ((p=.083) toward significant differences (with and without *school* and *age* as covariates) with a lower mean self reported *frustration*.response for those in the affect condition. Age correlated with overall frustration (*frustration scale 1*) with younger participants experiencing less frustration (p=.039, r=-.240). Testing was conducted over a two week period at the first school and one week at each of the other schools. This duration likely mitigated effects that might have been due to an experimenter's mood. Experimenter's mood is likely to vary over these longer durations, and thus average out the effects. There should not be differences in frustration at the time of intervention since participants in both groups *intervention* conditions had identical experiences up to that point, while there is only a trend toward significance (with participants in the *affect support* condition responding with slightly lower frustration than those in the *task support* condition), it may have contributed to unaccounted variance that might have diminished the ability of this experiment to support the primary hypotheses. *Boys* self-report using more *effort* than *girls*; other gender effects are discussed further in 6.5.

The rational for this investigation was that learners would be better equipped to manage frustration the more they believed and applied the *meta affective skills* message (that the mind is like a muscle and that even though it may be frustrating if you stick with it and try different strategies you can grow your intelligence) presented in the introduction (see Appendix C for a transcript of the message). Recall that two *frustration* scales were developed; one overall scale that combined measures from the time of intervention with post activity measures and the second measuring *frustration* after the *intervention*. The change in levels of *frustration* was also assessed by assessing the second scale while controlling for the level of *frustration* at the time of intervention.

The first opportunity to measure persuasion is to assess responses to questions asked during the introduction (see Table 6.2.1).

1. What do you think about intelligence, do you think it can change?
2. On a scale from 1 to 7 , how hard do you think it'll be for you to do new things if you get frustrated?
3. What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort?
4. If things get hard for you in this activity, do you think you will be able to try different ways of doing the activity?
5. On a scale from 1 to 7 , how hard do you think this activity will be for you?

Table 6.2.1. Questions that are assessed in the first opportunity to measure persuasion

Responses to question 3 showed a trend toward significant negative correlation with frustration scale 1. Responses to question 5 showed significant correlation with both frustration scale 1 and frustration quartiles.

The second opportunity to measure the relationship between *persuasion* and *frustration* investigated questions presented at the time of intervention (see Table 5.1.2.) that relate to participants beliefs about their ability to use the *strategies* presented in the intervention (either *task* or *affective* strategies) and separately their willingness to *stick with it* (i.e. the activity). While no significance was found for their willingness to *stick with it,* there was strong negative correlation between levels of *frustration*; participants that were less frustrated were more likely to indicate that they felt they could use the strategies. This finding can be interpreted in terms of Kort et al.'s affective model of the interplay of emotions and learning (Kort, Reilly et al. 2001) which suggests that there is a spiral trajectory of affect that learners progress upon that leads from curiosity to puzzlement, frustration, and finally to a fresh approach. At the time of frustration there might be an "anti-spiral" or a spiral of Stuck that can lead learners away from Kort's productive path.

The third opportunity to assess if *frustration* has a negative correlation with persuasion is to assess if the *character was easy to understand,* i.e. if the character is easier to understand for those experiencing lower levels of frustration than those experiencing higher levels. This relationship was not significant. The fourth opportunity to assess *frustration* is to correlate it with the *perseverance* measure; this showed very strong significance. The fifth opportunity to measure *frustration* showed no correlation with either of the *metacognitive/meta-affective* scales or with *changes in self theories.*

Summary

To summarize these findings there is mixed support of the negative correlations of *frustration* with the opportunities to measure *persuasion*. Several of the measures showed trends for significance and a few of them showed significance, while many more showed no significance. There may have been a ceiling effect, as only 16 participants pressed one of the quit buttons, and there may not have been sufficient distribution of frustration for these correlations to be measured.

6.3 Discussion of H3 Results

This section discusses the findings related to the character's *emotional intelligence (intervention, congruence,* and *mirroring*) and its prediction of the learner's *motivation* and *frustration.* Providing participants with the opportunity to reengage with the activity and coding their response as a measure of *intrinsic motivation* was shown to capture sufficient variance across the experimental population for it to be used in the analysis. This strategy is therefore encouraged in future investigations relating to learning and motivation.

Does emotional intelligence (intervention, congruence, and mirroring) predict motivation?

Emotional intelligence is a complex concept that is not yet fully understood, even in human to human interactions. One of the motivations for the study of emotional intelligence in the context of an affective agent research system is to have greater levels of control over some of the elements of emotional intelligence that the character displays and to study these with respect to motivation, learning, and perseverance. This research focused on just a few elements of emotional intelligence, specifically the type of intervention, *affect support* vs. *task support*; *congruence* or the appropriateness of the type of *intervention* with respect to the participant's level of *frustration* (with *affective support* hypothesized as being more appropriate for those with high levels of *frustration* and *task support* more appropriate for those with low levels of *frustration*; and *mirroring* (*sensor driven vs. pre-recorded non verbal interaction*). There are certainly many other elements of emotional intelligence that were not implemented. For example, while the character did attempt to look as if it cared, when a participant said they were really frustrated, it did not adjust its *intervention* timing to show respect for participants attention level by not interrupting when they were in the midst of concentrating on the task.

Summarizing the analysis set out in 5.3, there is almost no support for the link between the measured aspects of *emotional intelligence* and *motivation*. The only measure that shows a trend toward significance is the differences in *mirroring* with the *stick with it* measure. If this trend were to hold and to become significant in future experiments this would be quite interesting.

When *intervention* was assessed with *gender* most of the measures of *motivation* were not significant, however there was a main effect significant difference in *strategies* for *gender* indicating that *girls* report that they believe they will be able to use the *strategies* provided in the intervention, more than *boys* believe this.

Throughout the pilot studies were a range of participant's anecdotal responses to the character, with some saying: "Casey is dumb, it didn't do anything" and others "I liked her, she was really helpful". While this experimental protocol did not include an opportunity for a qualitative response from participants, further investigation of user's qualitative experience relative to the quantitative measures might help in determining the participant's perception of the emotional intelligence of the character and the effectiveness and impact of the character's emotional intelligence across participants. In spite of the many challenges that it entails, gathering open ended qualitative measures in future experiments would allow for the analysis to go beyond the modified Working Alliance Inventory measures used in this study, to capture and further elucidate some of the intricacies and complexities of emotional intelligence.

To critique why these elements of the character's *emotional intelligence* did not have the hypothesized effects they will now each be discussed. This intervention may not have been an appropriate intervention for both *boys* and *girls*. There may be sufficient gender differences in this age group to warrant different *interventions* based on *gender*. In terms of *congruence* the *gender* interactions may have again played a role in diminishing the ability of this experiment to prove its hypotheses. Additionally the adaptive response of the *affect support* may have introduced additional variance into the *congruence* measure. In the *affect support* condition participants with low frustration had interactions which addressed their current reduced levels of *frustration* and *effort*. While these interactions still contained no *task* based information they may have been more appropriate for these individuals than their "low *congruence*" encoding would indicate. Additional *gender* findings are discussed in section 6.5.

In terms of mirroring there are several factors that could have contributed. The mapping, intensity, and duration of the interactions may not be tailored sufficiently. For example, interactions in one channel may at times work against behaviors in another channel to appear, incongruous or inappropriate and may be off-putting or detrimental to the development of *social bond* or *persuasion*. This is one of the challenges of conducting multimodal sensor-to-behavior mapping, Bailenson's work, for example, had only a single sensor, a head tracker, to map a single well understood behavior. In contrast there were four sensors, and multiple behavior mappings that were conducted in this experiment. Mapping mouse pressure to character agitation and skin conductance to character skin tone for example may not be effective mirroring strategies. These mappings were chosen in this experiment, but they are not rooted in findings from interpersonal communication literature; reflecting the participant's unease in this way may have caused the unease to be amplified. In future studies this could be further examined.

Does the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*) *help learners better manage frustration*

There were no main effect significances between the *affect support* and *task support intervention,* the high and low levels of *congruence, or mirroring* (*sensor driven* vs. *pre-recorded non-verbal* *interactions*) with any of the measures of *motivation* (*intrinsic motivation, change in goal-mastery-orientation, strategy, stick with it,* and *like to try again*). There was however a main effect significance for *gender,* with *girls* reporting significantly less frustration at the end of the task than boys (*frustration* response at the time of intervention did not differ across *gender*). A significant interaction of *congruence* x *gender* was also found. *Congruence* was encoded based on participant's *frustration* response at the time of *intervention* and its relationship with the type of *intervention* they received. *Affect support* is coded as having high *congruence* with higher levels of *frustration* response while *Task support* is coded as having high *congruence* with lower levels of *frustration* response. (see chapter 4 and 5.1 for a description of the *congruence* measure). Further discussion of the gender findings are presented in section 6.5.

Summary

The overall conclusion is that H3, as formulated in this experiment, is not supported. This may have been due to several limitations in the *intervention, congruence,* and *mirroring* strategies used in this experiment. For example, this experiment assessed only two forms of intervention and it combined meta-affective skills messages with the empathetic intervention provided in the *affect support* condition, while it generally excluded elements of empathy from the *task support* intervention. It is likely the case that, even at times of low frustration, some elements of empathy are important to incorporate while providing *task support.* If a few of these elements were included in the *task support* intervention it may have lead to better outcomes for participants in this conditions, especially for those with lower frustration (i.e. higher levels of congruence, see chapter 3 for a discussion of congruence). Future work could further investigate the importance of including elements of empathy while providing task based support. There may also be gender differences and developmental differences with respect to participant's emotional intelligence that are particular to the 11-13 year old age group that were not adequately understood or incorporated into the design of the experiment (see section 5.5 and 6.5 for further discussion of gender results from this experiment). These differences may have contributed to unanticipated variance that could interfere with support for the primary hypotheses.

6.4 Discussion of H 4 Results

The investigation of the correlations of *metacognitive/meta-affective skill* with respect to *emotional intelligence, perseverance,* willingness to *stick with it,* and *intrinsic motivation* are discussed in this section.

Metacognitive/meta-affective skill will positively correlate with emotional intelligence

This investigation employed the two *metacognitive/meta-affective* scales (i.e. the *meta-affective skill* and *Flow/Stuck*), and three measures of the character's *emotional intelligence* (*intervention, congruence,* and *mirroring*). There were trends toward significance in a negative correlation between the user's reflection of their *Flow/Stuck* experience during the activity and congruence (one of the three elements of the character's *emotional intelligence. Congruence* is a measure that is coded by the level of *frustration* response at the time of intervention, and the type of *intervention* with high levels of *congruence* being encoded for individuals that experience higher levels of *frustration* and receive *affect support,* and likewise high levels of *congruence* encoded for those who experience lower levels of *frustration* and receive *task support.* The relationship between the *Flow/Stuck* scale and the *congruence* measure indicates that the higher the *congruence* of the intervention the less *Flow/* and more *Stuck* participants experience. This trend opposes the predicted relationship, as stated in H4 hypothesis. Further investigation of this negative correlation by gender shows that is significant for girls and is not significant for boys. The gender differences will be discussed in 6.5.

The *Flow/Stuck* scale also shows a trend toward significance with *intervention* indicating that the individuals that received *affective support* also reported significantly higher levels of *Flow/* lower levels of *Stuck* as they reflected upon their experience. Here too, there are related gender differences that will be discussed in 6.5. There was a significant correlation between *gender* and *Flow/Stuck* indicating that *girls* reported less *Flow,* or more *Stuck,* in their reflection on the experience, than the *boys* did. While there are a few trends toward significance there is not much evidence to support the correlation of the character's *emotional intelligence* as implemented in this experiment through *intervention, congruence,* and *mirroring* with the *meta-affective skill* or *Flow/Stuck* scales.

Metacognitive/meta-affective skill's relationship with perseverance, willingness to continue, and motivation

No significance was found for the correlation between either the *meta-affective skill* or the *Flow/Stuck* scales developed in this experiment and participants' *perseverance* (measured as the amount of time a participant engaged in the task). A trend toward significance with a negative correlation was found between the *meta-affective scale* and *stick with it.* This indicates that individuals that said they were willing to *stick with it* also showed lower meta-affect. While this may sound counterintuitive at first there is a reasonable explanation. These two measure address somewhat different topics – the *meta-affect scale* addresses the capability to use affect to pursue a task. *Stick with it* is a form of expectancy and is asked at the time of *intervention,* before the agent leaves. Perhaps the negative correlation arises because participants set high expectations for themselves (saying they will *stick with it*) then, as they engage in a difficult activity, this may lead them to report lower levels of *meta-affective skill* (i.e. lower ability to use their awareness of their frustration to continue to try the activity in different ways). In answering positively to the *stick with it* question, they may have set themselves up for a fall.

The rest of the measures of *motivation (strategies, changes in goal-mastery orientation, like to try the activity again,* and binary *intrinsic motivation*) do not correlate with the measures of *meta-affective skill* and *Flow/Stuck* except in two cases where there are trends toward significant correlation. *Strategies* showed a trend toward significant correlation with both *meta-affective skill* and *Flow/Stuck.* This is interesting since as seen above willingness to *stick with it* did not correlate significantly with *Flow/Stuck* and showed a negative trend in its correlation with *meta-affective skill.* In contrast to participants, discussed above, that might set themselves up for a fall, participants that believe that they could use the *strategies* effectively also indicated higher levels of self-reported *meta-affect,* as well as higher levels of self reported *Flow* upon reflection of their experience in the activity. This relationship supports the hypothesis that the *meta-affective skills* are important and apparently more productive than expectations (willingness to *stick with it*). In section 5.2 *strategies* was also shown to have a significant positive correlation with *perseverance* while willingness to *stick with it* does not.

While the scales of self reported *meta-affective skill* and *Flow/Stuck* were independent in the general population when assessed for boys and girls separately it was found that these measures were significantly correlated for each group, but in opposite directions from each other. For girls *meta-affective skill* and the more *Flow/* less *Stuck* scale were positively correlated, while for boys they were negative correlated. Gender differences in the sampled population may have contributed unaccounted for variance that lead to the lack of support for the primary hypotheses. Sections 5.5 and 6.5 further investigate and discuss the impact of gender differences and some interesting findings.

The second measure that supports the correlation of *metacognitive/meta-affective* skill with *intrinsic motivation* was *like to try the activity again* which correlated with *Flow/Stuck.* This finding makes intuitive sense since for participants that reported greater levels of *Flow* it would seem natural that they would also report that they would *like to try the activity again.* While there was no significant difference in the levels of *meta-affective skill* reported by those that had high and low *intrinsic motivation,* there was a significant difference in their reported levels of *Flow/Stuck.* Here again, this makes intuitive sense since, as it would seem natural for individual who had more optimal experiences to want to reengage in the activity. As presented in section 5.5 and discussed in 6.5 girls that received affective support had much higher levels of *Flow/*and lower levels of *Stuck.* While there was not direct support in this study for *affective support* leading to higher levels of *intrinsic motivation,* it is expected that further research on this connection would be productive.

Summary

While there are a few trends toward significance there is little to mixed support for the correlation of *emotional intelligence*, *perseverance*, willingness to *stick with it*, and *intrinsic motivation* with *metacognitive/meta-affective skill*. There are positive correlations between individuals' beliefs that they are equipped with *strategies* they can use and their levels of *meta-affective skill* and their actual *perseverance* in the activity while those that simply have an expectation that they are willing to *stick with it* do not have significant correlations with *perseverance* or *meta-affective skill*. With respect to boosting time spent persevering in a learning activity it appears more productive to have meta-affective strategies than to have higher levels of willingness to stick with it.

6.5 Discussion of Gender Results

In the process of the analysis of the primary hypothesis several gender differences were discovered. Once the primary hypotheses were found not to hold true across the population, it was anticipated that these and other gender differences may have played an important and unaccounted for role in the experience and responses of the participants, that lead to the lack of support for the primary hypotheses. Further investigation considered that while the primary hypotheses did not hold true across the sampled population they might hold true for *girls* and *boys*, when these groups were assessed separately; this was not the case. The exploratory analysis of H1, H2, H3, and H4 for the separate *gender* groups did not support the primary hypothesis.

In the processes of this further investigation, numerous and substantial *gender* differences were found that may have contributed to the lack of support for the primary hypotheses across the sampled population. First, it should be understood that when the population is separated by *gender* it halves the sample size, reducing the ability of this experiment to find support for the hypotheses. In spite of the general lack of support for the primary hypotheses in any of the analyzed populations groups (across the population or when separated by *gender*), the further investigation yielded several interesting results that support strong and potentially important recommendations for further study. This section will summarize the results of the gender specific analysis presented in section 5.5 and argue for the importance of a deeper understanding of the impact of *mirroring* and of *affect* and *task support,* as these relate to the *frustration, meta-affective skill* and *Flow/Stuck* of the 11-13 year-old sampled population. In particular this section argues for the need for better understanding of the *gender* differences in the impact of the elements of a learning companion's *emotional intelligence* and for the importance of the appropriate coordination of these elements with each other, for both *girls* and *boys.*

As presented in 5.5 there were a few differences in the pattern of the social bond that *girls* and *boys* develop with the character, with respect to the type of *intervention* the participants received. *Boys* responded more positively to the character and had more positive impressions of the character that provided *task support* than the character that provided affect support; *girls* had the opposite pattern. Differences in the social and emotional skill developments of girls and boys at these ages (11-13 year olds), with girls typically maturing earlier than boys, may have contributed to these differences. (The effects for both male and female adults who received affect support from machines in non-learning environments have previously been positive.) *Boys* also self-report using more *effort* than *girls*. This finding and the *frustration* finding for *girls* discussed later in this paragraph, may have influenced different levels of interest in this activity, for *girls* and *boys*. There were very few differences found in the *motivation* measures with respect to the different elements of the character's *emotional intelligence* for either *girls* or *boys*. It was found that the *girls* that were more frustrated at the time of intervention also showed higher levels of *intrinsic motivation*, regardless of *intervention.* A possible explanation for this may be related to how much a participant cares about the activity. *Girls* that care more about doing this activity may also find it more frustrating. Independent of the frustration and independent of the type of *intervention* they receive, the caring may also lead to their increased *intrinsic-motivation.* In contrast to the *girls, boys* showed a strong difference in their levels of frustration due to the type of *intervention*, with much lower levels of frustration occurring in the *task support* conditions. This is probably related to the *social bond* differences discussed above, in which boys responded better to the character in the *task support* intervention. Likewise it is likely related to the finding that *boys* and *girls* that received interventions that were not *congruent* had similar levels of post activity *frustration*; while *boys* that received *congruent interventions* had higher levels of post activity frustration and *girls* that received *congruent interventions* had substantially less frustration.

One of the biggest gender differences was found in the relationship between *meta-affective skill* and *Flow/Stuck*. In contrast to the 5.4 result that no significant correlation between *meta-affective skill* and *Flow/Stuck* was present when assessed across both genders, the assessment with only *girls* shows a strong correlation between *meta-affective skill* and more *Flow*/less *Stuck*, while for boys these measures show a strong correlation in the opposite direction. This is a clear instance where the grouping of the genders clearly mixes different gender effects, yielding no significance when assessed together. This and other *gender* differences must be further investigated and better understood in order to develop affective learning companions for this age group.

While *girls* showed no main effect difference in the level of frustration based on the type of intervention, a further analysis indicated that this masked a more complex relationship that showed highly significant differences due to the interaction of the type of *intervention* and the presence of *mirroring*. These differences can be explained in terms of the coordination of the different elements of the character's *emotional intelligence*. *Girls* that experienced an *affective support* intervention in conjunction with *mirroring* (condition 1) had lower levels of frustration than *girls* who received either *affective support* without *mirroring* or *girls* who received *task support* with *mirroring*. Condition 1 is a condition in which the *mirroring* and *intervention* are coordinated so that the character displays higher levels of *emotional intelligence* (as defined in this experiment as the presence of *intervention, congruence,* and *mirroring*) than in the other two conditions. One might argue that *girls* that received *task support* without *mirroring* were also in a coordinated condition that presents a character with higher levels *emotional intelligence;* they could also argue that in this condition *girls* experienced similar low levels of frustration when compared to the *girls* in condition 1. Extending this argument one might then argue that the existing capabilities of Intelligent Tutoring Systems, to provide *task support* without *mirroring* have similar benefits to girls, and the effort to develop *affect support* and *mirroring* are unwarranted. However the importance of *affect support* for girls is bolstered by the exploratory analysis of H4 showing that *girls* that receive *affective support* have higher levels of *meta-affective skill* and more *Flow*/less *Stuck* (these relationships were not found for *boys*). *Meta-affective skill* correlated significantly with beneficial changes in *goal-master orientation* and there was a trend toward significance in the positive relationship between *Flow/Stuck* and *intrinsic-motivation*. The findings from H3 and H4, taken together, support an argument not only for the further development of *affective support* and its benefits for *girls,* but also for the appropriate coordination of the elements of the character's *emotional intelligence*. These findings indicate that there are important opportunities to increase *girls' meta-affective skills*, increase their experience of *Flow* and decrease their experience of *Stuck,* increase their *mastery orientation,* and increase their *intrinsic-motivation.*

Data from the *boys* also supports the argument for coordinating the elements of the character's *emotional intelligence*. The significant interaction between *congruence x mirroring* indicated that the *boys* that experience *congruent* intervention without *mirroring* also experienced twice as much post activity *frustration* as boys in the other three *mirroring x congruence* conditions. This particular form of discordant *emotional intelligence* displayed by the character (i.e. *congruent* intervention without *mirroring*) seems to have had a negative impact on these *boys*. The *congruence* construct and some of the issues relevant to its development and implementation in this experiment will be discussed later in this section.

Other findings for *boys* include a main effect trend toward significance for *mirroring* leading to less *frustration* and a trend toward significant in the correlation of *meta-affective skill* with *mirroring*. There was "almost" a trend toward significance in the negative correlation of meta-affect skill with intrinsic-motivation; indicating that meta-affect skill for boys "may" hinder their reengagement. For *boys* the *meta-affective skill* scale correlated significantly with *perseverance* indicating that even though *meta-affective skill* "might" hinder *intrinsic-motivation* it has a positive relationship with *perseverance*.

Other findings for *girls* include a significant negative correlation between *Flow/Stuck* and *congruence,* indicating that if *girls* are not *frustrated* and receive *affect support* (low congruence) then they are also more likely to show high *Flow/*and *less Stuck.* If the *girls* were frustrated and received *task support,* which was also encoded as low *congruence,* then *girls* were also likely to show higher *Flow/*and lower *Stuck.* This may be a circumstance in which the *task support* benefited *girls. Girls* with higher levels of congruence (*girls* that were *frustrated* and received *affective support* or *girls* that were not *frustrated* and received *task support*) experienced less Flow/more Stuck. The *girls* that were *frustrated* and received *affective* may have lower levels of *Flow/*higher of *Stuck* because they were frustrated. The *girls* that were not *frustrated* and received *task support* may have had lower levels of *Flow/*higher of *Stuck* because the *task support* intervention may have interrupted their *Flow.*

Throughout the experiment the investigation of the effects of *congruence,* such as those mentioned above for *girls* and *boys,* may have been compounded by two aspects of the implementation and development of the *congruence* construct. First, since the *affect support* is adaptive (see chapter 3), it was designed to be more "congruent" to participants' levels of *frustration* and *effort* than the *task support* intervention. Second, as noted in the footnote in section 5.5 providing "no intervention" might be a more appropriate or more "congruent" response when individuals are not *frustrated* than the *task support* intervention which the participants of this experiment received. For the purposes of this experiment and its interest in contrasting *affect support* with *task support* this condition was encoded as "congruent". Future work should further investigate improvements to the *congruence* construct and to the development of additional forms of intervention. In particular it is recommended that future investigations combine empathy, meta-affective skill, and task support toward an understanding of an integrated approach to supporting learners.

Through a better understanding of the effects of the diverse elements of the character's *emotional intelligence (intervention, congruence,* and *mirroring*) and the differences in their impact on *girls* and *boys,* as well as through improvements in the types of interventions that are developed these results indicate that there are substantial and important opportunities to improve learners' *meta-affective skills*, increase their experience of *Flow* and decrease their experience of *Stuck,* increase their *mastery orientation,* and increase their *intrinsic-motivation* through the advancement of affective learning companions that are capable of *mirroring* and *affective support.*

7 Contributions

This thesis has made contributions to Educational Psychology and to Intelligent Tutoring Systems in three areas: theory, system development, and experimental findings.

New and productive theories of affect and self awareness of affect that relate to frustration, learning, perseverance, and intrinsic motivation have been developed. There are two primary theoretical contributions of this thesis: first, defining Stuck as a state of non-optimal experience that directly parallels Flow (chapter 2), and second, defining meta affective skills based on existing theories of metacognition, self theories of intelligence, and affect awareness (chapter 2). Through the experimental methodology (chapter 4), the analysis (chapter 5), and the discussion (chapter 6), these two theoretical contributions were shown to be productive to the design, development, and evaluation of an advanced affective sensing system that is coupled with the ability to drive an Intelligent Tutoring System's affective agent's expressive responses.

This thesis has advanced the pedagogical and communicative abilities of Intelligent Tutoring Systems by demonstrating that they can productively use affective interventions, affective sensor channels, and affective expressions in their interactions with learners. There are four primary system contributions. First, the Affective Agent Research Platform was developed with both real-time multimodal affective sensing and as a responsive social agent capable of sensor driven and pre-recorded non-verbal social interactions (chapter 3). This system is broadly applicable to the study of interpersonal communication and in this thesis it has been specifically refined as an affective learning companion. The second system contribution is the development and training of a classifier that predicted learner's frustration or help seeking with 79% accuracy (chapter 3). Third, interactions that contrast task support vs. affect support (a combination of empathy and meta-affect support) were developed and evaluated in carefully measured and controlled learning environments. The fourth contribution, developed in preparation for the experimental evaluation, was a difficulty assessment for the Towers of Hanoi activity for the 11-13 year-olds (chapter 6, Appendix A).

There are four primary experimental contributions of this evaluation. First, reliable scales were developed for social bond (in terms of positive and negative impressions of the character and in terms of the participants social behavior when saying "bye") and meta-affective skill (that incorporated both learner's awareness of frustration and their ability to use strategies to overcome frustration). Second, the experiment demonstrated that the primary hypothesis was not supported for this age group when genders are combined (chapter 5 and chapter 6). Third, further analysis illuminated gender differences to help explain the outcomes. Fourth, affective interventions were positively associated with girl's meta-affective abilities, higher levels of Flow, and lower levels of Stuck. Fifth, it was demonstrated that the various elements of a character's emotional intelligence should be presented in a coordinated manner. Inconsistencies between the presence or absence of non-verbal social mirroring and the presence or absence of other elements of emotional intelligence (congruence or affective support intervention) were associated with both girl's and boy's frustration.

In the experiment conducted in this thesis, the type of intervention (affect support or task support, see chapter 3), the level of congruence of the intervention with respect to a learner's frustration, and the presence or absence of social non-verbal mirroring played several important and different roles with respect to girl's and boy's frustration, meta-affective abilities, increased Flow and reduced Stuck, and intrinsic motivation. If these findings are confirmed by further studies and if they generalize to broader populations than the participants used in this study, then as Intelligent Tutoring Systems, and other systems that use relational agent strategies, advance to incorporate greater levels of emotional intelligence, developers and researchers should be able to make considerable advances to their systems and to learners' experiences by incorporating these elements of emotional intelligence. At the same time developers and researchers must be careful to appropriately coordinate the diverse elements of emotional intelligence and be well aware of the differences in the impact of these elements on boys and girls.

8 References

Adcock, A. B. and R. Van Eck (2003). The development of an affective measure for use in pedagogical agent tutoring environments. 15th Annual University of Memphis Student Research Forum. Memphis, TN.

Aimeur, E., C. Frasson, et al. (2000). "Cooperative Learning strategies of intelligent tutoring systems." Aplied Artificial Intelligence 14: 465-489.

Allanson, J. and S. H. Fairclough (2004). "A research agenda for physiological computing." Interacting with Computers 16(5): 857-878.

Ambady, N. and R. Rosenthal (1993). " Half a minute: Predicting teacher evaluations from thin slices of behavior and physical attractiveness " Journal of Personality and Social Psychology 64 431-441.

Bailenson, J. N. (2005). Personal Conversation. W. Burleson.

Bailenson, J. N. and N. Yee (2005). "Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments." Psychological Science 16: 814-819.

Baker, R., A. T. Corbett, et al. (2004). Off-task behavior in the cognitive tutor classroom: when students "game the system". SIGCHI Conference on Human Factors in Computing Systems, ACM Press.

Bates, J., A. B. Loyall, et al. (1992). An architecture for action, emotion, and social behavior. The Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World, S. Marino al Cimino, Italy.

Baylor, A. L., E. Shen, et al. (2003). Which pedagogical agent do learners choose? The effects of gender and ethnicity. E-Learn World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education. Phoenix, Arizona.

Baylor, A. L., E. Shen, et al. (2004). Supporting learners with math anxiety: The impact of pedagogical agent emotional and motivational support. International Conference on Intelligent Tutoring Systems. Maceió, Brazil.

Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships, Massachusetts Institute of Technology. Ph.D.

Bickmore, T. and R. W. Picard (2004). "Establishing and Maintaining Long-Term Human-Computer Relationships." Transactions on Computer-Human Interaction 12(2): 293 - 327.

Bishop, D. V. M., G. Aamodt-Leaper, et al. (2001). "Individual differences in cognitive planning on the Tower of Hanoi task: Neuropsychological maturity or measurement error?" Journal of Child Psychology and Psychiatry 42: 551-556.

Blanck, P. D., R. Rosenthal, et al. (1990). "The measure of the judge: An empirically-based framework for exploring trial judges' behavior." Iowa Law Review 75: 653-684.

Boucsein, W. (1992). Electrodermal Activity. New York, NY, Plenum Press.

Brandt, G. (2004). Using Virtual Agents, Avatars and Worlds for Treatment and Healing. Medicine Meets Virtual Reality Conference. Long Beach, CA.

Bull, P. (1983). Body Movement and Interpersonal Communication. New York, John Wiley & Sons.

Burleson, W., R. W. Picard, et al. (2004). A Platform for Affective Agent Research. International Conference on Autonomous Agents and Multiagent Systems. New York, NY, Columbia University

Cassell, J., H. Vilhjálmsson, et al. (2001). BEAT: The Behavior Expression Animation Toolkit. Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH '01). Los Angeles, CA.

Chan, T. W. and A. B. Baskin (1990). Learning companion systems. Intelligent Tutoring Systems: At the Crossroads of Artificial Intelligence and Education. C. Frasson and G. Gauthier. Norwood, NJ, Ablex Publishing.

Conati, C. (2002). "Probabilistic Assessment of User's Emotions in Educational Games." Journal of Applied Artificial Intelligence 16(7-8): 555-575.

Conati, C. (2004). Keynote Presentation. 7th International Conference on Intelligent Tutoring Systems. Maceio, Brazil.

Conati, C. and H. Maclaren (2004). Evaluating A Probabilistic Model of Student Affect. 7th International Conference on Intelligent Tutoring Systems, Maceio, Brazil.

Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. New York, Harper & Row.

Czerwinski, M., E. Horvitz, et al. (2001). Subjective Duration Assessment: An Implicit Probe for Software Usability. IHMHCI.

D'Mello, S. K., S. D. Craig, et al. (2005). Integrating affect sensors in an intelligent tutoring system. 2005 International Conference on Intelligent User Interfaces, New York, AMC Press.

Dennerlein, J., T. Becker, et al. (2003). "Frustrating Computers Users Increases Exposure to Physical Factors." Proceedings of the International Ergonomics Association.

du Boulay, B. and R. Luckin (2001). "Modeling Human Teaching Tactics and Strategies for Tutoring Systems." International Journal of Artificial Intelligence in Education 12(3): 235-256.

Dweck, C. S. (1999). Self-Theories: Their role in motivation, personality and development. Philadelphia, PA, Psychology Press.

Dweck, C. S. (2004). Personal Conversation.

Efklides, A. (2002). "Feelings as subjective evaluations of cognitive processing: How reliable are they?" Psychology: The Journal of the Hellenic Psychological Society 9: 163-184.

Efklides, A. and M. Vauras (1999). "Metacognitive experiences and their role in cognition." European Journal of Psychology of Education XIV(4).

Elliott, C., J. Rickel, et al. (1999). Lifelike pedagogical agents and affective computing: An exploratory synthesis. Artificial Intelligence Today. M. Woodridge and M. Veloso. Chicago, IL, Springer-Verlag. 1600 of Lecture Notes in Computer Science: 195-212.

Espy, K., R. Bull, et al. (2004). "A comparison of performance on the Towers of London and Hanoi in young children." USA Journal of Child Psychology and Psychiatry 45(4): 743-754.

Faste, R. (1996). Personal Communication: Teaching Ambidextrous Thinking.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. The Nature of Intelligence. L. B. Resnick. Hillsdale NJ, Erlbaum.

Flavell, J. H. (1979). "Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry." American Psychologist 34: 906-911.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. Metacognition, Motivation and Understanding. F. E. Weinert and R. H. Kluwe. Hillside, NJ, Erlbaum: 21-29.

Giuseppe, L. and D. Brass (2003). "Exploring the Social Ledger: Negative Relationships and Negative Asymmetry in Social Networks in Organizations." Academy of Management Review.

Goleman, D. (1995). Emotional Intelligence. New York, NY, Bantam Books.

Gottman, J. M. and R. W. Levenson (1992). "Marital Processes Predictive of Later Dissolution." Journal of Personality and Social Psychology 63(2): 221-233.
ABSTRACT. Behavior, Physiology, and Health

Haro, A., I. Essa, et al. (2000). Detecting and Tracking Eyes by Using their Physiological Properties. Computer Vision and Pattern Recognition.

Healey, J. (2000). Wearable and Automotive Systems for the Recognition of Affect from Physiology, Massachusetts Institute of Technology. Ph.D.

Higgins, M. C. (2001). "When is Helping Helpful? Effects of Evaluation and Intervention Timing on Individual Task Performance." Journal of Applied Behavioral Science 37(3): 280-298.

Hill, R., J. Gratch, et al. (2001). Toward the holodeck: integrating graphics, sound, character and story. Fifth International Conference on Autonomous Agents, Montreal, Quebec, Canada.

Johnson, L., J. Rickel, et al. (2000). "Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments." International Journal of Artificial Intelligence in Education 11: 47-78.

Johnson, W. L., S. Kole, et al. (2003). Socially Intelligent Learner-Agent Interaction Tactics. 11th International Conference on Artificial Intelligence in Education (AIED), Amsterdam, IOS Press.

Kapoor, A., W. Burleson, et al. Automatic Prediction of Frustration or Asking for Help. Working Paper.

Kapoor, A., S. Mota, et al. (2001). Towards a learning companion that recognizes affect. AAAI Fall Symposium.

Kapoor, A. and R. W. Picard (2003). Fully-Automatic Upper Facial Action Recognition. EEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003), ICCV, Nice, France.

Kapoor, A. and R. W. Picard (2005). Multimodal Affect Recognition in Learning Environments. ACM MM'05, Singapore.

Kapoor, A., R. W. Picard, et al. (2004). Probabilistic Combination of Multiple Modalities to Detect Interest. International Conference on Pattern Recognition, Cambridge, U.K.

Kay, A. (1991). Computers, Networks and Education. Scientific American.

Klein, J., Y. Moon, et al. (2002). "This Computer Responds to User Frustration: Theory, Design, Results, and Implications." Interacting with Computers 14: 119-140.

Koda, T. and P. Maes (1996). Agents with Faces: The Effect of Personification. HCI'96.

Kort, B., R. Reilly, et al. (2001). An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion International Conference on Advanced Learning Technologies.

LaFrance, M. (1982). Posture Mirroring and Rapport. Interaction Rhythms: Periodicity in Communicative Behavior. M. Davis. New York, Human Sciences Press, Inc.: 279-298.

Lepper, M. R. (2004). Personal Conversation.

Lepper, M. R., M. Woolverton, et al. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. Computers as cognitive tools S. P. Lajoie and S. J. Derry. Hillsdale, NJ, Erlbaum: 75-105

Lester, J., C. Callaway, et al. (2001). Animated Pedagogical Agents in Knowledge-Based Learning Environments. Smart Machines in Education. . K. Forbus and P. Feltovich. Menlo Park, CA, AAAI Press: 269-298.

Lester, J., S. G. Towns, et al. (1999). "Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents." International Journal of Artificial Intelligence in Education 10: 278-291.

Lester, J. C., S. A. Converse, et al. (1997). The persona effect: Affective impact of animated pedagogical agents. CHI'97, New York, NY, ACM Press.

Liu, K. and R. W. Picard (2005). Embedded Empathy in Continuous, Interactive Health Assessment. CHI Workshop on HCI Challenges in Health Assessment, Portland, Oregon.

Malone, T. (1981). Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games, Association for Computing Machinery.

Marci, C. and H. Riess (2005). "The clinical relevance of psychophysiology: Support for the psychobiology of empathy and psychodynamic process." American Journal of Psychotherapy 59: 213-226.

Marsella, S. C. and W. L. Johnson (2003). Interactive pedagogical drama for health interventions. Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies. U. Hope et al. Amsterdam, IOS Press: 341-348.

Monk, A. (2000). "Computer's and Fun."

Moon, Y. (2000). "Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers." Journal of Consumer Research: 323-339.

Morency, L. P., A. Rahimi, et al. (2003). Adaptive View-based Appearance Model CVPR 2003, Wisconsin.

Morency, L. P., P. Sundberg, et al. (2003). Pose Estimation using 3D View-Based Eigenspaces. ICCV Workshop on Analysis and Modeling of Face and Gesture, Nice, France.

Mota, S. and R. W. Picard (2003). Automated Posture Analysis for Detecting Learner's Interest Level. 1st IEEE Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI.

Ortony, A., G. L. Clore, et al. (1988). The Cognitive Structure of Emotions, Cambridge University Press.

Papert, S. (1980). Mindstorms: children, computers, and powerful ideas. New York, NY, Basic Books.

Perlin, K. (1997). Layered Compositing of Facial Expression. ACM SIGGRAPH 97.

Picard, R. W. (1997). Affective Computing. Cambridge, MA, MIT Press.

Reeves, B. and C. Nass (1996). The Media Equation. Cambridge, Cambridge University Press.

Reynolds, C. (1999). The Sensing and Measurement of Frustration with Computers, Massachusetts Institute of Technology. Masters.

Rich, C., C. Sidner, et al. (2001). "Collagen: Applying collaborative discourse theory to human--computer interaction." AI Magazine, Special Issue on Intelligent User Interfaces 22(4): 15-25.

Robinson, D. and L. Smith-Lovin (1999). "Emotion Display as a Strategy for Identity Negotiation." Motivation and Emotion 23(2).

Rottenberg, J., R. D. Ray, et al. (in press). Emotion elicitation using films. The handbook of emotion elicitation and assessment. J. A. Coan and J. J. B. Allen. London, Oxford University Press.

Rueda, C. (1997) An optimal solution to the Towers of Hanoi Puzzle. Volume, DOI:

Ryokai, K., C. Vaucelle, et al. (2003). "Virtual Peers as Partners in Storytelling and Literacy Learning." Journal of Computer Assisted Learning 19(2): 195-208.

Salovey, P. and J. D. Mayer (1990). "Emotional intelligence." Imagination, Cognition and Personality 9(3): 185-211.

Schank, R. and A. Neaman (2001). Motivation and Failure in Educational Systems Design. Smart Machines in Education. K. Forbus and P. Feltovich. Cambridge MA, AAAI Press/ The MIT Press.

Schneider, K. and I. Josephs (1991). "The expressive and communicative functions of preschool children's smiles in an achievement situation." Journal of Nonverbal Behavior 15: 185-98.

Shaw, E., C. LaBore, et al. (2004). "Animating 2D Digital Puppets with Limited Autonomy." Smart Graphics: 1-10.

Simon, H. A. (1975). "The Functional Equivalence of Problem Solving Skills." Cognitive Psychology 7: 268-288.

Strain, P. (1981). The Utilization of Classroom Peers as Behavior Change Agents. New York, Plenum Press.

Strauss, M., C. Reynolds, et al. (2005). The HandWave Bluetooth Skin Conductance Sensor. 1st International Conference on Affective Computing and Intelligent Interaction, Beijing, China.

Tao, J. and Y. Kang (2005). Features Importance Analysis for Emotional Speech Classification. ACII 2005.

Tekscan (1997). Tekscan Body Pressure Measurement System User's Manual. South Boston, MA, USA. , Tekscan Inc.

Tudge, J. and D. Caruso (1988) Cooperative Problem-Solving in the Classroom Volume, DOI: 198, EDO-PS-88-1, 1988

Videre Design (2004) MEGA-DCS Stereo Camera website. Volume, DOI:

Webb, M. (1987) Peer Helping Relationships in Urban Schools. ERIC Digest, US Volume, DOI: 1987-12-00. ERID/CUE-D-37

Welsh, M. C., T. Satterlee-Cartmell, et al. (1999). "Towers of Hanoi and London: Contribution of working memory and inhibition to performance." Brain and Cognition 41: 231-242.

Wentzel, K. (1997). "Student Motivation in Middle School: The Role of Perceived Pedagogical Caring." Journal of Educational Psychology 89(3): 411-419.

Weybrew, B. B. (1984). "The Zeigarnik phenomenon revisited: Implications for enhancement of morale." Perceptual and Motor Skills 58: 223-226.

Zeigarnik, B. (1967). On finished and unfinished tasks. A sourcebook of Gestalt psychology. W. D. Ellis. New York, NY, Humanities Press.

Appendix A. Pilot Studies

Prior to the experiment discussed in chapter 4 a series of pilot studies were conducted that served to test and refine the system and the experimental evaluation. In the process of creating such a complex system and in the preparation of the experimental protocol there were numerous parameters that were investigated (e.g. the design of character interactions, the nature and difficulty of the task, the duration of the protocol, the training of the classifier algorithms, the robustness of the system, etc.). Although a great deal of testing and debugging was conducted during the programming and development, pilot testing was essential to inform the design direction with experience gained from participant's interactions with the system, the Towers of Hanoi activity, the character, and the experimental sequence.

The initial study was conducted with 10 children, six recruited to come to the Media Lab and four who used the system during an initial test deployment, at the Charlestown Computer Club House. At this time the character system was still under development, so it was not used in this test. This test provided initial impressions of the difficulty of the towers of Hanoi, some information on the duration it would take for participants to complete the survey and the activity as well as much insight on the functionality of the system. The study had a pre-test, full screen slide show, the Towers of Hanoi activity with 3, 4 and 5 disks with questions on the level of frustration during each experience, and a post-test.

This initial study found that the head-tracking equipment was not robust enough to provide sustained tracking reliably for periods longer than 10 minutes. The experiment would require the sensors to work for more than 20 minutes, so the Head-Tracking equipment was removed from the system. It would be interesting to reincorporate such a device in the future, especially in light of Bailenson's research showing increased liking of and persuasion by an agent that mirrored participant's head movements (Bailenson, 2005). It was found that trying to gather so many datagram packets from five different sensor sources, simultaneously, resulted in many packets being dropped. Through successive programming efforts the receipt of datagrams was substantially improved. Eventually the datagram module of the Data Logger became a multi-threaded application in which one thread would continuously look for new information on the socket and the other would package the current information into a TCP packet and transmit it to the System Server (see appendix D).

The second pilot study was conducted at Lanesborough School, in Lanesborough MA with 21 participants over three days. With the basic functionality of the sensor system improved, the character and integrated Towers of Hanoi activity were introduced to participants. This pilot study was conducted with iterative improvements to the character interactions along with several changes to the presentation of the activity, between each of the three days.

On the first day participants were presented with the Towers of Hanoi with four disks, followed by five disks, followed by six disks. Many participants were able to complete all three activities. The second day participants were presented with five disks followed by six disks. If they finished each of these they were then asked if they would like an easier or harder activity. Regardless of their answer they were presented with six disks again and told to try to do the activity in as few moves as possible. It was interesting to note that the perception of participants regarding the difficulty of the second presentation of the six disk activity was biased by their expectation that the experimenter would comply with their request for an easier or harder activity. Although the activity they received was identical to the one they had just completed, participants justified why they believed it was easier or harder with comments such as, it was easier because I had already done it, or, it was harder because I had to do it in fewer moves this time.

On the third day the option to disengage from the activity was introduced. At the beginning of the activity the experimenter provided participants with paper cards that they could use to indicate they wanted to "quit", two of the three subjects used the cards. On all three days, after the post-test survey, subjects were informally interviewed about their experience and their impressions of the character. Some found the character to be dumb while others liked having it present. Several indicated that they thought the character could understand how they felt. Most described the character as male while a few described it as female. Impressions of the age of the characters voice were also obtained to see if this aspect of the character would lead participants to consider it as a peer learning companion. The impressions of the age of the voice ranged from some participants saying that it was younger older and others saying that it was older; the decision was made to continue using the same voice.

The third pilot study was conducted at the Wellesley Middle School, in Wellesley MA, with 31 participants (Kapoor, Burleson et al.). Since the design of the behavior for the *affect support* and the *task support* interventions were still being developed, this pilot study presented the Towers of Hanoi activity with six disks and used a two factor design with *sensor-driven non-verbal mirroring* as one condition and *Perlin noise - static character* as a second condition. The *Perlin noise - static character* condition applied Perlin Noise using variable values which changed pseudo-randomly to create life-like subtle movements in the character over time (Perlin, 2002). This technique tends to give the viewer the impression that the image of the character is actually a life-like character that is standing relatively still, rather than a two dimensional image of a character that looks more like a picture. In the *Perlin noise - static character* condition the character stands and observes the participant engaged in the activity as a control for the *Sensor-driven non-verbal mirroring* character in which the characters interactions are driven by the sensors, as described in chapter 4. In this study the quit buttons were presented by the character and were present on the screen throughout the activity. While there were 31 participants in this study, there were complete sensor data sets for 24 participants. The data sets were used to train the classifier algorithms (see chapter 3 for a discussion of the classifier) resulting in the ability to detect "quitting" with 79% accuracy (chance = 58%).

It was found that several participants were able to complete the six disk Towers of Hanoi activity so this lead to the decision to use seven disks in the final experiment. It was also found that the mapping for the head nod/head shake detection in the blue eyes camera sensor was causing the character in the in *sensor-driven non-verbal mirroring* condition to shake its head excessively. This effect seen in the observers of ping-pong or tennis, while participants engage in the Towers of Hanoi they look back and forth as the move the disk laterally from pole to pole. The mapping was altered so that head tilt would be mirrored rather than head nod / head shake. This individual interaction was then tested on several volunteers at the Media Lab prior to its deployment in the subsequent experiment; the methodology of the experiment is described in chapter 4.

Appendix B. Survey Instruments used in the Experiment

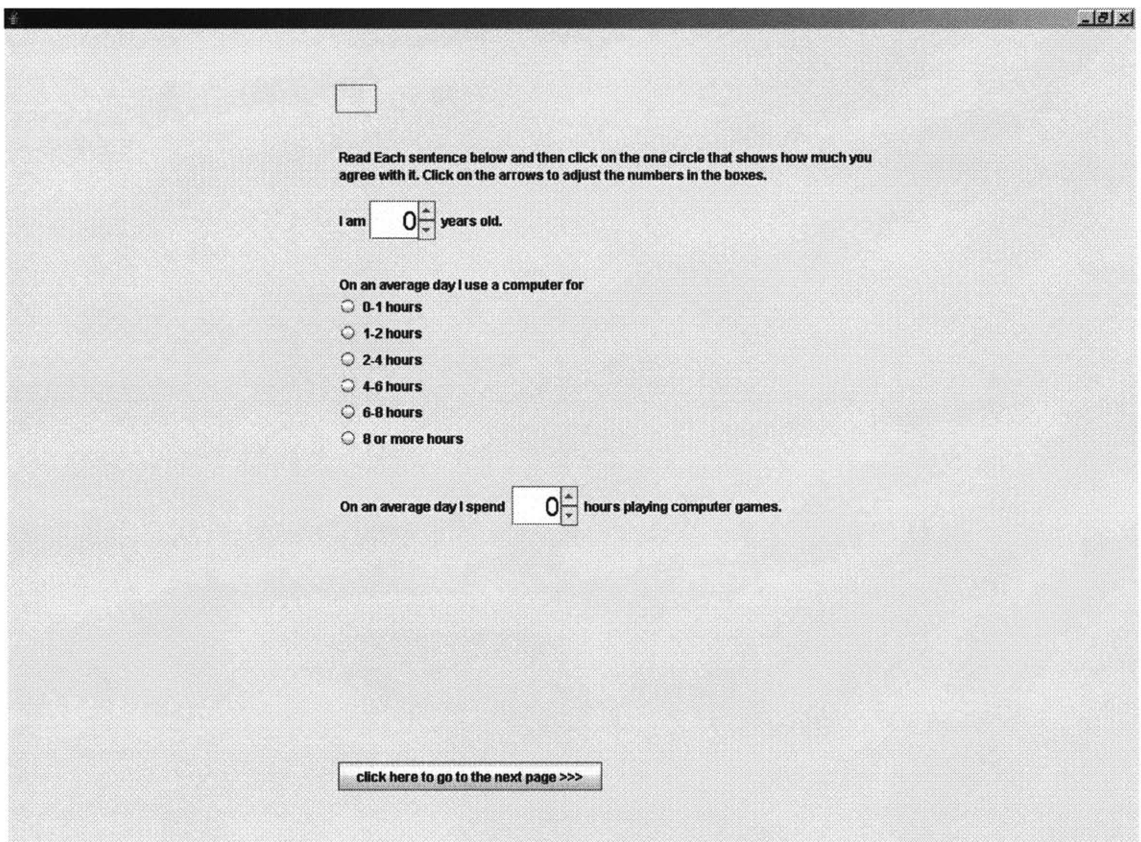
This question was provided to the participant's teachers:

Imagine a scenario where you have the opportunity to work one-on-one with each of the following students and to present him or her with a task that would be challenging to him or her. In this scenario after presenting the task you leave the room. The student now has the opportunity to decide when he or she will stop doing the task.

For each of the following students please rate on a scale of 1-7 how long you think he or she will spend on the task that you presented to him or her.

Please provide a rating only for students you have taught within the last year.
Participant #, First Name, Last Name.
(not long at all) 1 ---- 2 ---- 3 ---- 4 ---- 5 ---- 6 ---- 7 (very long)

The following questions were provided to participants in the experiment.


Read Each sentence below and then click on the one circle that shows how much you agree with it. Click on the arrows to adjust the numbers in the boxes.

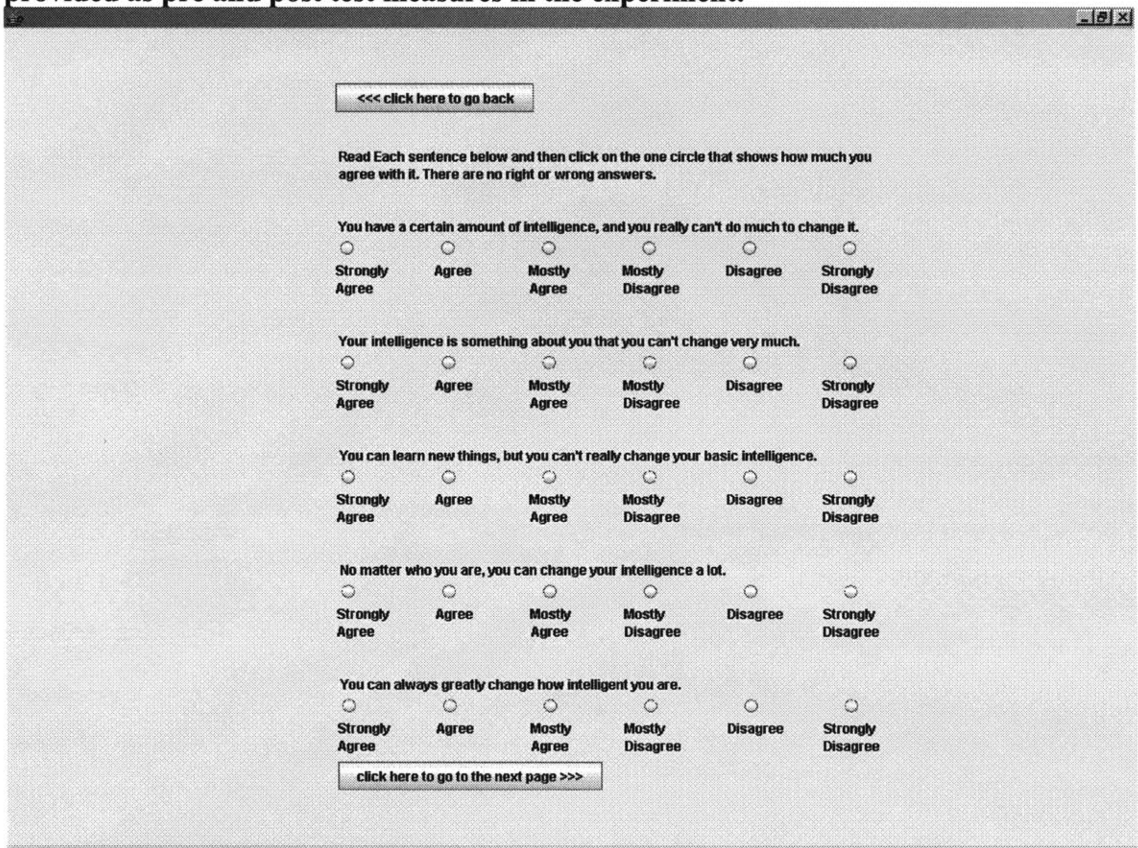
I am 0 years old.

On an average day I use a computer for
- 0-1 hours
- 1-2 hours
- 2-4 hours
- 4-6 hours
- 6-8 hours
- 8 or more hours

On an average day I spend 0 hours playing computer games.

click here to go to the next page >>>

The following questions measure Self theories of Intelligence (Dweck 1999). They were provided as pre and post test measures in the experiment.



<<< click here to go back

Read Each sentence below and then click on the one circle that shows how much you agree with it. There are no right or wrong answers.

You have a certain amount of intelligence, and you really can't do much to change it.

○ Strongly Agree ○ Agree ○ Mostly Agree ○ Mostly Disagree ○ Disagree ○ Strongly Disagree

Your intelligence is something about you that you can't change very much.

○ Strongly Agree ○ Agree ○ Mostly Agree ○ Mostly Disagree ○ Disagree ○ Strongly Disagree

You can learn new things, but you can't really change your basic intelligence.

○ Strongly Agree ○ Agree ○ Mostly Agree ○ Mostly Disagree ○ Disagree ○ Strongly Disagree

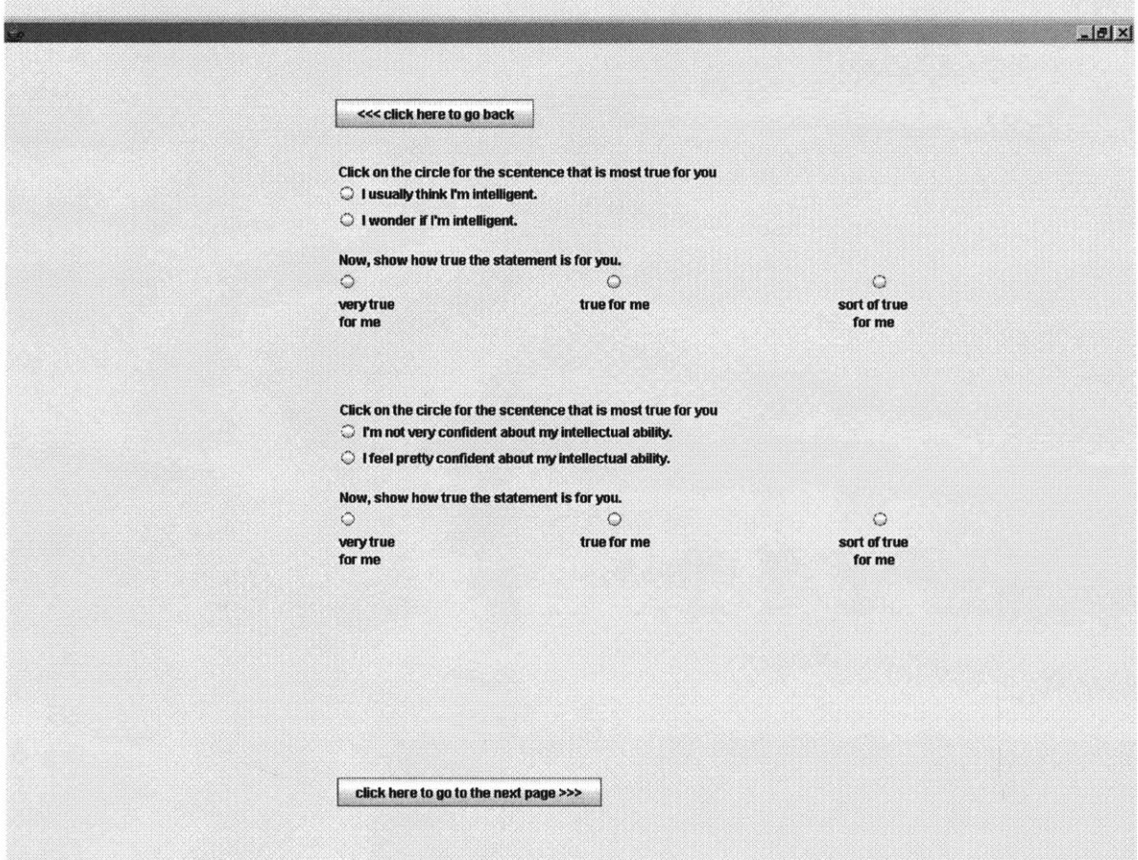
No matter who you are, you can change your intelligence a lot.

○ Strongly Agree ○ Agree ○ Mostly Agree ○ Mostly Disagree ○ Disagree ○ Strongly Disagree

You can always greatly change how intelligent you are.

○ Strongly Agree ○ Agree ○ Mostly Agree ○ Mostly Disagree ○ Disagree ○ Strongly Disagree

click here to go to the next page >>>



<<< click here to go back

Click on the circle for the scentence that is most true for you

○ I usually think I'm intelligent.

○ I wonder if I'm intelligent.

Now, show how true the statement is for you.

○ very true for me ○ true for me ○ sort of true for me

Click on the circle for the scentence that is most true for you

○ I'm not very confident about my intellectual ability.

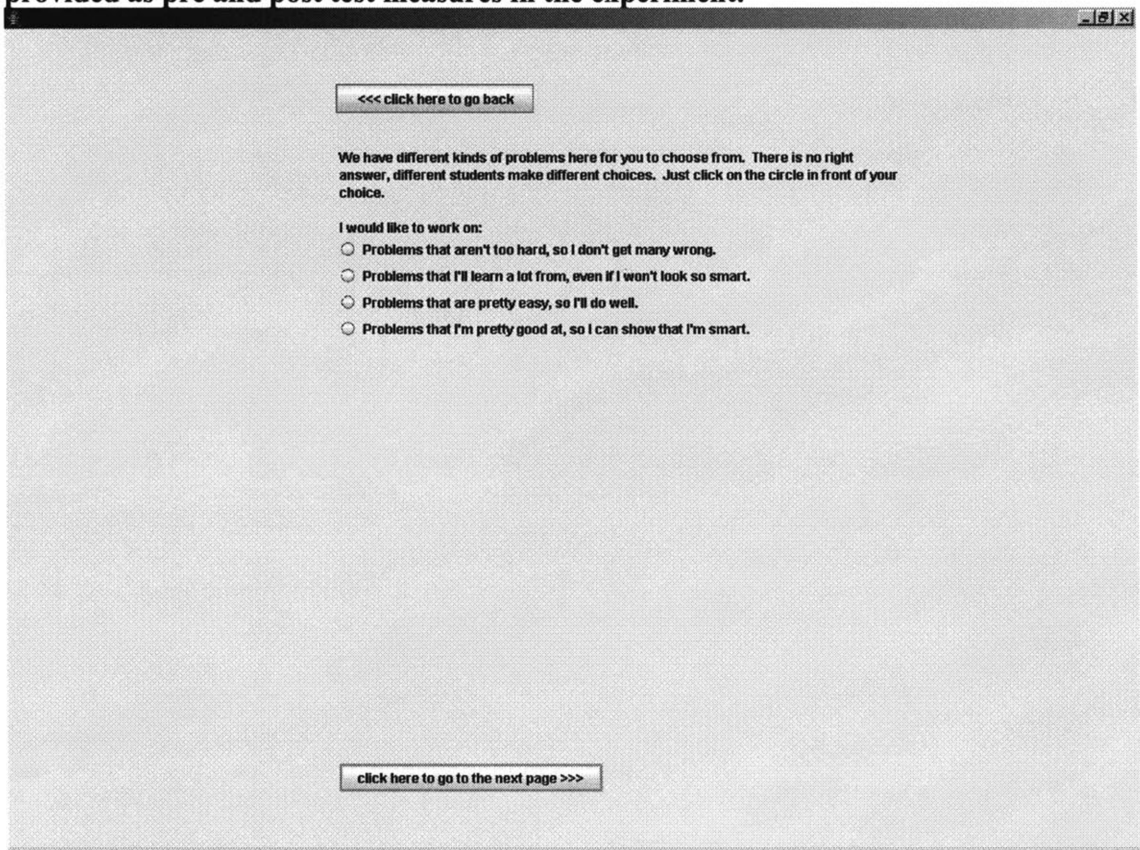
○ I feel pretty confident about my intellectual ability.

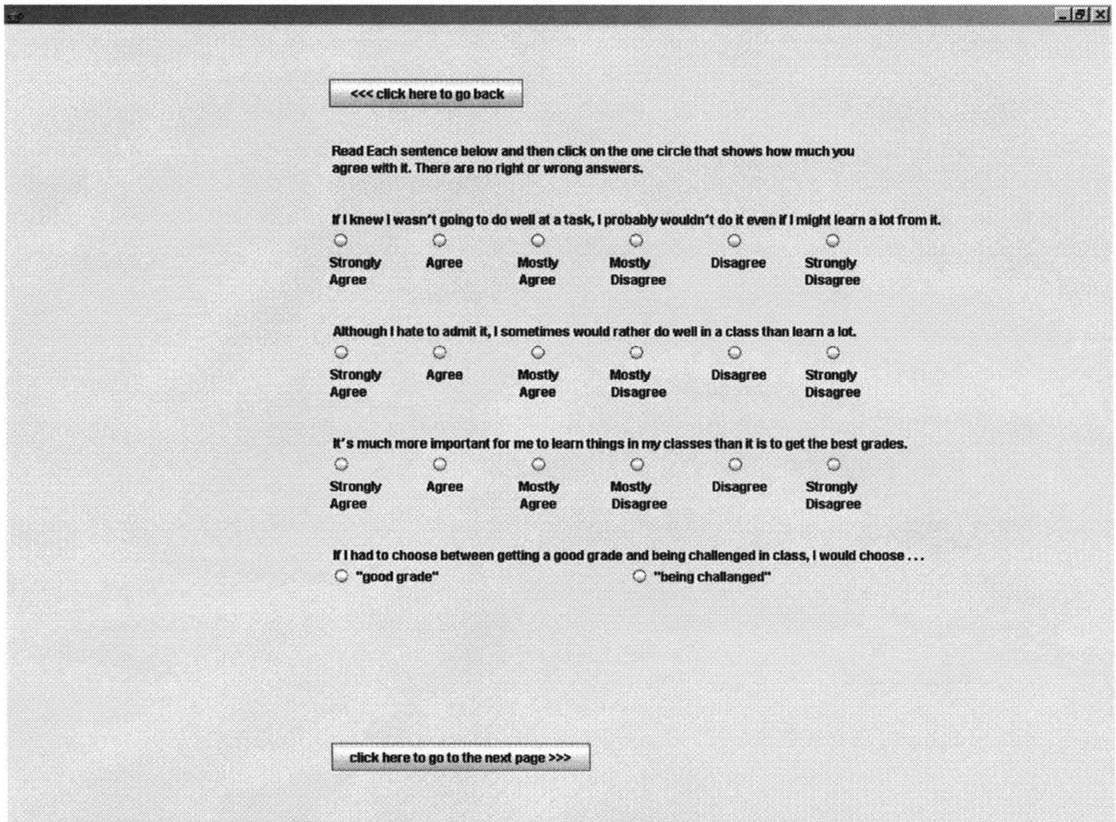
Now, show how true the statement is for you.

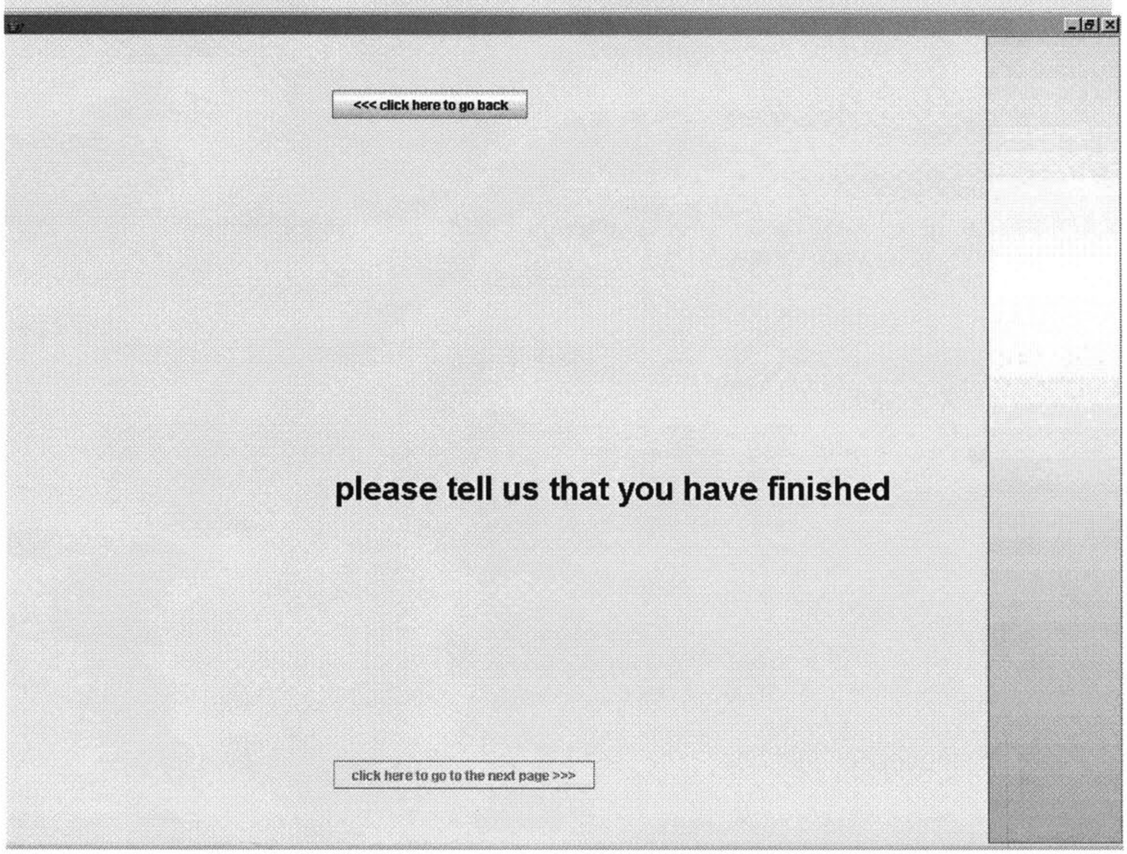
○ very true for me ○ true for me ○ sort of true for me

click here to go to the next page >>>

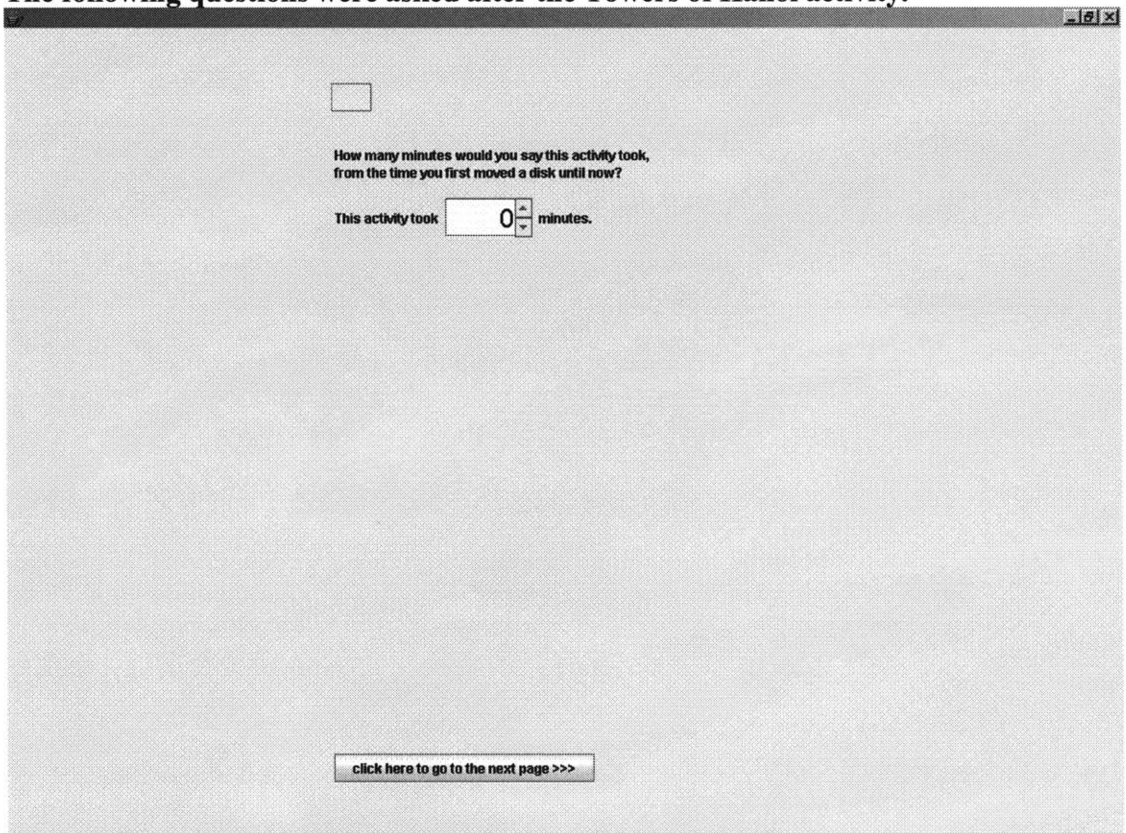
The following questions measure Goal Mastery Orientation (Dweck 1999). They were provided as pre and post test measures in the experiment.




<<< click here to go back
Read Each sentence below and then click on the one circle that shows how much you agree with it. There are no right or wrong answers.
If I knew I wasn't going to do well at a task, I probably wouldn't do it even if I might learn a lot from it.
Strongly Agree
Agree
Mostly Agree
Mostly Disagree
Disagree
Strongly Disagree
Although I hate to admit it, I sometimes would rather do well in a class than learn a lot.
Strongly Agree
Agree
Mostly Agree
Mostly Disagree
Disagree
Strongly Disagree
It's much more important for me to learn things in my classes than it is to get the best grades.
Strongly Agree
Agree
Mostly Agree
Mostly Disagree
Disagree
Strongly Disagree
If I had to choose between getting a good grade and being challenged in class, I would choose . . .
"good grade"
"being challanged"
click here to go to the next page >>>


<<< click here to go back
please tell us that you have finished
click here to go to the next page >>>

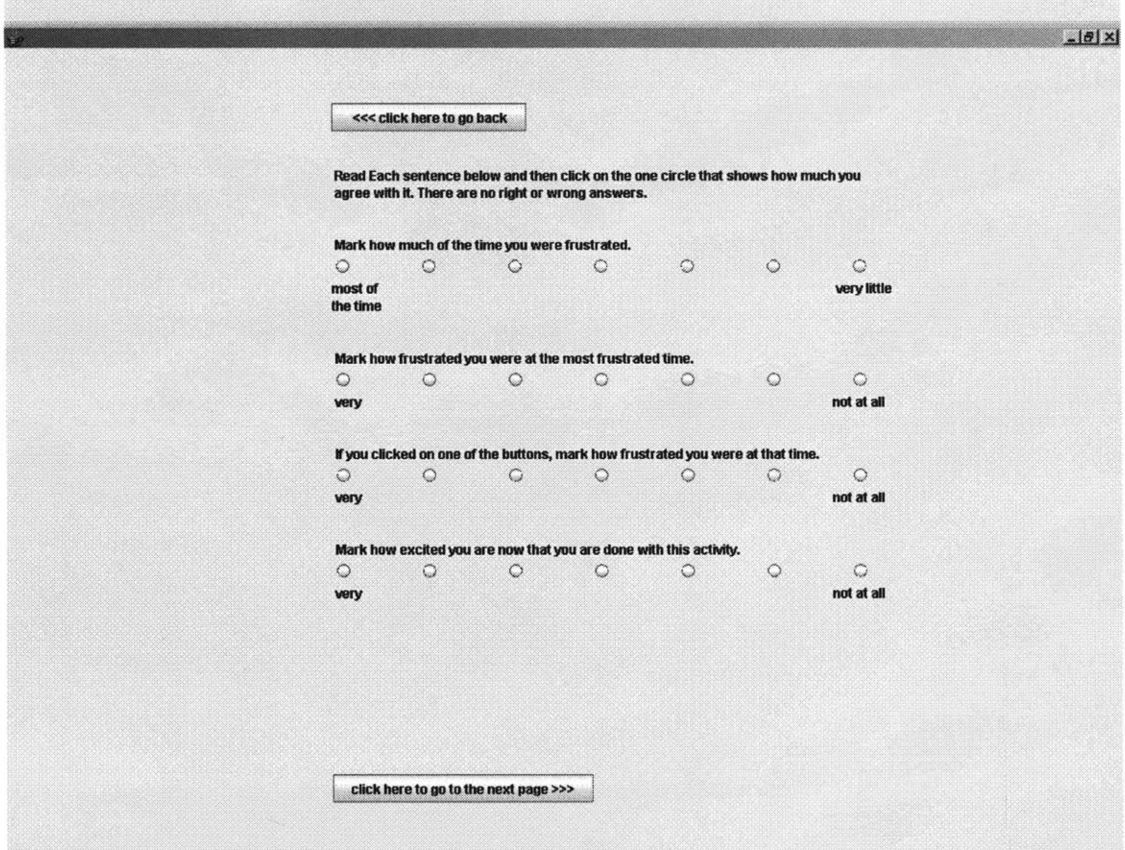
The following questions were asked after the Towers of Hanoi activity.



How many minutes would you say this activity took,
from the time you first moved a disk until now?

This activity took 0 minutes.

click here to go to the next page >>>



<<< click here to go back

Read Each sentence below and then click on the one circle that shows how much you agree with it. There are no right or wrong answers.

Mark how much of the time you were frustrated.

most of the time — very little

Mark how frustrated you were at the most frustrated time.

very — not at all

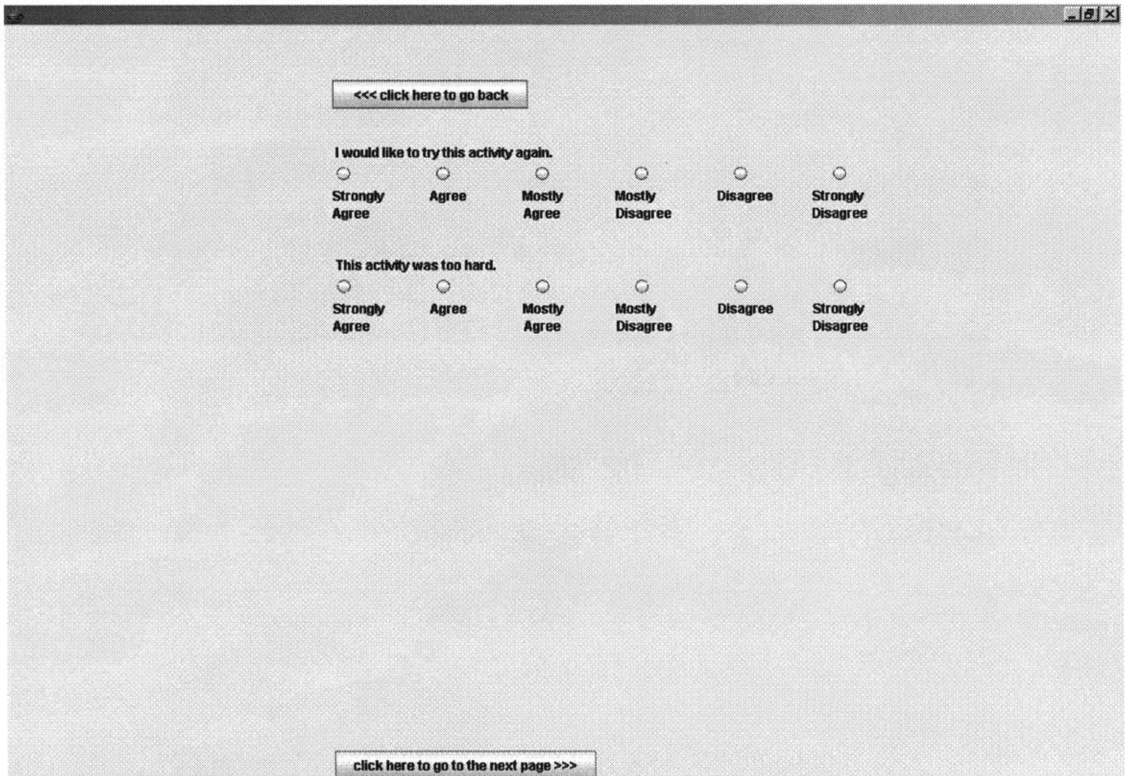
If you clicked on one of the buttons, mark how frustrated you were at that time.

very — not at all

Mark how excited you are now that you are done with this activity.

very — not at all

click here to go to the next page >>>



<<< click here to go back

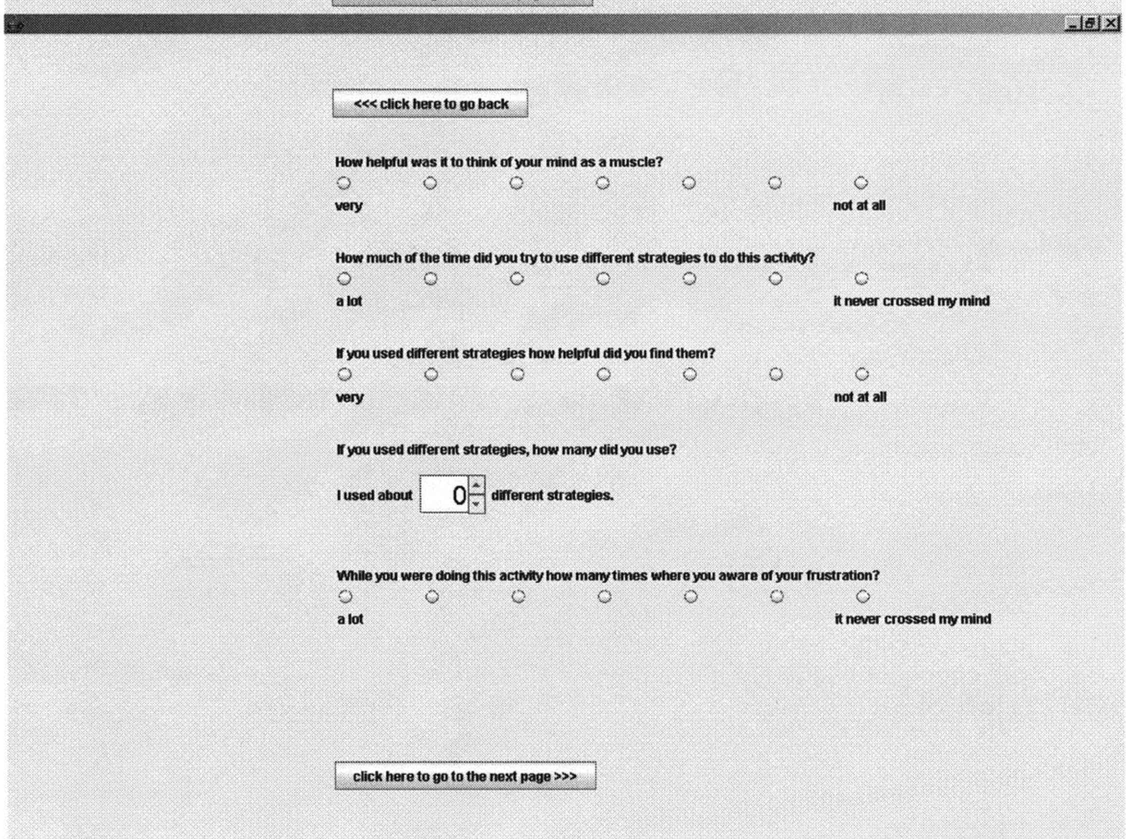
I would like to try this activity again.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

This activity was too hard.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

click here to go to the next page >>>



<<< click here to go back

How helpful was it to think of your mind as a muscle?

very … not at all

How much of the time did you try to use different strategies to do this activity?

a lot … it never crossed my mind

If you used different strategies how helpful did you find them?

very … not at all

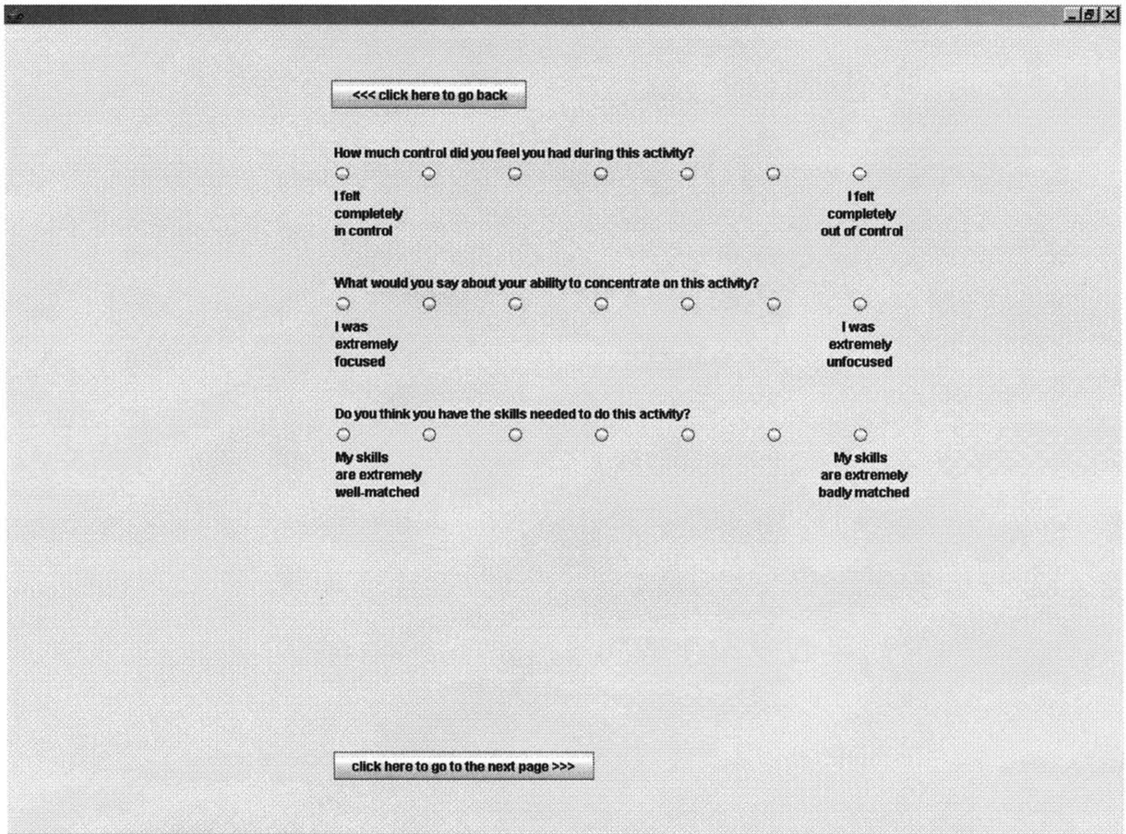
If you used different strategies, how many did you use?

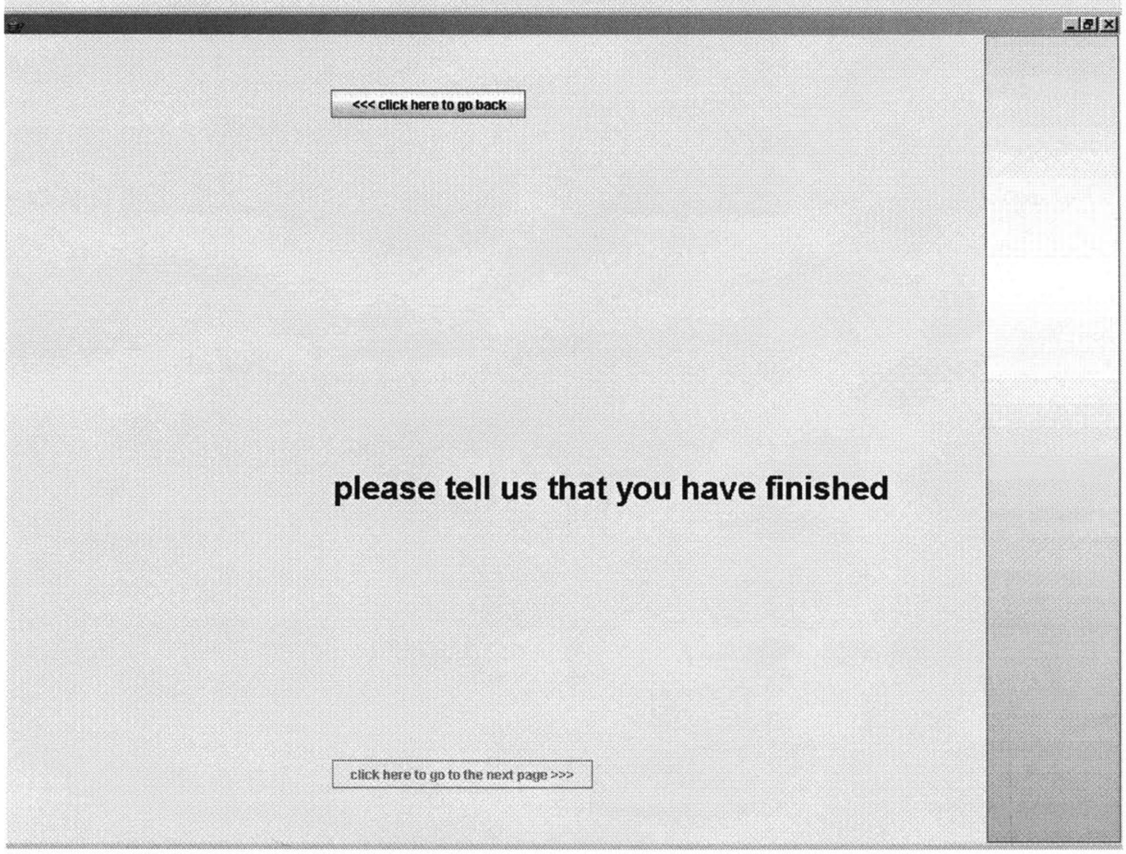
I used about 0 different strategies.

While you were doing this activity how many times where you aware of your frustration?

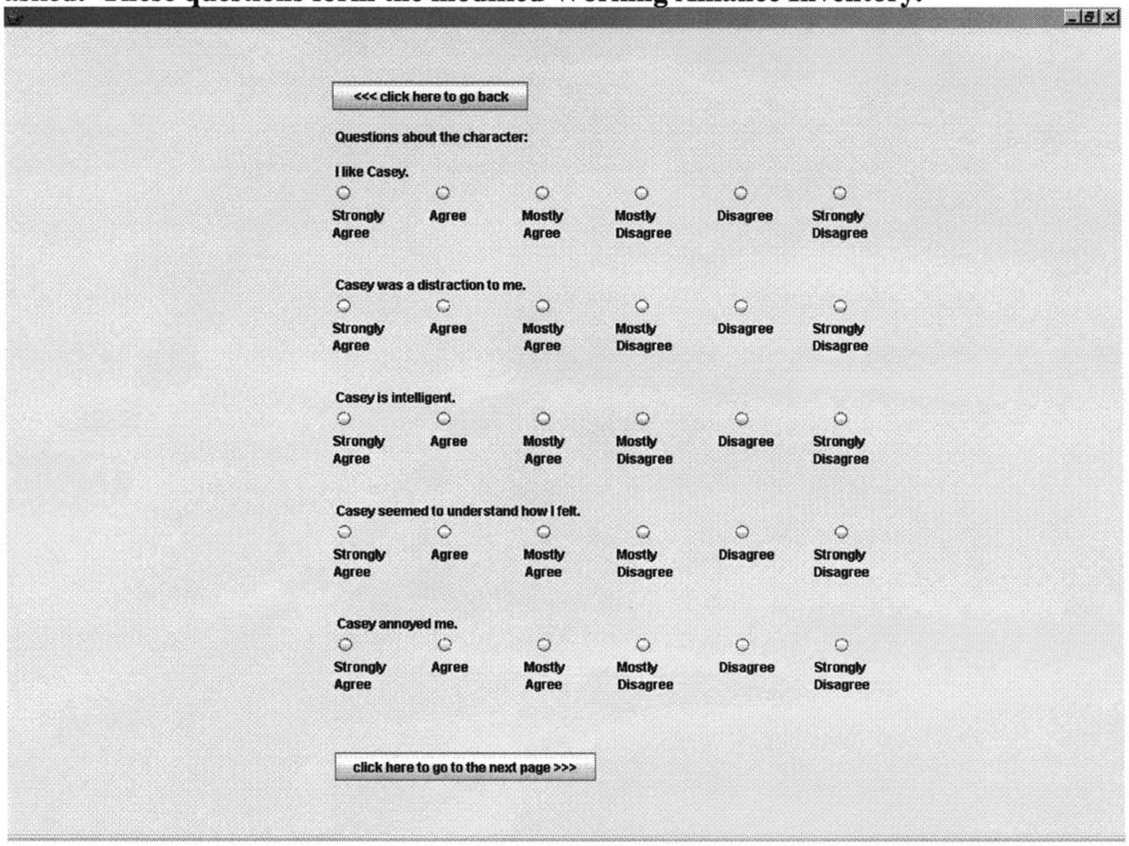
a lot … it never crossed my mind

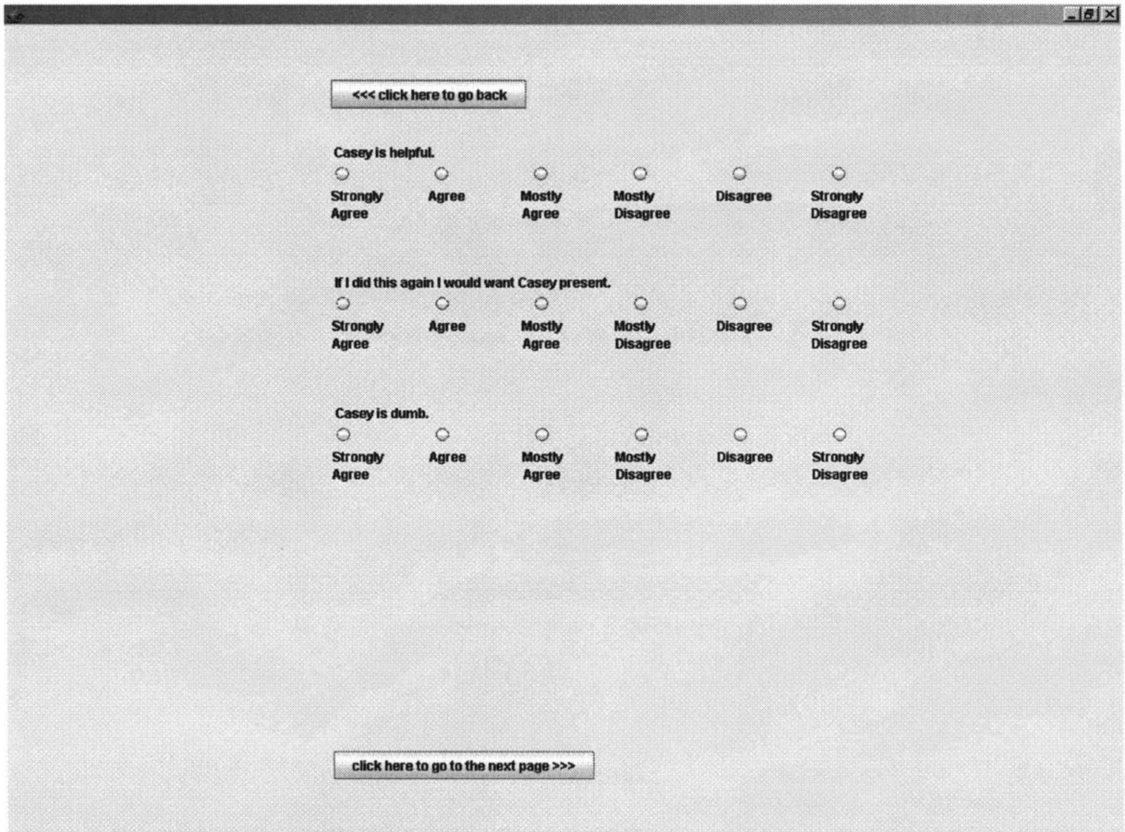
click here to go to the next page >>>


<<< click here to go back
How much control did you feel you had during this activity?
I felt completely in control
I felt completely out of control
What would you say about your ability to concentrate on this activity?
I was extremely focused
I was extremely unfocused
Do you think you have the skills needed to do this activity?
My skills are extremely well-matched
My skills are extremely badly matched
click here to go to the next page >>>


<<< click here to go back
please tell us that you have finished
click here to go to the next page >>>

The previous questions were followed by a 1.5 minute video to neutralize the effects of participant's frustration (see chapter 4 for more detail on the video). The post test of the self theories of intelligence and goal mastery orientation were then provided (the questions were identical to the screens shown above). The following questions were then asked. These questions form the modified Working Alliance Inventory.





<<< click here to go back

Casey is helpful.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

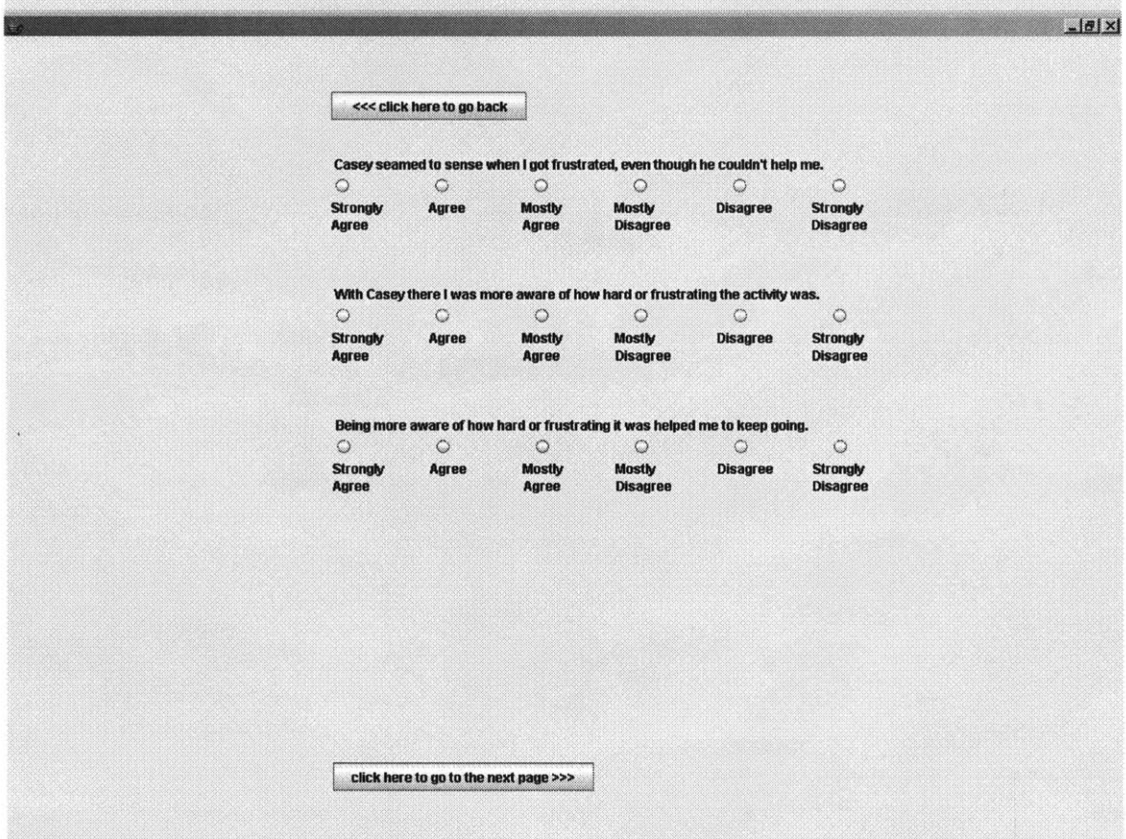
If I did this again I would want Casey present.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

Casey is dumb.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

click here to go to the next page >>>



<<< click here to go back

Casey seamed to sense when I got frustrated, even though he couldn't help me.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

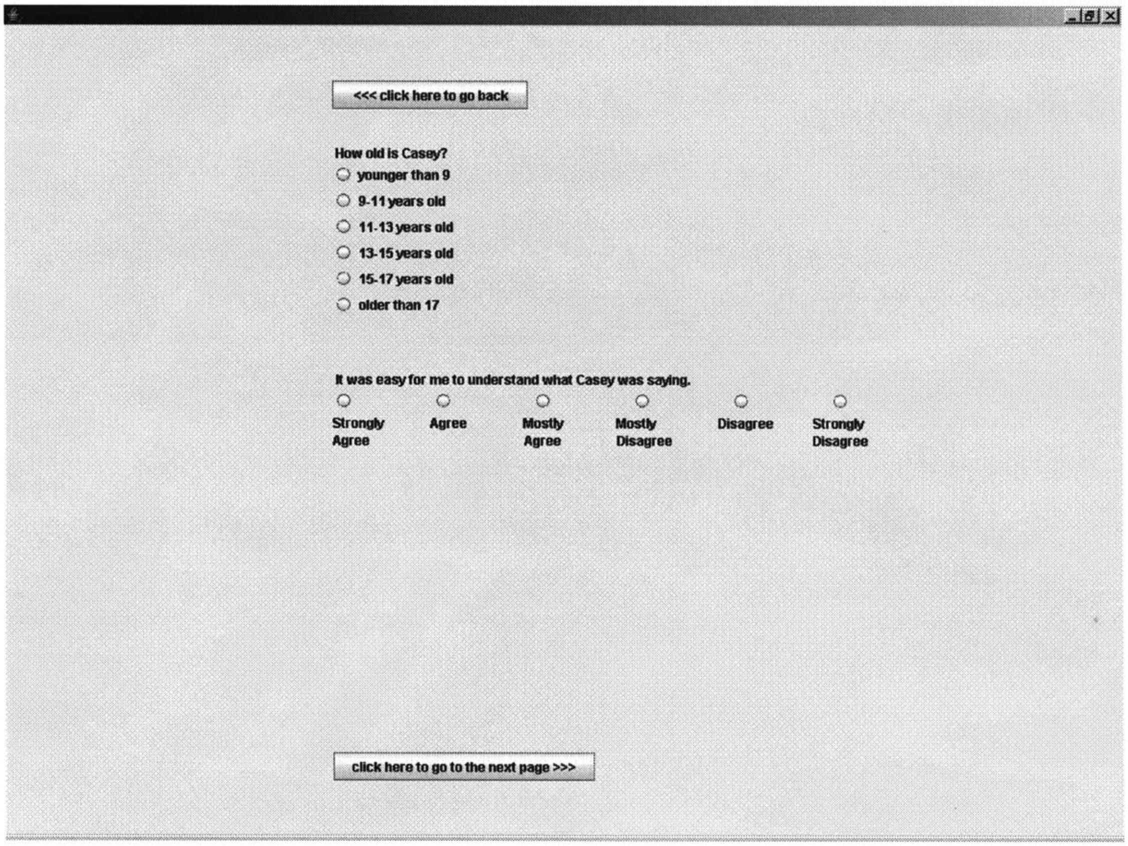
With Casey there I was more aware of how hard or frustrating the activity was.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

Being more aware of how hard or frustrating it was helped me to keep going.

Strongly Agree | Agree | Mostly Agree | Mostly Disagree | Disagree | Strongly Disagree

click here to go to the next page >>>



<<< click here to go back

How old is Casey?

- younger than 9
- 9-11 years old
- 11-13 years old
- 13-15 years old
- 15-17 years old
- older than 17

It was easy for me to understand what Casey was saying.

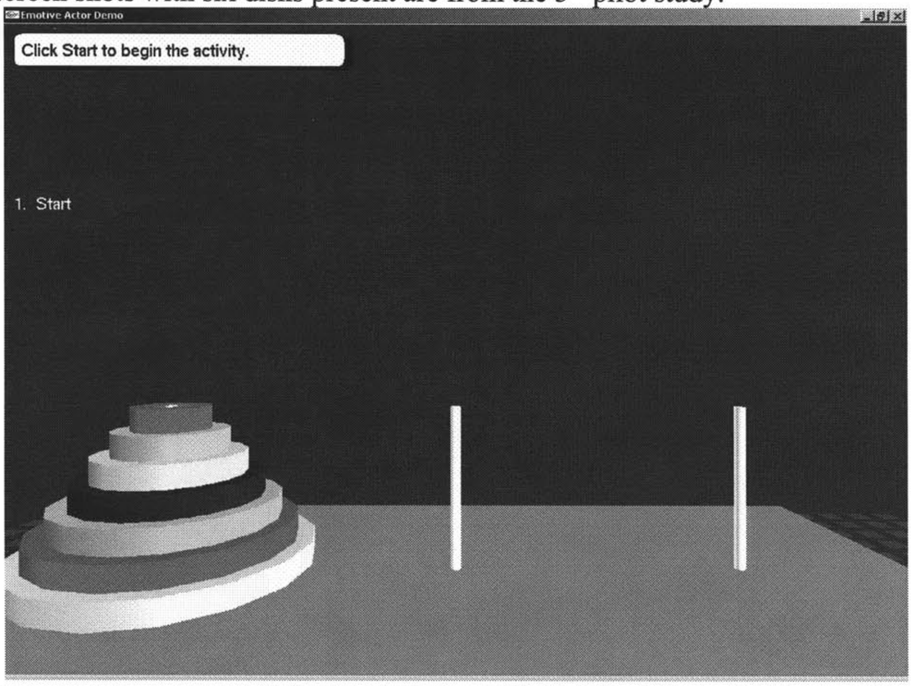
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○

click here to go to the next page >>>

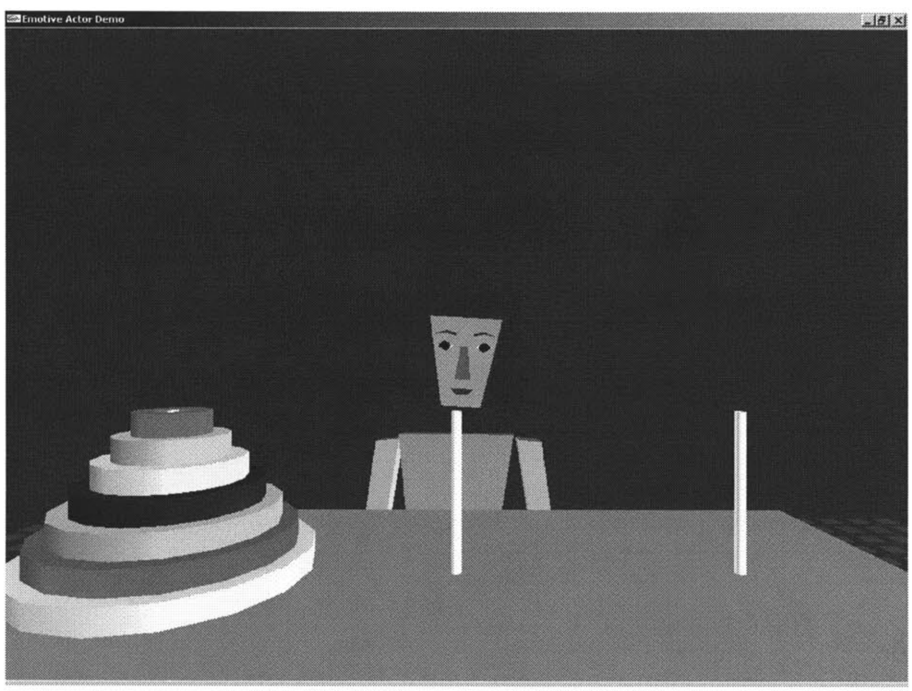
Appendix C. Character Interactions

Screen Shots of the Character Interactions and User Experience

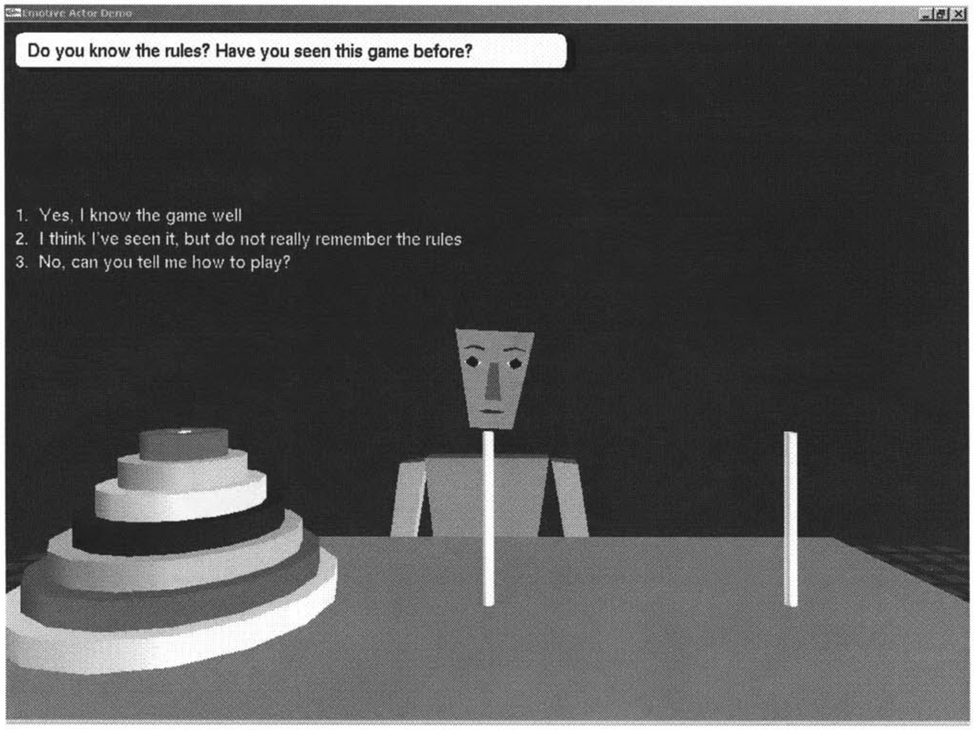
The following are screen shots of the major events in the character interactions and user experience. The final screen shots with six disks present are from the 3rd pilot study.



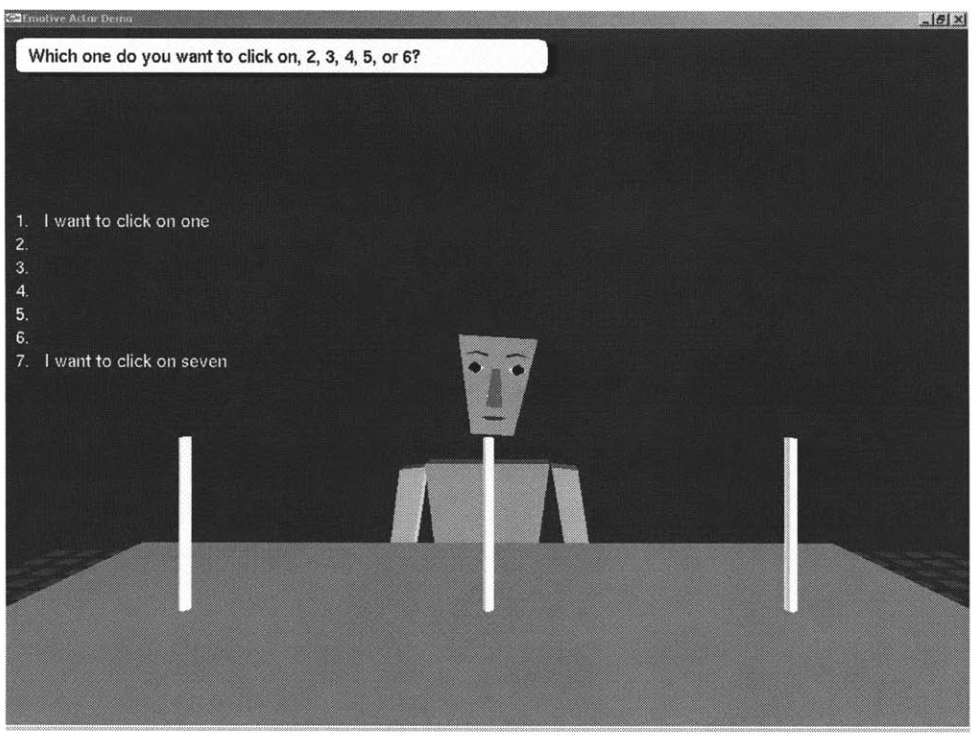
Experimenter instructs participant to click on Start. The Character then appears and introduces itself, the activity, and presents the slide show.



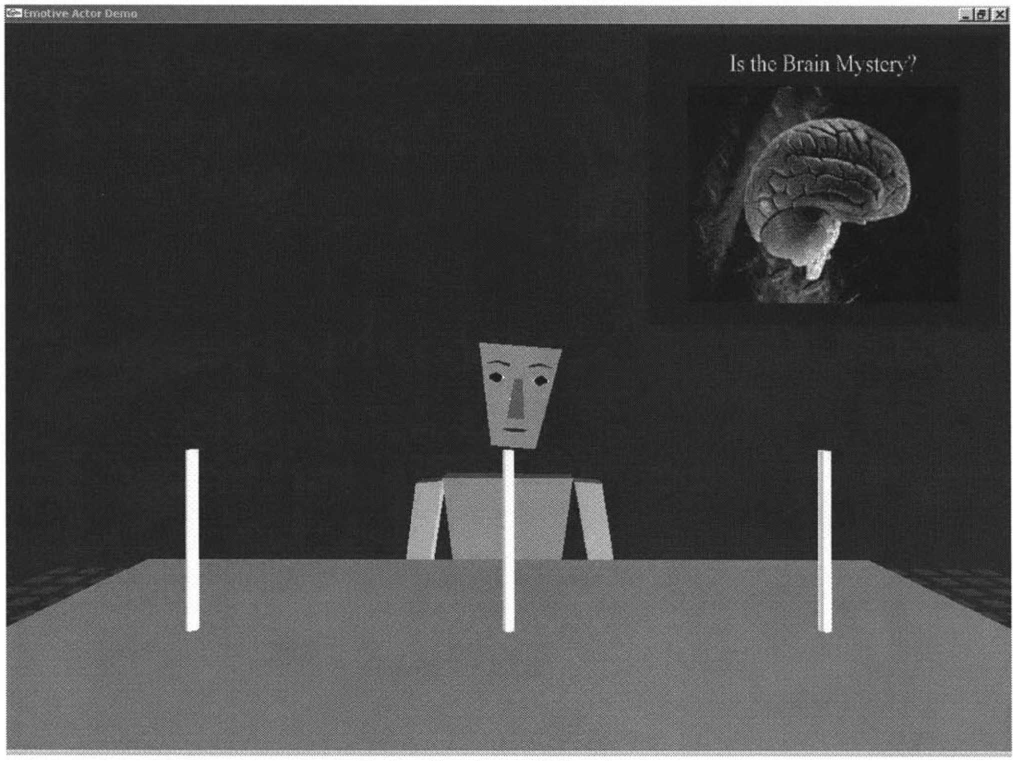
The character appears and introduces the activity



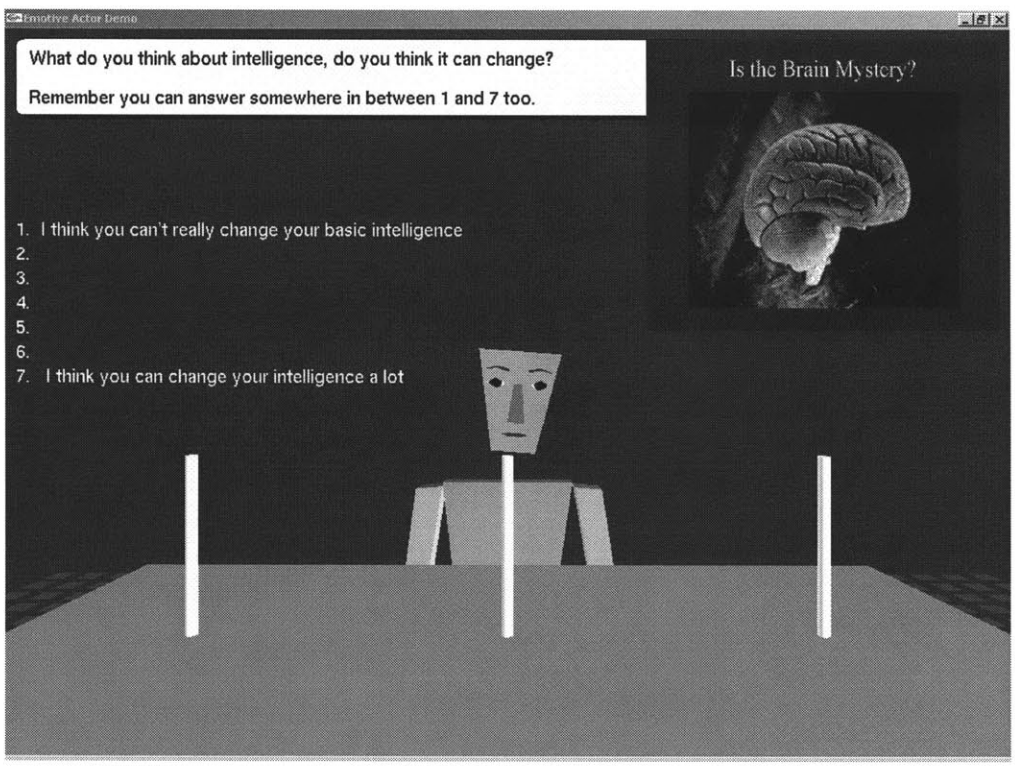
Introductory question (reverse coded)



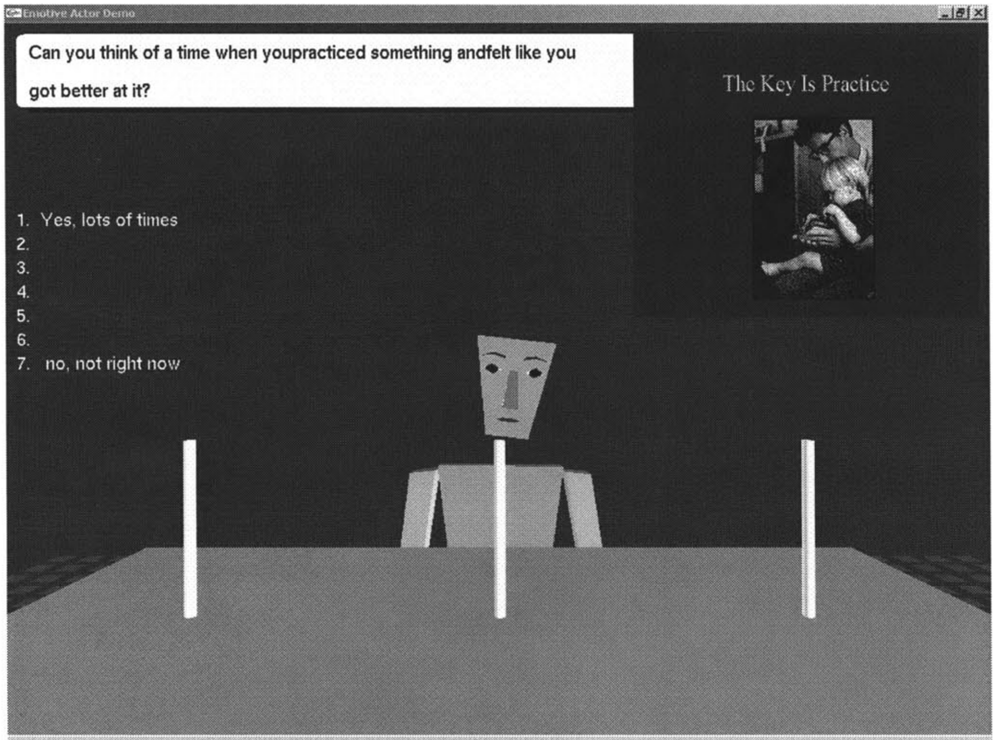
Introductory question



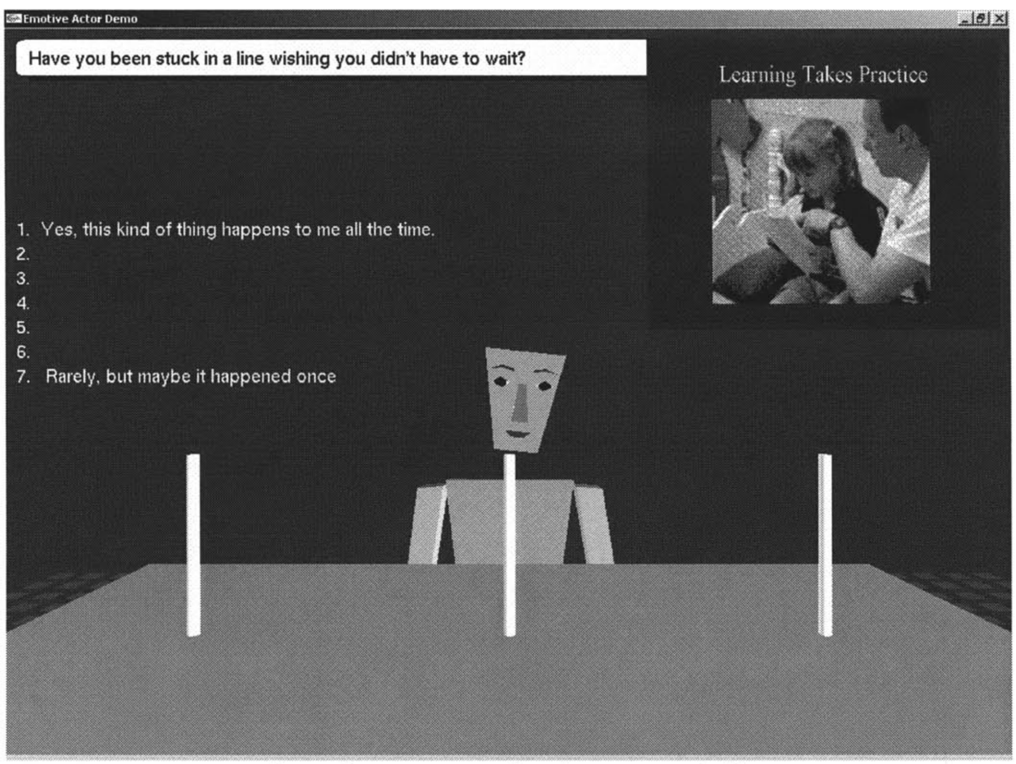
Introductory slide show



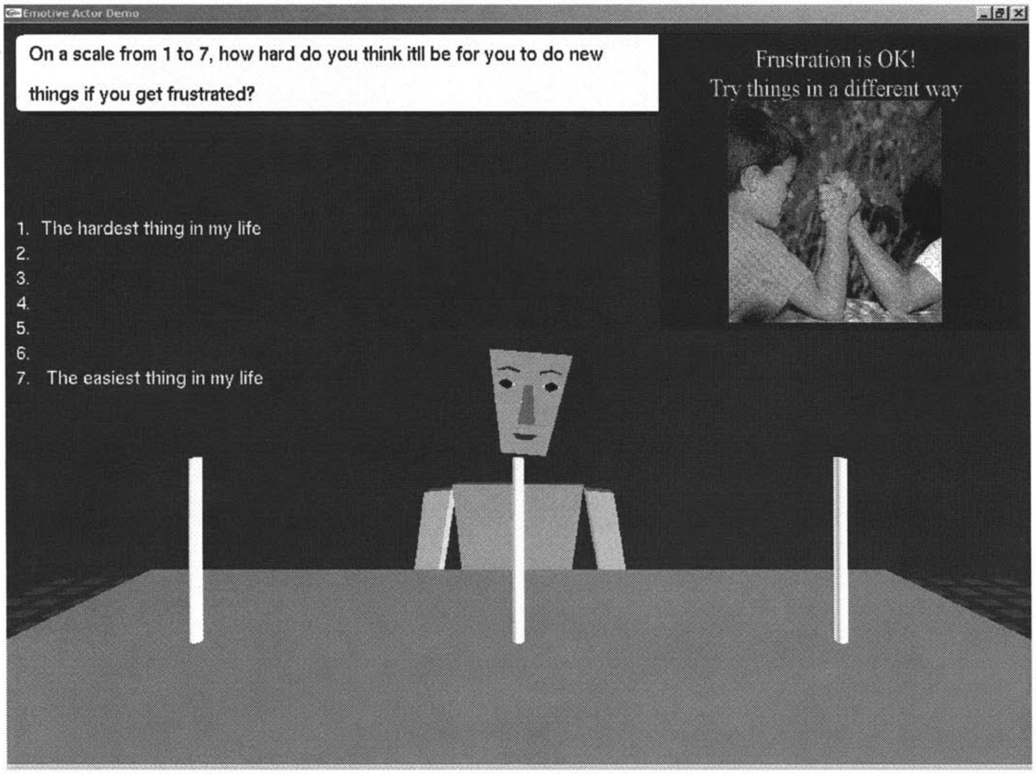
Introductory question



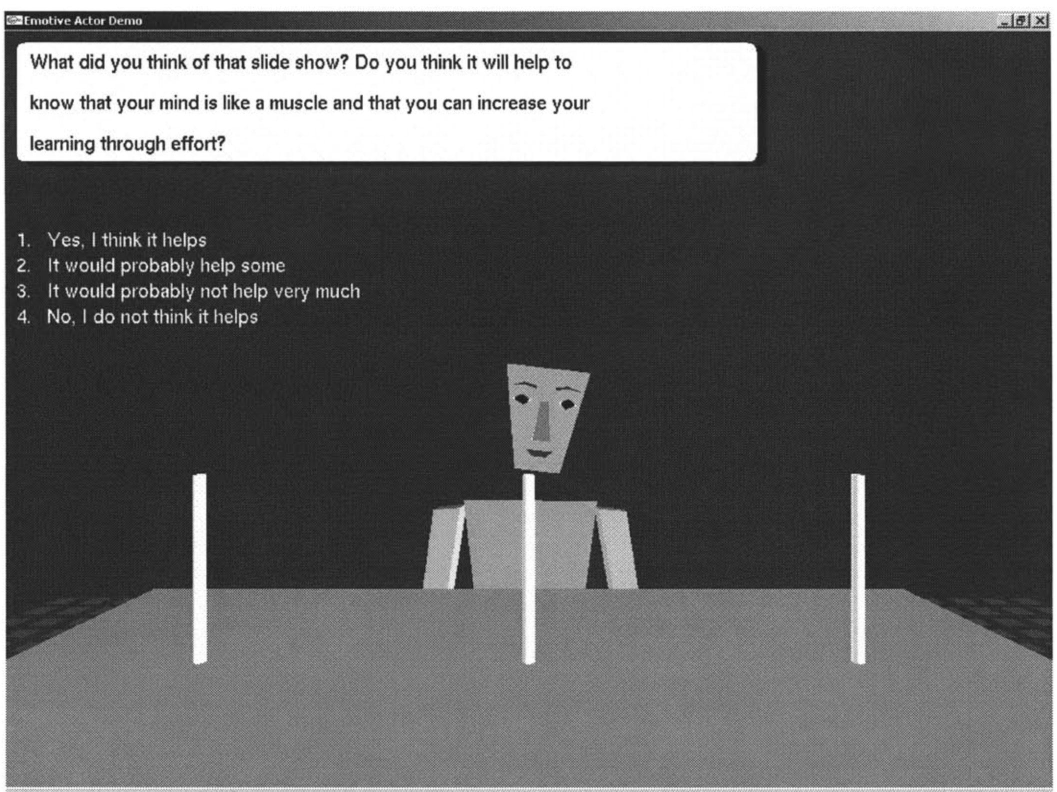
Introductory question (reverse coded)



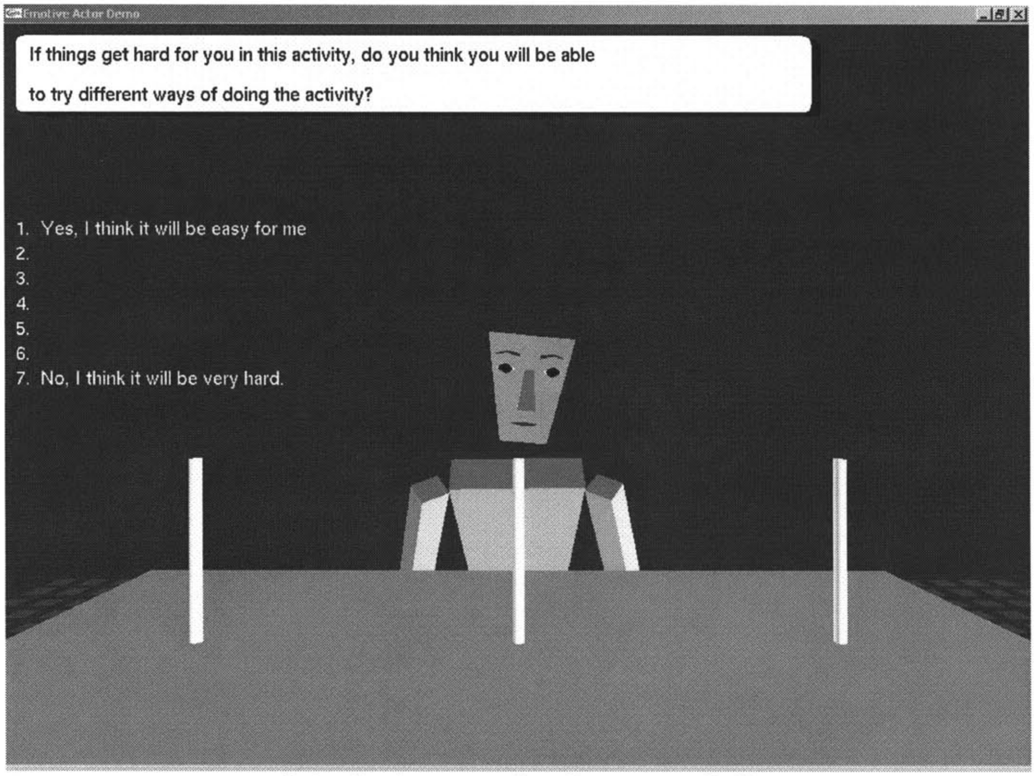
Introductory question (reverse coded)



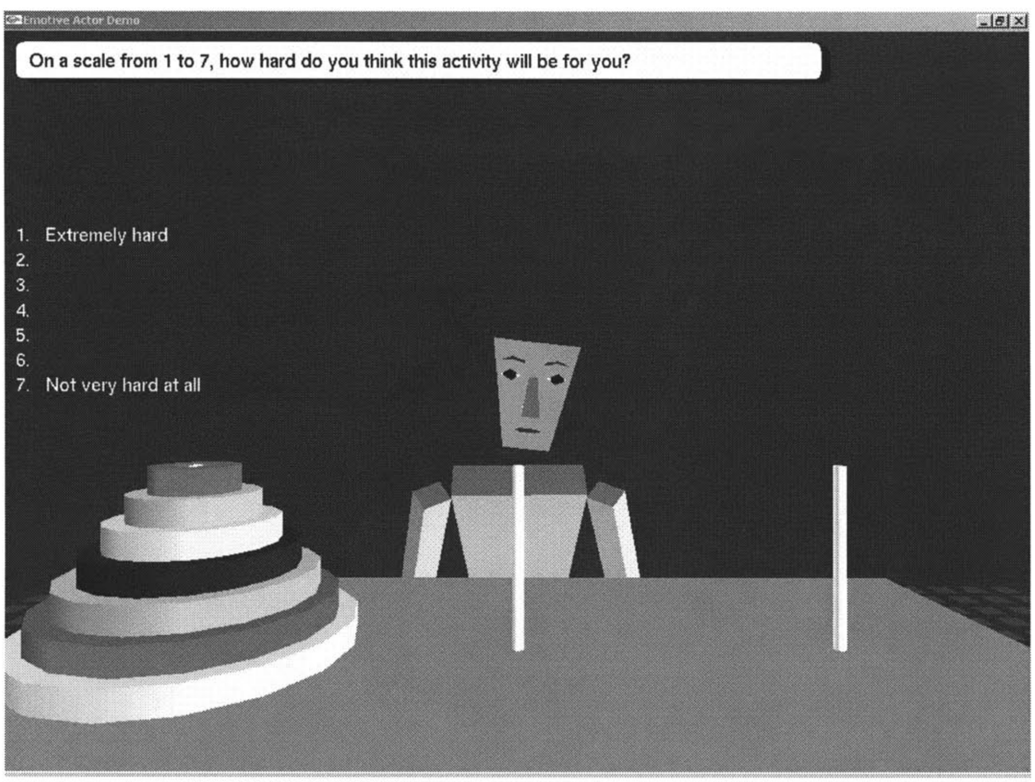
Introductory question (reverse coded)



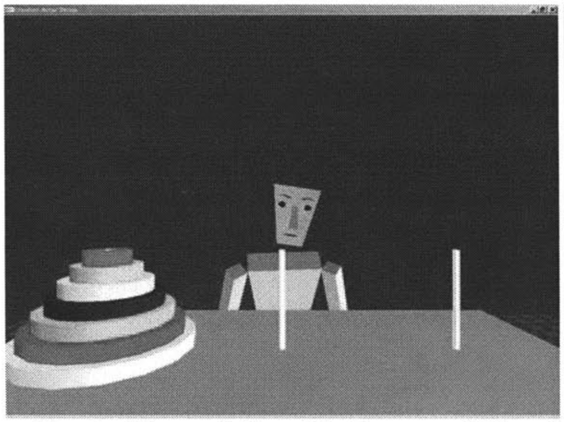
Introductory question (reverse coded)



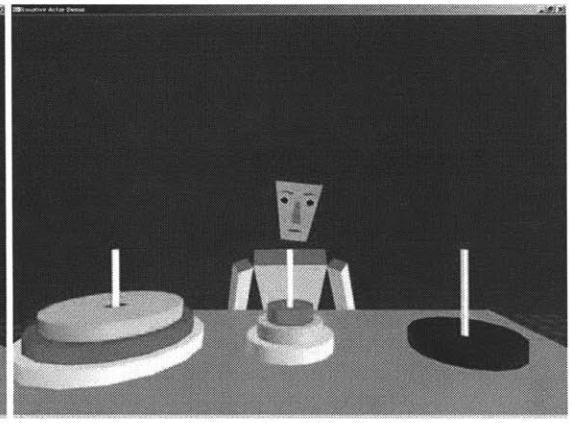
Introductory question (reverse coded)



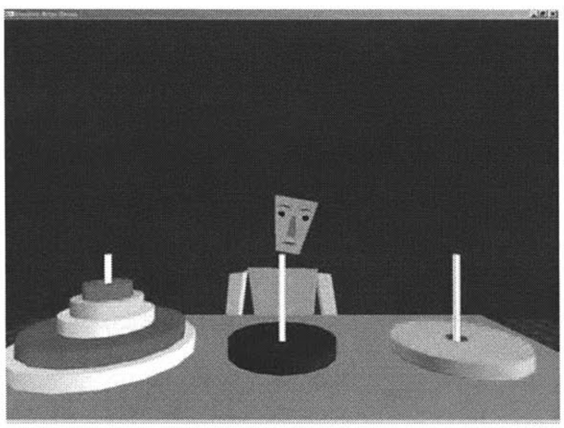
Introductory question (reverse coded)



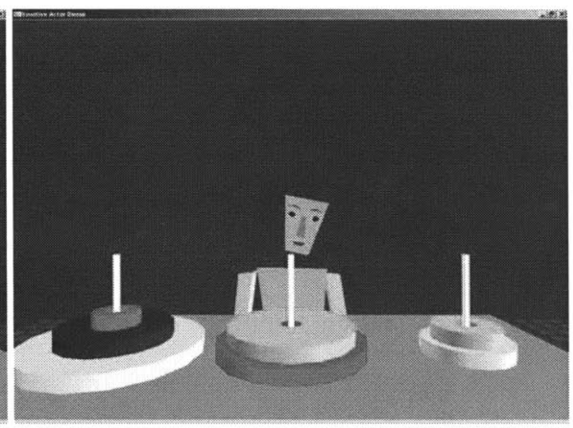
Start of the activity with Towers of Hanoi in the initial configuration



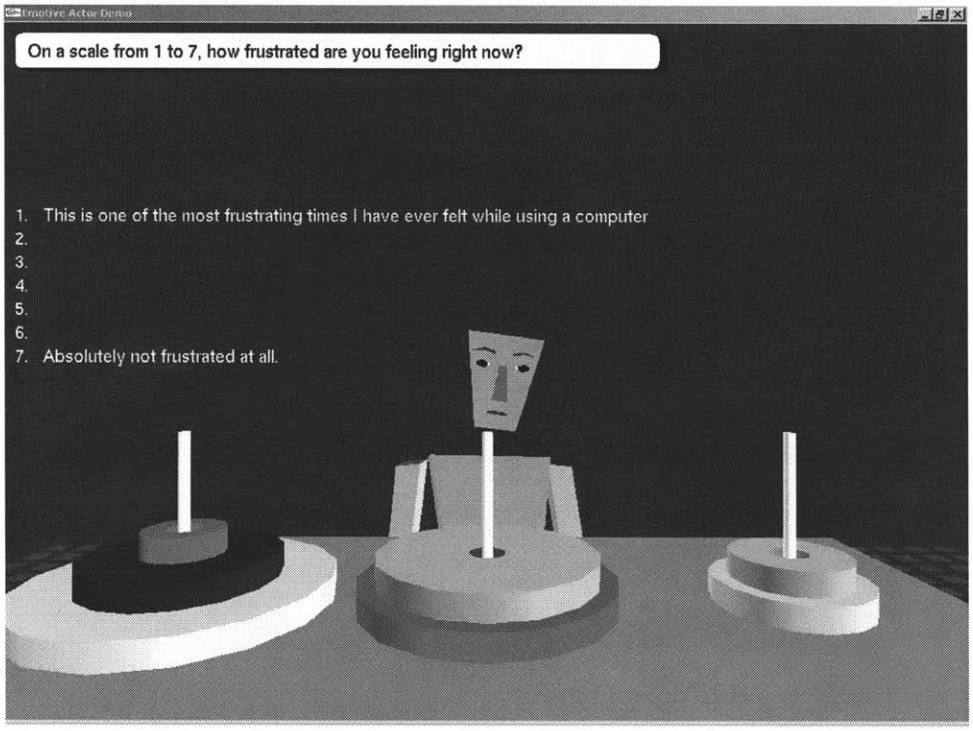
Towers of Hanoi with partial progress



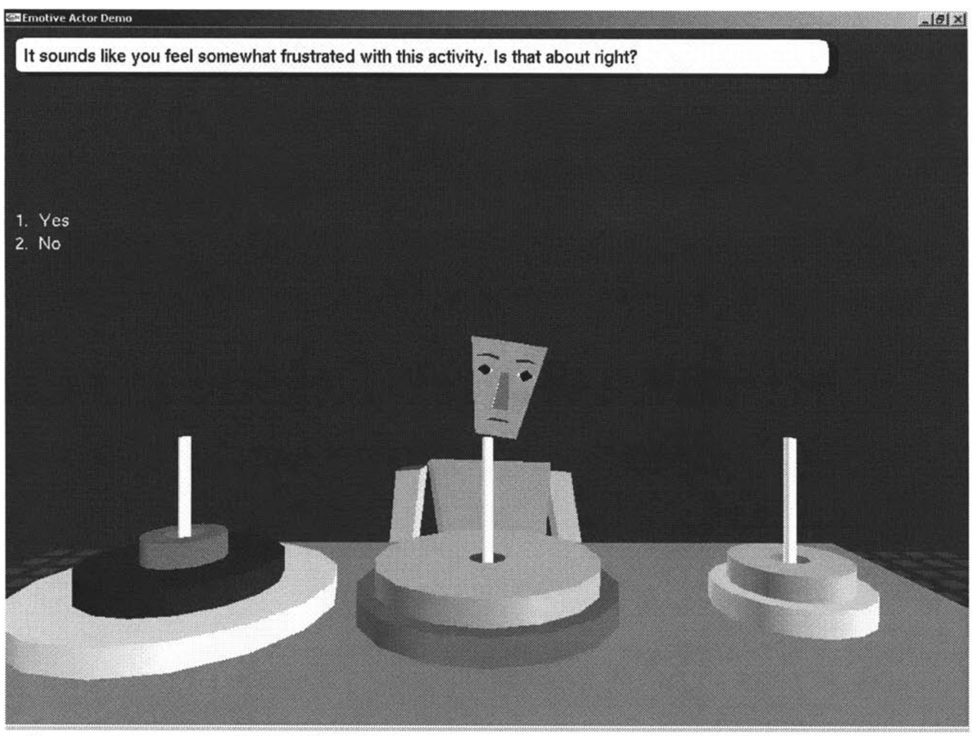
Towers of Hanoi with partial progress



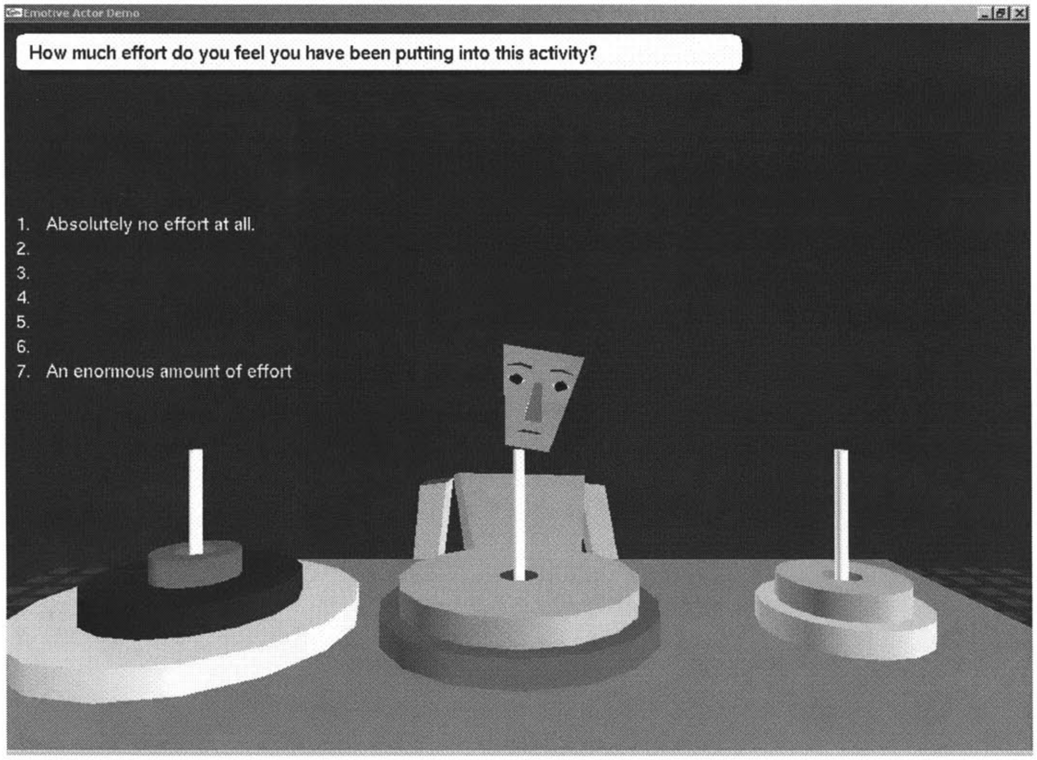
Towers of Hanoi with partial progress



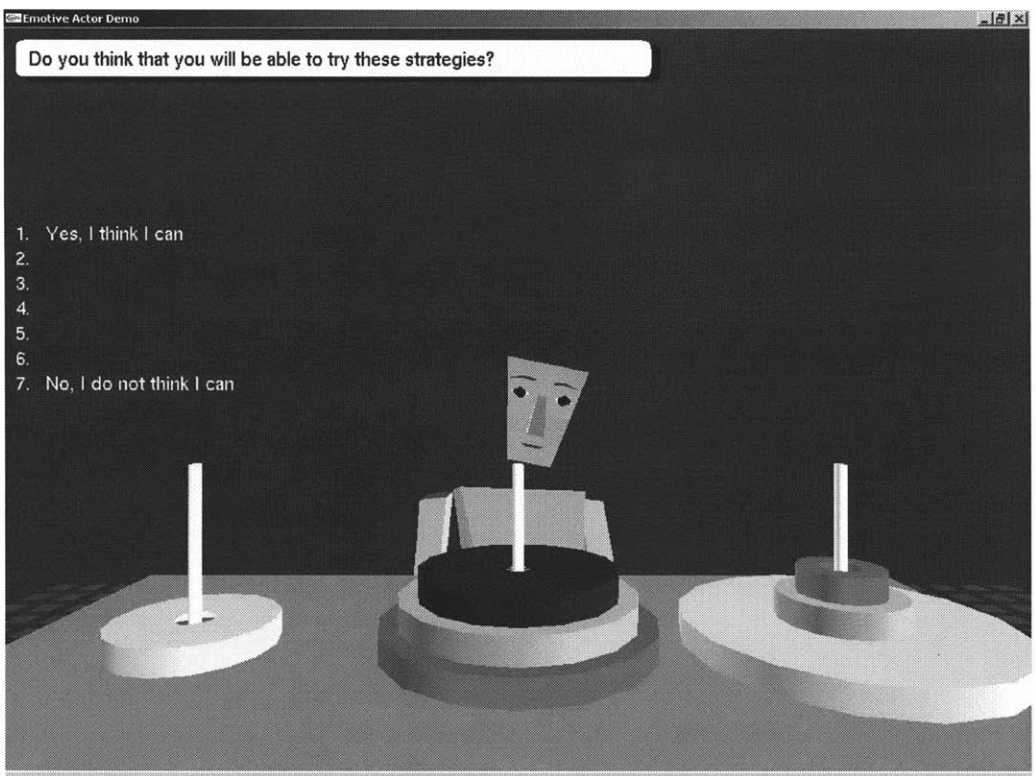
Intervention question (reverse coded)



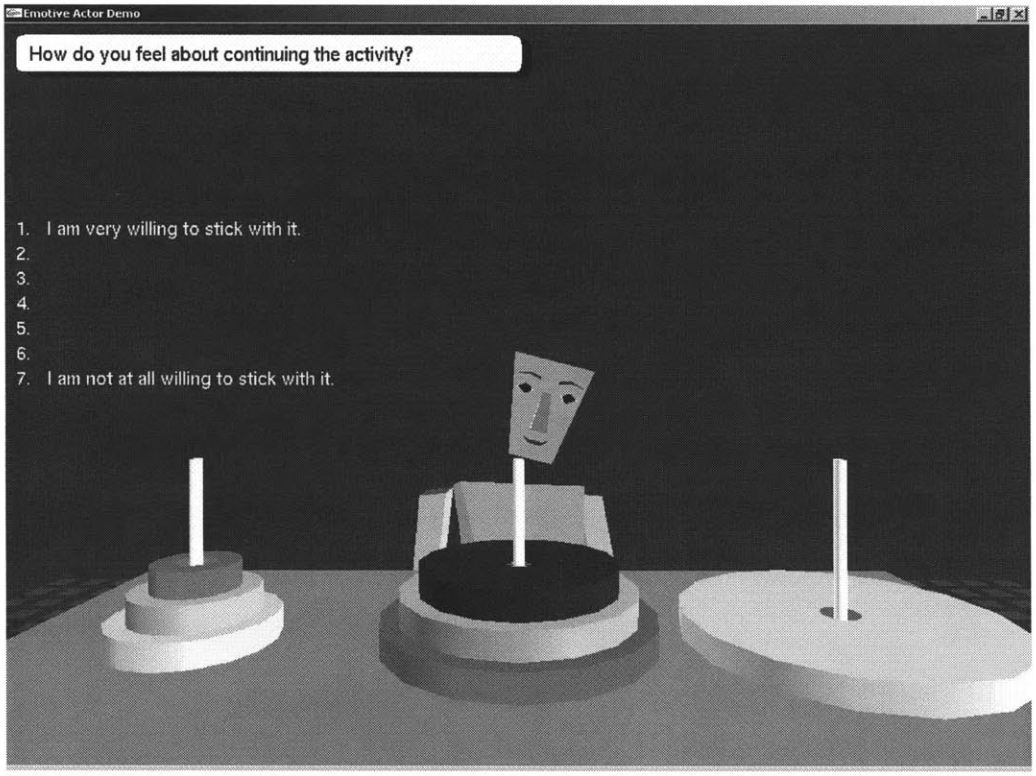
Affective support intervention checking to see if its interpretation of the participant's response to the question about their current frustration is "about right." If no is answered the character says, "I'm sorry," and repeats the previous frustration question. In terms of screen shots this is the only difference between the two interventions. Other differences manifest themselves in terms of subtle character expression and differing voice dialogue.



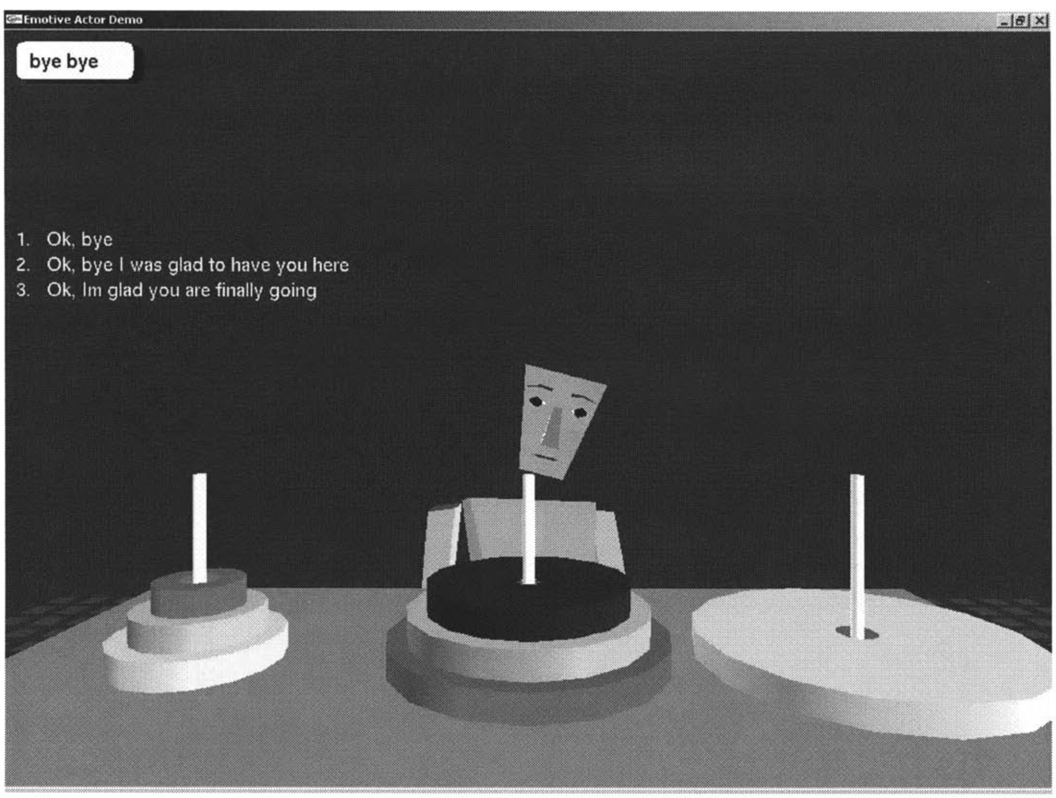
Intervention question



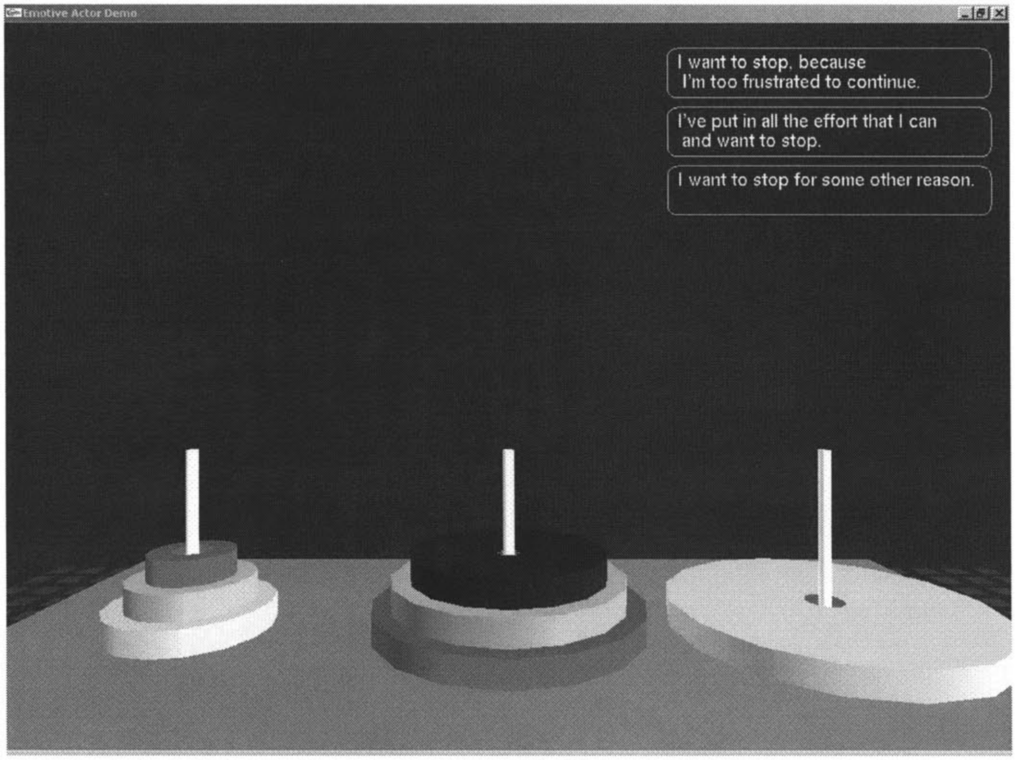
Intervention question (reverse coded)



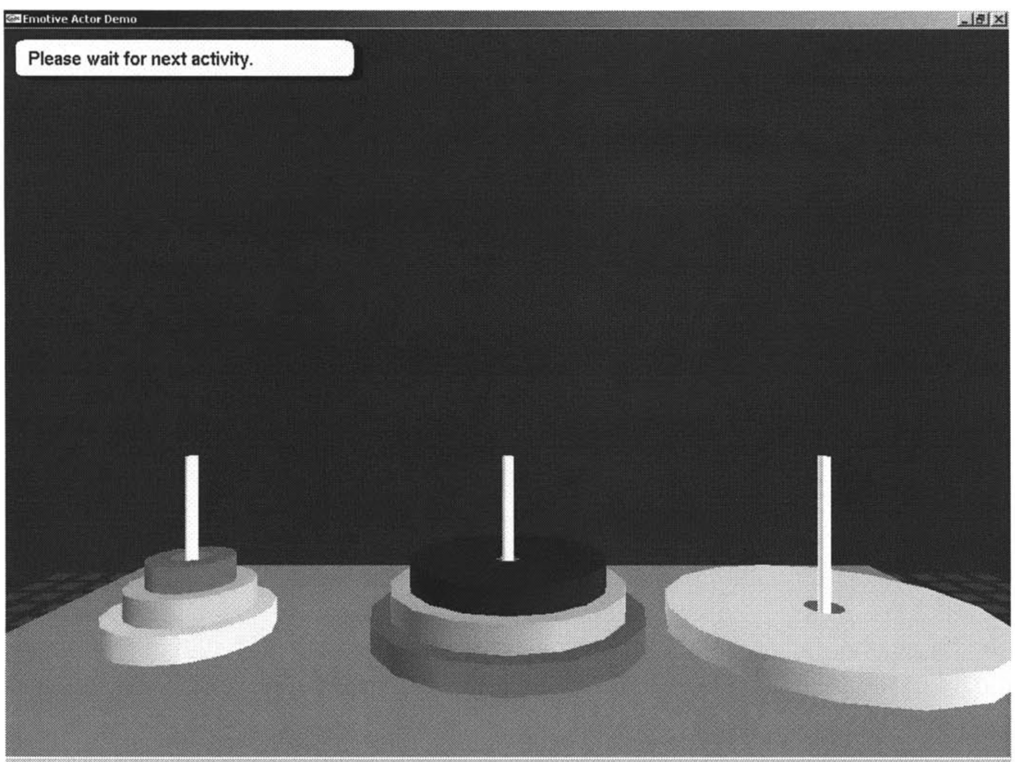
Intervention question (reverse coded)



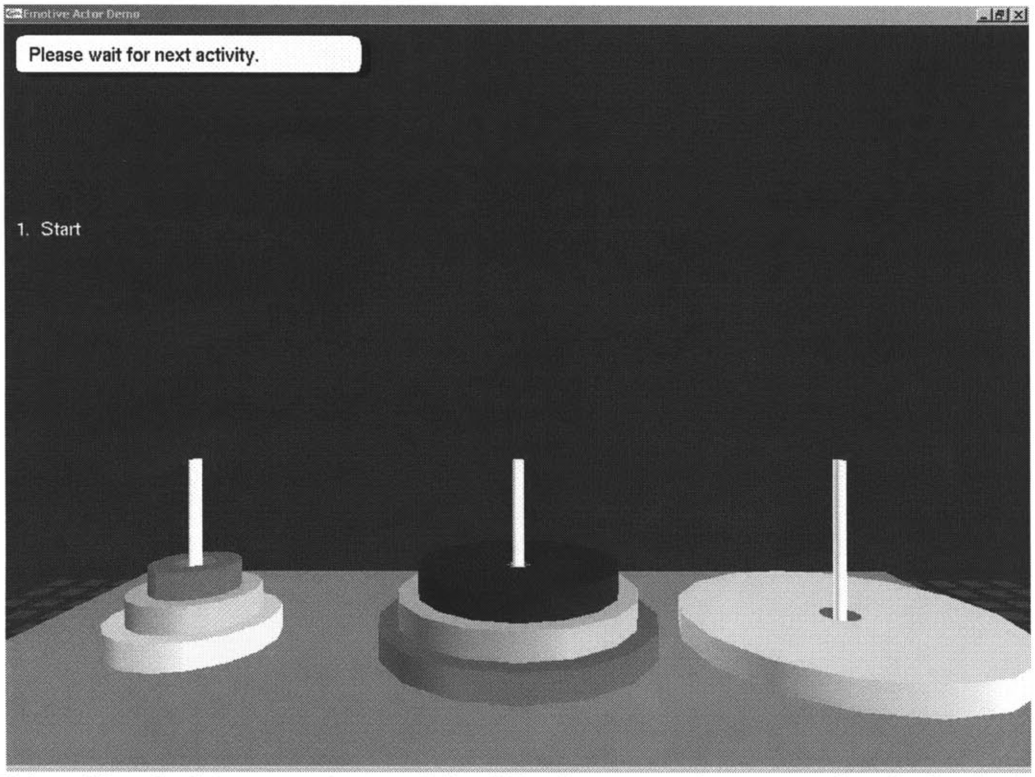
Intervention bye.button response screen, the presentation order is randomized



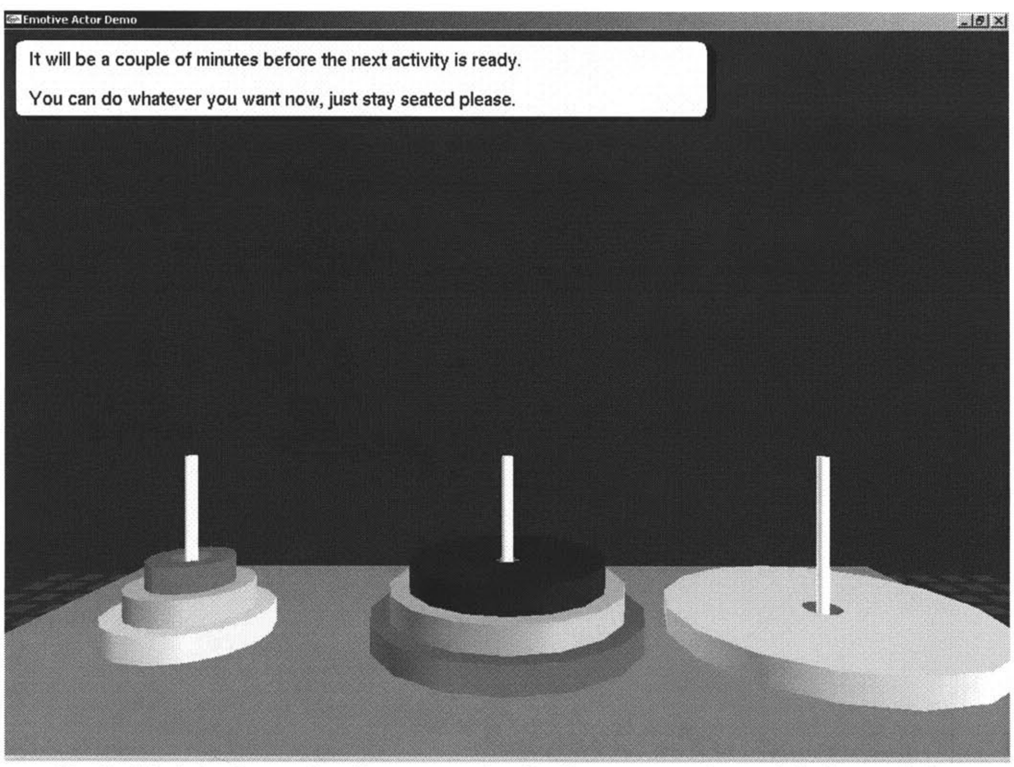
Post intervention persistence with three quit buttons present



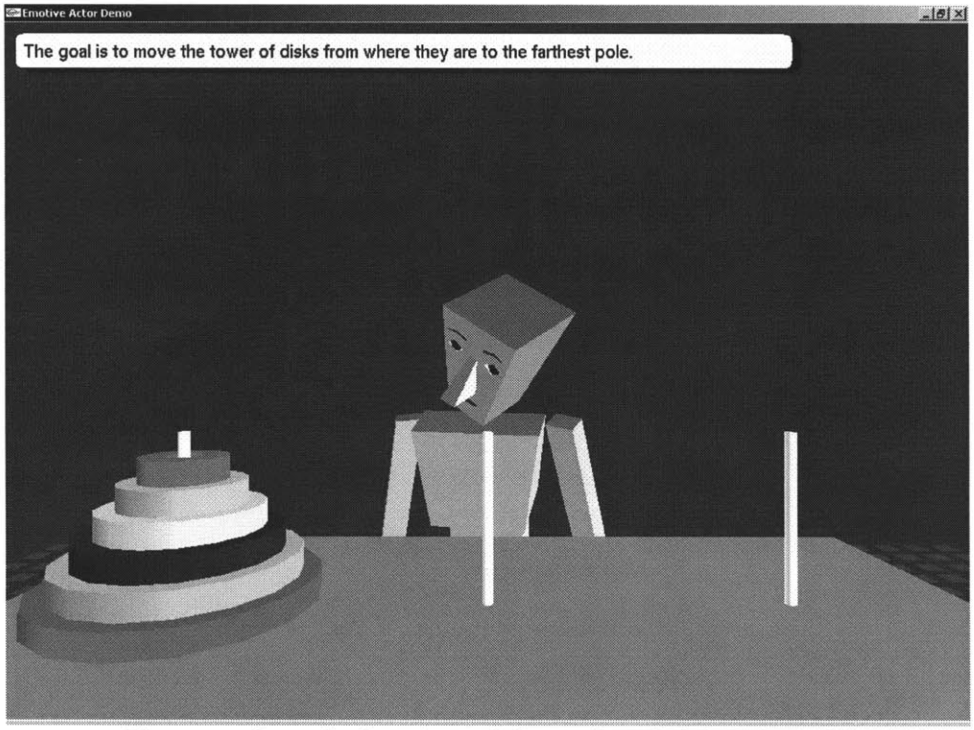
"Please wait for the next activity" appears on the screen 15 minute after the start of the activity, if participants have not already pressed one of the three quit buttons. The experimenter then asks the participant to answer the post-activity questions.



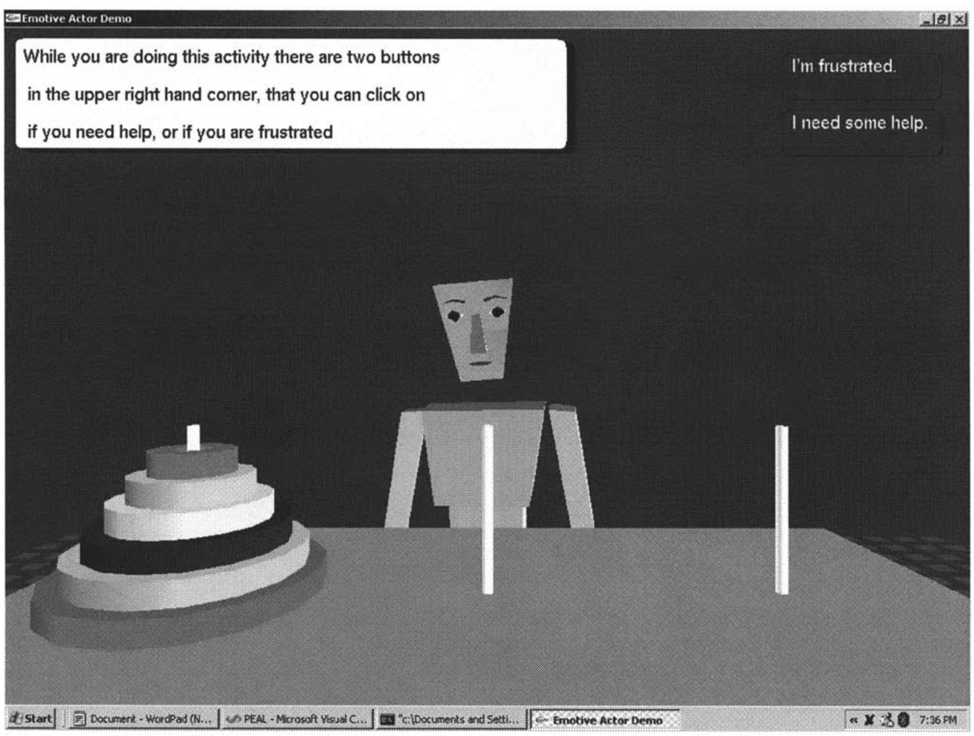
Participant is asked to press start, and the following screen is presented at the beginning of the intrinsic motivation opportunity to assess reengagement with the activity



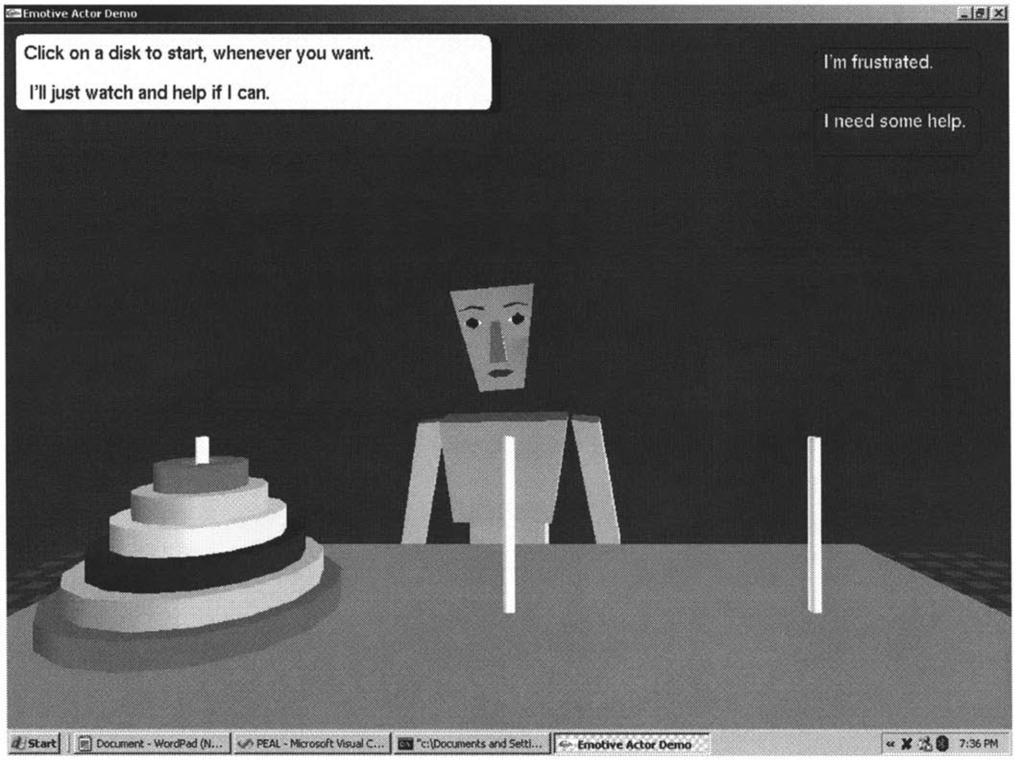
End of intrinsic motivation opportunity for reengaging with the activity
The following screen shots are from the Third Pilot Study.



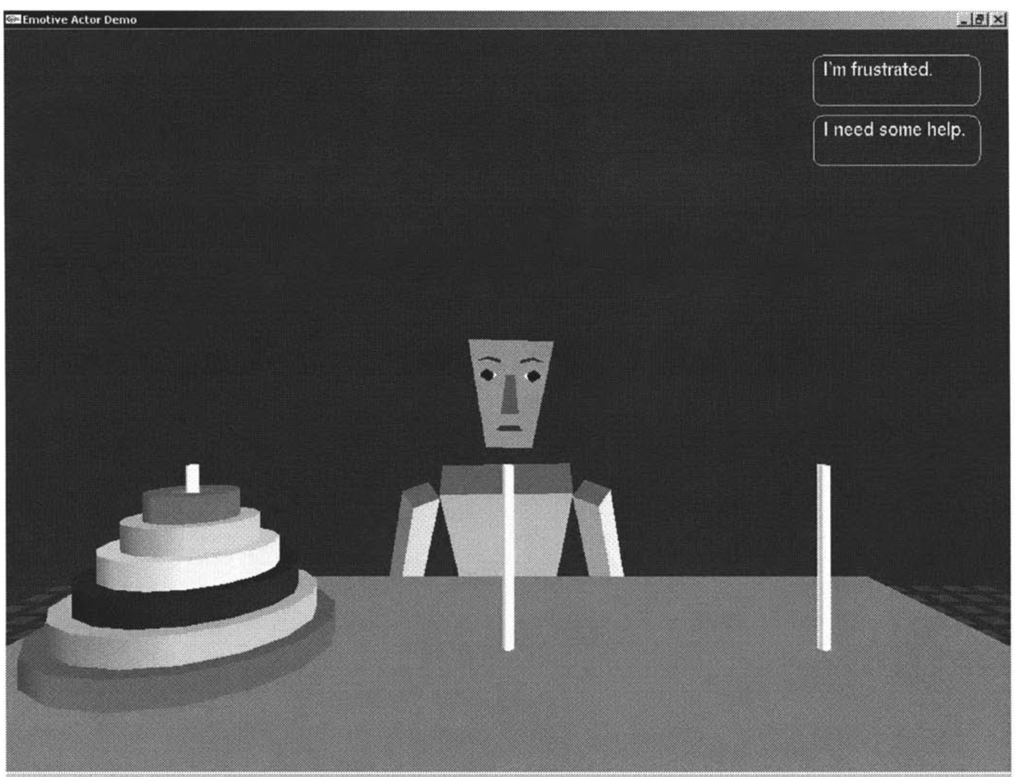
Character introducing the activity and explaining the rules



Character introducing the two quit/help buttons



Character suggesting that participants start the activity.



The character is present, along with two quit/help buttons, throughout the activity.

Character Dialogue in the Introduction:

Hi there. My name is Casey. I'm a digital character. The people who created made this game. I have not been able to try it yet because they forgot to program me to be able to move the disks. I can't even move my arms. However, they did tell me the rules. Do you know the rules? Have you seen this game before?

Participant can click on one of these options:
1 Yes, I know the game well
2 I think I've seen it, but do not really remember the rules
3 No, can you tell me how to play?"

Ok then, the rules are really pretty simple. The goal is to move the tower of disks from where they are to the farthest pole. The only rule is that a big disk can not go on top of a smaller one.

Oh yeah, I almost forgot, before you start, they wanted me to show you this slide show about how you can use this activity to increase your intelligence. It lasts about four minutes so please pay close attention.

There will be some questions for you too. When you answer them you can click on 1 or 7 if these represent what you think, or you can click on 2, 3, 4, 5, or 6 even though they don't have words. You should click on one of these if your answer is somewhere in between, if it's right in the middle, or a bit closer to 1 or 7 but not completely there. Go ahead, try clicking on one of the numbers that doesn't have text to see that they work. Which one do you want to click on, 2, 3, 4, 5, or 6.

Participant can click on one of these options:
1 I want to click on one
.
.
7 I want to click on seven

If the participant clicks on 1 or 7 the character says:
Thanks. Remember you can hit the other numbers too.

Otherwise the character says:
Thanks! You wanted to click 5. **(or 2, 3, 4, or 6)**

Then the character starts the slide show:
Many people think of the brain as a mystery. They don't know much about what it does or how it works. When they do think about what intelligence is, many people believe that a person is either smart, average, or dumb, that they are born smart or just slow.

What do you think about intelligence, do you think it can change? Remember you can answer somewhere in between 1 and 7 too.

Participant can click on one of these options:
1 I think you can't really change your basic intelligence

.

.

7 I think you can change your intelligence a lot

New research shows that the brain is more like a muscle than many people think. It changes and gets stronger when you use it. When you use your brain in different ways it gets better at handling new challenges. And scientists have been able to show just how the brain grows and gets stronger when you learn.

Everyone knows that when you lift weights, your muscles get bigger and you get stronger. A person who can't lift 20 pounds when they start exercising can get strong enough to lift more than 100 pounds after working out for a long time. That's because their muscles become larger and stronger with exercise. And when you stop exercising, the muscles shrink and you get weaker. That's why they say, "use it or lose it."

But many people don't know that when they learn and practice new things, parts of their brain change and get larger.

When you learn new things, tiny connections in the brain actually multiply and get stronger. Then, things that you once found very hard or even impossible to do, like speaking a foreign language or doing algebra seem to become easy. The result is your brain can get stronger with exercise.

Participant can click on one of these options:
Can you think of a time when you practiced something and felt like you got better at it?
1 Yes, lots of times

.

.

7 no, not right now

Scientists have shown practice can make the brain stronger. Inside the brain are billions of tiny nerve cells called neurons. The nerve cells have branches connecting them to other cells. The cells send messages to each other in the form of chemicals. This communication between brain cells is what allows us to think and solve problems.

Scientists have found that the key to growing the brain is to keep trying to learn. When you try new activities you have the opportunity to learn. Sometimes these activities may feel very hard. You may even get stuck, and you may feel very frustrated. When you feel frustrated or stuck, it might help to try the activity in a different way.

Let's try to think of an example. Have you ever been late for something that really mattered to you, and found yourself frustrated that you were stuck somewhere that you didn't want to be? Maybe you were stuck in traffic or lost because you didn't have directions with you. Have you been stuck in a line, wishing you didn't have to wait?

Participant can click on one of these options:
1 Yes, this kind of thing happens to me all the time.

.

.

7 Rarely, but maybe it happened once

Well, most people get stuck at some point. At that moment, when you are stuck, the frustration you felt might have made you think of different things you could have tried: perhaps you could have checked ahead about the traffic, or called first for directions, or done other things. Frustration is a signal that it might be time to try to think of how to do something differently than the way you were just trying. It is not a signal that you should quit, or that you aren't good at something. People who are very good at things are people who have learned to deal with frustration well. They know that being stuck is ok; what matters is to keep trying new things when the thing you're doing doesn't work.

On a scale from 1 to 7, how hard do you think it'll be for you to do new things if you get frustrated?

Participant can click on one of these options:
1 The hardest thing in my life

.

.

7 The easiest thing in my life

"Well it is always good practice and well worth trying, I know you can do it."

If it helps, keep in mind that babies, who are learning at an incredibly fast rate, are trying things in different ways all the time. From the first day they are born, babies are hearing people around them talk all day, every day, to the baby and to each other. In a way, they get exercise in listening. Later, when they want to tell their parents what they want, they start practicing talking themselves. At first, they just make goo-goo sounds.
Babies also get very frustrated when learning, but they do not quit trying. They keep trying new things. It is frustrating for them at first but since they keep trying they get better and better at it. Then, words start coming. And by the time they are three years old, most can say whole sentences almost perfectly.

Once children learn language, they don't forget it. The child's brain has changed; it has actually gotten stronger. This can happen because learning causes the babies brain cells to get larger and grow new connections between them. These new, stronger connections make a child's brain stronger and smarter, just like weightlifters' big muscles make them strong. No one thinks that babies are stupid because they cannot talk. They just haven't learned how to yet. Similarly, if somebody can't solve math problems, or spell a word right, or read fast, they are not dumb, they simply need more practice, and each brain may need different kinds of practice. Remember if you feel that something is frustrating, or if you get stuck, then take a breath and think about how you can try it in a different way. Then go ahead and see what you learn by doing it that way. In the process you will probably learn something.

The student everyone thinks is the smartest may not have been born any different from anyone else. But before they started school, they may have already started learning to read. Then, in the classroom, everyone said, that's the smartest student in the class, because they could read more than the other students. They just don't realize that any of the other students could learn to do the same if they exercised and practiced as much.

What can you do to get smarter? Just like a weightlifter or basketball player, you have to exercise and practice. By practicing, and doing activities in different ways, especially when you get frustrated, you can make your brain stronger and better at new challenges. You also learn skills that let you use your brain in a smarter way, just like a basketball player leans new moves. But many people miss out on the chance to grow a stronger brain because they think they cannot do it, or that it's too hard or frustrating. It does take work, just like becoming stronger physically or becoming a better ball player does. Sometimes it even hurts. But when you feel yourself get better and stronger, all the work is worth it.

The slides go away and the character asks:

What did you think of that slide show? Do you think it will help to know that your mind is like a muscle and that you can increase your learning through effort?

Participant can click on one of these options:
1 Yes, I think it helps
2 It would probably help some
3 It would probably not help very much
4 No, I do not think it helps

Well, I think you will get a chance to see in a moment.

If things get hard for you in this activity, do you think you will be able to try different ways of doing the activity?

Participant can click on one of these options:
1 Yes, I think it will be easy for me
.
.
7 No, I think it will be very hard

Well if you give it your full effort, you will be practicing your learning skills and increasing your intelligence.

Some people express themselves and their feelings by squeezing something, like the mouse. Others move around a lot in their chair. Other people move their head a lot. I hope you will be comfortable expressing any feelings you have during this activity in as natural a way as possible. Anyways, I think you are ready to start. Here are the disks, remember you can only move one at a time and a big one can not go on top of a smaller one.

And the goal is to move these disks to the far pole.

The people who created me sometimes call me away, so I may have to leave while you are doing this activity. If this happens, I hope you won't mind. I would much rather stay.

On a scale from 1 to 7, how hard do you think this activity will be for you?
Participant can click on one of these options:
1 Yes, I think it will be easy for me
.
.
7 No, I think it will be very hard

Well give it your best shot. Click on a disk to start, whenever you want. I'll just watch and help if I can.

Character Dialogue in Affect Support Intervention:

I'm sorry I don't know more about this activity so I could help you through it. I do know that many people find it frustrating. On a scale from 1 to 7, how frustrated are you feeling right now?

1. This is one of the most frustrating times I have ever felt while using a computer
2.
3.
4.
5.
6.
7. Absolutely not frustrated at all.

Participants receive one of the following 4 responses in accordance to their level of frustration.

(1) It sounds like you are not very frustrated with this activity. Is that about right?
(2-4) It sounds like you feel somewhat frustrated with this activity. Is that about right?
(5-6) It sounds like you are very frustrated with this activity. Is that about right?
(7) It sounds like you are extremely frustrated with this activity. Is that about right?

Participants can answer:

Yes
No

If they answer "No", then the character says:

Sorry about that, to clarify, how frustrated are you?
Participants can select from the same options as before, see above, and they recieve the same 4 responses.

The character says the following in accordance to their answer:

(1) Good, I'm glad this activity is not making you feel that way.

(2-5) I'm sorry you feel that way, I know it can be hard to have to do things that make you feel frustrated.

(6-7) Wow, that must be really tough, I'm really sorry doing this activity is making you feel that way.

Then the character asks:

On a scale from 1 to 7, how much effort do you feel you have been putting into this activity?

1. Absolutely no effort at all

.

.

7. An enormous amount of effort

The character says the following in accordance to their answers to the frustration and effort questions (effort is reverse coded so that 1 is low and 7 is high effort). The (frustration)(effort) numbers correspond to the levels of frustration and effort indicated by the participant:

(1)(1) Hummm.... I wonder if you are putting enough effort into this. It is great that you are AWARE of how you feel. Remember, if it does get frustrating it sometimes helps to try things differently. Sometimes it might be good to "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

(1)(2-5) It sounds like this is a good challenge for you. Just remember that your effort does help you learn. It is great that you are AWARE of how you feel. Remember, if it does get frustrating it sometimes helps to try things differently. Sometimes it might be good to "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

(1)(6-7) It sounds like this is a good challenge for you. Just remember that all of your effort does help you learn. It is great that you are AWARE of how you feel. Remember, if it does get frustrating it sometimes helps to try things differently. Sometimes it might be good to "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

(2-5)(1) It is probably aggravating to have to stick with this activity when you're finding it somewhat frustrating. Please remember it is ok to be frustrated. It is great that you are aware of how you feel. Remember, frustration sometimes tells you to try things differently. It is like a navigation sign that says "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying as hard as you can, when it gets challenging, you will get better and smarter.

(2-5)(2-5) It is probably aggravating to have to stick with this activity when you're already putting in effort and finding it somewhat frustrating. Please remember it is ok to be frustrated. It is great that you are AWARE of how you feel. Remember, frustration sometimes tells you to try things differently. It is like a navigation sign that says "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

(2-5)(6-7) It is probably aggravating to have to stick with this activity when you're already putting in a lot of effort and finding it somewhat frustrating. Please remember it is ok to be frustrated. It is great that you are AWARE of how you feel. Remember, frustration sometimes tells you to try things differently. It is like a navigation sign that says "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

(6-7)(1) It is probably really aggravating to have to stick with this activity when you're finding it so frustrating. Please remember that it is ok to be frustrated. It is great that you are aware of how you feel. Remember, frustration sometimes tells you to try things differently. It is like a navigation sign that says "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

(6-7)(2-5) It is probably really aggravating to have to stick with this activity when you're already putting in effort and finding it so frustrating. Please remember that it is ok to be frustrated. It is great that you are AWARE of how you feel. Remember, frustration sometimes tells you to try things differently. It is like a navigation sign that says "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

(6-7)(6-7) It is probably really aggravating to have to stick with this activity when you're already putting in a lot of effort and finding it so frustrating. Please remember that it is ok to be frustrated. It is great that you are AWARE of how you feel. Remember, frustration sometimes tells you to try things differently. It is like a navigation sign that says, "slow down as you might want to change direction." Take a breath and be determined to keep thinking of different ways to solve the problem. You are creative and there are always many things you can try. Maybe one of them will work. Remember, the mind is like a muscle that when exercised may not feel good, but it is getting stronger through exercise. If you stick with it and keep trying hard, you will get better and smarter.

The character then asks:
Do you think that you will be able to use these strategies?
1. Yes, I think I can
.
.
7 No, I do not think I can

The character says the following in accordance to their answer:
Great, good luck!
It can be hard, but remember that's a sign that you are learning, stick with it and you will learn a lot.

I have to go now. Thank you for letting me watch you do this activity. Watching you has helped me learn too. Sorry that I have to leave now. How do you feel about continuing the activity?
1. I am very willing to stick with it.
.
.
7. I am not at all willing to stick with it.

The character says the following in accordance to their answer:
(1-2) Great, good luck. Please try as hard as you can. If you feel like you would like to stop there will be a few buttons in the upper right hand corner that you can press. Bye Bye.

(3-5) Well try as hard as you can and try to use these strategies. If you feel like you need to stop, there are a few buttons in the upper right hand corner that you can press.

(6-7) I'm sorry that I have to ask you to continue anyway. Please just try as hard as you can. If you feel like you would like to stop there will be a few buttons in the upper right hand corner that you can press. Bye Bye.

Appendix D. Affective Agent Research Platform System Description

By John Rebula

Introduction

The Affective Learning Companion system is divided into several constituent components: System Server, Sensor outputs, Sensor Monitor, Processors and DataAnalyiser (refered to in the thesis as the Behavior Engine), Character Monitor, Character/Game Engine, Data Logger.

The system server (SS) is the nexus for decision-making and message passing in the overall system. The SS instantiates objects known as monitors that serve as interfaces to different input/output sources in the character system. Because the possible input sources are sensors with hardware components that output data in a format unique to each device, unique processors are also needed to handle the information received from the sensors. It also has a GameManager class which keeps track of the current game. Then, the SS instantiates objects known as processors to analyze data. Appropriate instructions are delivered by the processors through the character monitor to the Character/Game Engine (CGE). The Data Logger (DL) is responsible for logging all the information moving through the whole system.

The establishment of this modularity was partly our decision, and partly determined by external factors. For instance, the CGE had to be a separate module because it was a third-party application with predefined interfaces. The separation of SS, DL, monitors, and processors was due to clear functional differences: The DL was responsible for keeping logs of packet receipt from the sensors. The monitors were responsible for gathering data from different components in the system. The processors were responsible for analyzing collected information, and the SS was responsible for coordinating the actions of every component in the system.

Another consideration in deciding modular boundaries was performance. Our basic idea was to have each of these components running on separate computers, so as to avoid overload and unnecessary slowdown of any particular machine. However, we refrained from further subdividing these units because eventually, the cost of communicating over the network would outweigh the performance benefit of subdivided machines.

System Server

Role & Responsibilities

The SS is the central component of the entire system, and has many modules. Its major tasks are:

Start TCP Server
Begin Video Monitoring
Instantiating monitors to handle data collection and delivery

Instantiating processes to handle data analysis

The SS collects information through the use of monitors. These monitors will read information from their respective buffers and send the information in the form of a TCP packet to the SS. Many different monitors, (CharacterMonitor, SensorMonitor, Playbackmonitor, ...) are created and run in their own thread. These monitors will send packets to the SS, containing information from either the sensors or the CGE. The SS upon receiving the message will dispatch the appropriate processor to analyze the data. Each processor, after analysis is complete, has a handle to the SS's CharacterMonitor, and will use it to deliver appropriate instructions to the CGE.

TCP Server

The SS receives TCP packets from the monitors as data is collected. The former's constructor sets up the TCP server on the local machine for this purpose. Users starting the system from the command line can specify a port number as the first argument; this will then be used in the SystemServer class constructor.

Video Recording

The first task for the SystemServer class constructor is to initialize video recording of the user attempting the task. On another machine, there is a recording program which upon receiving a specially formatted datagram, sets the connected video camera to 'record' mode and begins storing the video. The following code fragment in the SystemServer class illustrates this process:

```
DatagramSender udpSender = new DatagramSender(20000);
udpSender.sendPacket("test", videoMachineIP, videoMachinePortNum);
```

The exact contents of this packet are inconsequential as long as it is sent to the correct port.

Monitors

The following code shows part of the SystemServer class constructor:

```
this.charMonitor = new CharacterMonitor(this, servSocket, studentNum);
this.characterThread = new Thread(charMonitor);
...
this.sensorMonitor = new SensorMonitor(this, servSocket, studntNum);
this.sensorThread = new Thread(sensorMonitor);
...
gsrMonitor = new GSRPostProcessingMonitor(this);
gsrThread = new Thread(gsrMonitor);
headMonitor = new HeadPostProcessingMonitor(this);
headThread = new Thread(headMonitor);
...
sensorThread.start();
characterThread.start();
```

```
gsrThread.start();
headThread.start();
```

As you can see, the SystemServer class instantiates monitors and runs each monitor in its own thread. The monitors are used by the SS to gather information from the sensors, and will be more fully discussed in section *Monitors*.

Game Manager

At this point, the SystemServer class also creates a new GameMananger class, via the following code:

```
GameManager.newGame(true);
GameManager.resetGame();
```

The GameManager class keeps track of the current state of the Towers of Hanoi; how many moves are left, and whether a move is legal or not.

Data Analyzer

The data analyzer is the component of the SS which is in charge of extracting and processing the information from received TCP packets. The data analyzer is created and initialized in the SystemServer class constructor. Upon initialization, the data analyzer creates numerous processors. These processors are in charge of processing the gathered information, and will be more fully discussed in section *Processors*.

Sensor outputs

Roles and Responsibilities

Receiving information from user
Delivering information to system server

There are four sensors in the system, the BlueEyes smile sensor, the mouse pressure sensor, the GSR skin conductivity sensor, and the chair pressure sensor. Each of these sensors is responsible for sending data to the SS. The following is the Python code that collects and dispatches the mouse pressure data:

```
def run():
    serial.Serial(0).close()
    port = serial.Serial(0) #ACTUALLY using COM4 (serial.Serial is off)
    #print "Opened COM 1"
    fileout = open("./PSMdata.txt", 'w') # create and clear file
    host = "18.85.1.94"
    s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
    s.bind(("", 0))
    running_index = 0
    while 1:
        ...
```

```
line = port.readline()
...
client_port = 4950
s.sendto((line), (host, client_port))
```

The other sensors behave similarly. Each initializes the server and the port, and prepares a temporary file to record data to. The file is then read, line by line, and sent to the SS.

The sensors on the whole are very reliable with the exception of the BlueEyes sensor, which has a very small field of vision. Increasing the distance of the sensor from the user increases the field of vision, but reduces the accuracy of readings. One fix we tried was to set up cardboard around the sensor to block out any light that may be interfering. But in general, we've found that if the user moves, it is very likely he/she will move out of the field of vision.

Sensor Monitor

The sensor monitor is used as a communication controller/filter between the SystemServer class and sensor data output buffers. Its primary roles are:

Collecting information from the sensors
Sending the information to the SS

A SensorMonitor is created with a specific SystemServer that it reports to. Duplicity of SystemServers and SensorMonitors is not allowed.

The most important part of the SensorMonitor class is its run() method. The following code in the run() method reads the data, and tells the SystemServer class what to do with it:

```
String line = read_socket.readLine();
DataRecorder.record(line);
server.process(SensorMsgFactory.parse(line, this);
```

The process method is used by the System Server to call its data analyzer. The data analyzer is in charge of dispatching the information to the appropriate processor, which will be discussed later.

SensorMsgFactory is used to create a message object from the string representation sent over the network. It creates the correct subclass of SensorMsg (BlueEyesMsg, MouseMsg, etc.) with the correct data, which the appropriate Processor can handle.

Processors & DataAnalyzer

The DataAnalyzer is a dispatcher class for all messages received by the SensorMonitor. It registers listening processors with specific message types (subclasses of SensorMsg). Whenever a message is received, it is sent to the DataAnalyzer, which subsequently dispatches it to the appropriate Processor. Each Processor is responsible for taking appropriate actions on the given message. For example, the SeatProcessor accepts a SeatMsg, factors the posture into a running average and sends the appropriate posture message to the CharacterMonitor. Typically, these processors are relatively simple, however, in two cases some post processing on the data was desirable.

The Gsr (skin conductivity) data was found to be too jumpy even when factoring it into a long running average, so GsrPostProcessingMonitor was created as an intermediate step between the GsrProcessor and the CharacterMonitor. The GsrProcessor, instead of sending message to the CharacterMonitor, sends them to the post processor. The PostProcessor is a subclass of Monitor, and it runs in its own thread which is started with all of the others in the SystemServer. The PostProcessor continuously sends messages to the CharacterMonitor based on the last value sent from the GsrProcessor. If the value sent from the Processor is close enough to the last value sent by the PostProcessor, then the PostProcessor sends the target value, other wise it sends a new value which is a maximum step away from the last value sent in the direction of the target value. This effectively places an upper limit on the change in the skin pigmentation of the character, which gives a smoother, less disruptive response. The BlueEyes PostProcessor works similarly, placing maximum limits on the speed of the head movement and trying to filter out jumpy behavior that occurs when the subject's face is starting to leave the field of view of the camera.

Character Monitor

The character monitor is responsible for all communication from the SS and related modules to the character module. The Character Monitor exposes a static method that can be used to send a message directly to the character. Messages can be delayed to facilitate, for example, an appropriate delay between receiving notification of a change in posture and changing the character's posture. As the character engine responds immediately to all parameter sent to it, the Character monitor is responsible for the delay.

In addition, the CharacterMonitor ignores those messages that it receives while a message of the same type is still "active". For example, a posture message has a certain time duration, and during this time, any changes in posture are not sent to the character.

Character/Game Engine

The Character/Game Engine is third part software responsible for displaying and animating the character and game. It is written in C++ and includes a scripting language for controlling the character. The engine listens for data from the CharacterMonitor and updates the appropriate variable in the scripts. It is also responsible for all GUI interaction between the character and the user.

Data Logger:

Role & Responsibilities

The role and responsibilities of this component are:

Aggregating disparate UDP datagrams from sensors;
Logging receipt of datagrams with timestamp;
Combining these into TCP packets;

The distinction between TCP and UDP should be noted here. UDP is a stateless, connection-less protocol; packets between two UDP sockets (no client/server designation) can be received in any order, or possibly even dropped en route with no notification. TCP is the standard protocol for data transport, since it establishes a connection between a client and server, and guarantees in-order delivery of packets in either direction. Naturally, there is some additional performance overhead in the latter approach to reflect its additional robustness and reliability.

Our sensors are designed to transmit UDP datagrams, yet we needed reliable delivery to the system server. Therefore, we created the DL to group incoming datagrams into TCP packets. Note, however, that this grouping was not cleanly separated among different sensor outputs. For instance, a datagram sent by sensor *A* might be combined with other sensors' datagrams (or possibly even more of its own) into a single packet, a problem which had to be addressed in the design of the SS.

Code Analysis

The DataAggregatorLogger class performs the major functions of the DL. The constructor is the most important method of this class. Its job is to start the execution of two threads that govern communication with other modules. The first thread, DatagramListener, opens a UDP socket on a port specified in the DataAggregatorLogger constructor. DatagramListener's run() method simply blocks on this socket, waiting for any datagrams to come in from the DL and storing them in a queue data structure. The following code, excerpted from DatagramListener, illustrates the salient features of this method:

```
while (true) {
        socket.receive(packet);
        synchronized(semaphore) {
                curr_queue.enQueue(new String(packet.getData(),
                                               packet.getOffset(),
                                               packet.getLength()));
                semaphore.notify();
        }
}
```

The synchronization mechanism (semaphore) is necessary because DatagramListener starts a complementary thread, QueueManager, which handles the datagrams in the queue. With the semaphore in place, we can guarantee that only one thread, either DatagramListener or QueueManager, will access the thread at any on time. QueueManager's main task is to iterate through all the datagrams in the queue and instruct DataAggregatorLogger to handle them with its handleString() method. The following code in QueueManager illustrates this looping process:

```
for (i=0; i<numelements; i++) {
        // next item off queue
        String s = (String)unload_queue.deQueue();
        Sh.handleString(code, s);
}
```

The handleString() method first logs the packet received and the time elapsed since the beginning of the program. Through this log file and timestamp, we can check to make sure each sensor is operational and sending data in a timely fashion. The main function of the handleString() is the sending of these datagrams to the SS as a TCP packet. The following code in handleString() illustrates how DataAggregatorLogger uses a ModifiedTCPListerner (dlss_manager, explained in the next section) to send packets to the SS.

```
if (code == CODE_SENSORS_DATA)
        {
        if (dlss_manager != null) {
                dlss_manager.sendDataThroughSocket(s);
        }
}
```

The *code* field used above is used to identify the source of the call to handleString(). Every class that is granted a handle to DataAggregatorLogger in its constructor (currently DatagramListener and ModifiedTCPListener) is passed a unique code argument during instantiation in DataAggregatorLogger's constructor. When calling handleString(), the listener passes in its own code argument. The conditional statement shows how handleString() differentiates among the various different sources with the code argument. Note that because it currently acts on only one code (CODE_SENSORS_DATA, from DatagramListener), handleString() is not prepared to handle calls from ModifiedTCPListener, whose code is CODE_DLSS_COMM.

The DL is responsible for setting up TCP communication with the SS. The ModifiedTCPListener class is a thread that connects to the SS on the specified remote host and port, and waits for any TCP packets from this module. We do not currently send packets in this direction, so the main use of ModifiedTCPListener is to send packets to the SS through its sendDataThroughSocket() method (as shown in the handleString() code fragment above).

A key design decision made in this module (and indeed, throughout the entire system) was the judicious use of threads to handle network connection activities. Having DataAggregatorLogger block on UDP datagram receipt would have been unacceptable, because it would never stop receiving datagrams and get around to sending TCP packets. Thus, separating datagram receipt and packet transmission into separate threads was a sensible choice.

An additional side benefit of this approach is the ability to test out individual modules separately. For instance, even if we cannot make a connection to the SS, DatagramListener can continue to receive and log UDP datagrams.

Appendix E. The Towers of Hanoi Algorithm for Least Moves Solution

The following is presented from Carlos Rueda's website (Rueda, 1997): http://yupana.autonoma.edu.co/publicaciones/yupana/003/hanoi/hanoi_eng.html

Letting current and final be the initial and final states, the algorithm is as follows:

```
solve ( current, final )
{
   let max be the number of disks
   let dest be the final place of max
   let disk = max
  repeat
     while disk > 0 do
        if disk is already on dest,
        or, moving it succeeds then
           if disk = max then
              decrement max by 1
              if max = 0 then
                 return   // done
             end if
              let dest be the final place of max
          end if
       else
           let dest be the alternative place between dest and
          the current place of disk
       end if
         decrement disk by 1
     end while
      let p and q be the places different of dest
      let disk be the smaller of the disks on top of p and q
      let dest be the place between p and q with greater disk on top
  end repeat
}
```

Appendix F. Committee On the Use of Humans as Experimental Subjects

In this appendix the 2005 renewal, approving the current protocol described in Chapter 4, and the original 1999 protocol are provided on the following pages. The forms used by the subjects in this experiment were the ones were the ones approved on 11/17/2005. The 1999 forms have been omitted.

Application No. 2591

MASSACHUSETTS INSTUTUTE OF TECHNOLOGY

Committee on The Use of Humans as Experimental Subjects

Application for Approval to Use Humans as Experimental Subjects*

PART I

DATE Oct, 1999

Title of Study: Detecting and Interpreting Emotions of Students: Crafting Models and Methods for Intelligent Mentoring Through Affective Computing

Principal Investigator: Rosalind Picard, Associate Professor

Department: Media Arts and Sciences

Room No.: E15-392 The Media Laboratory
E-mail address: picard@media.mit.edu

Telephone No.: 253-0611

Associated Investigators (name & telephone number): Rob Reilly, 253-0153, rob@ceci.mit.edu

Collaborating Institution(s): None

Financial Support: Directorate of Social, Behavioral and Economic Sciences, National Science Foundation, Arlington, Virginia, NSF proposal #9983335

Anticipated Dates of Research: Start Date: Feb 10, 2000 Estimated Date of Completion: Feb 9, 2003

Purpose of Study:

We propose to model the interplay of emotions and learning to demonstrate how to vary the learning environment in response to the transient emotions of the learner so as to optimize the learning process. We will begin by developing tools and methods to recognize and assess emotional expressions of the learner, such as boredom, interest, confusion, and frustration. We will then explore and craft strategies for intervening in an intelligent manner as cued by real-time observation of the emotional state of the learner. To evaluate this system we will first establish a baseline for learning in the absence of any intervention(s) and compare that to learning with interventions that respond to a student's emotions. We envision modeling the learning process as a *feedback control process* where the state is twofold: the partial knowledge of the learner and the emotional state of the learner. The goal is to facilitate the learning journey: complete and accurate learning and the learner enjoys the experience. We expect to demonstrate substantial gains in the quality of the learning experience when we intervene in an optimal manner. In part, to achieve this goal, we need to observe human subjects playing a computer *game*, record their various reactions when the *game* becomes frustrating, boring, too easy, too hard, confusing, etc.

* If you plan to use the facilities of the Clinical Research Center at MIT, please use this for simultaneous submission to the CRC Advisory Committee.

PART II

EXPERIMENTAL PROTOCOL: Please provide an outline of the experiments to be performed.

The physical set-up of the study will consist of two separate computer systems, the sensing system and the game system as well as a video tape camera. One computer system will run the *game*, which is a software simulation created for this study by the Media Lab. The experimental software (the computer *game*) is a simple scenario of a character seeking treasures that are lying about in an adventure-game-style maze of paths. The *game* features some sophisticated-appearing 3-D graphics and animations. The game itself is relatively simple. The subject uses 8 of the numeric keypad keys to move the character through the maze in any of 8 different directions. The on-screen character's movements are constrained to occur only on the 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 position 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. When the character gets near a piece of treasure, it bends down and picks the treasure up. The game concludes when all the treasures are "picked up" by the on-screen character. The *game* is designed to take 25-35 minutes to collect all the treasures in the entire maze. Each subject will be asked to run the *game* 4 times, one trial per day over a 6 week period. The *game*, at various stages becomes frustrating, boring, too easy, too hard, and/or confusing. One group of subjects will be left to their own devices to complete the game. The other group of subjects, based upon their emotion state (e.g., bored, confused, frustrated), will be given social, emotional-content feedback strategies in an effort to help relieve/support their emotional state. These strategies are designed to provide many of the same cues that skilled human listeners employ when helping guide students through a learning journey. Our goal is to identify which cues are most effective, which are least effective and which may be irrelevant.

PART III Please answer each question below, and indicate "NA" where not applicable to your application. Positive answers should be briefly explained, with detailed information included in Part II.

1. How will you subjects be obtained? Voluntary participation by children at 2 public schools
   Number of subjects? Approximately 40
   Age(s) of subjects? 8-12 years old (2^{nd} thru 6^{th} graders)

2. Will female and minorities be recruited? Yes If *not* explain why.

3. Will subjects receive any payment or other compensation for participation? No

4. Will your subjects be studied outside MIT premises? Yes
   If so, please indicate location. Computer Classroom, Lanesborough Elementary School, Lanesborough, Mass.
   Computer Classroom, Hancock Central School, Hancock, Mass.

5. Will facilities of the Clinical Research Center be used? No

---

For proposed investigations in social sciences, management, and other non-biomedical areas, please continue with question 9.

---

6. Will drugs be used? NA

7. Will radiation of radioactive materials be employed: NA

8. Will special diets be used: NA

9. Will subjects experience physical pain or stress? NA

10. Will a questionnaire be used? No

11. Are personal interviews involved? No

12. Will the subjects experience psychological stress? No

13. Does this study involve planned deception of the subjects? No

14. Can information acquired through this investigation adversely effect a subject's relationship with other individuals (e.g., employee-supervisor, patient-physician, student-teacher, co-worker, family relationships)? No

15. Please explain how the subject's anonymity will be protected, and/or confidentiality of the data will be preserved.

    The subjects will be referred to only by their assigned ID number. The only link between their actual identity and their ID number will be the consent forms, which contain both name and ID number. During the experiment the consent forms will be secured in a private archive that will not be available to anyone except the PI and Co-PI. Upon completion of the experiment, the ID numbers will be removed from the consent forms.

PART IV

A. *Please summarize the risks to the individual subjects, and the benefits, if any*; include any possible risk of invasion of privacy, embarrassment or exposure of sensitive or confidential data, and explain how you propose to deal with these risks.

   There is "minimal risk" to the individual subjects as they are merely running computer software as they would in any typical computer education or other academic class. They are being observed, and their behavior is being recorded, but the "risk" in this study is virtually non-existent. There may be some vicarious benefit/learning occur as a result of this research, but it is, in all likelihood, negligible.

B. *Detection and reporting of harmful effects*: Please describe what follow-up efforts will be made to detect any harm to subjects, and how this Committee will be kept informed.

   During each session, the investigator, who is an experienced educator, will be attentive to the emotional state of the subject, which is the focus of the study. If the emotional state of the subject rises to lowers to an unacceptable level or other problems are identified (e.g., subject becomes too frustrated, too anxious about succeeding or failing, anxiety about "how they did" in the study, felt badly because they did not receive any experimenter assistance where other children in the study did), the investigator will terminate the session, document the events, and then inform the PI of this event. The PI will then inform COUHES. The subject's parents will also be alerted to such an event. The parents will also be advised to be watchful for any problems resulting from the research study through the "parent cover letter" and the "parent's end-of-subject's-participation letter."

Signature of Principal Investigator ________________________________ DATE__________

Print Full Name: Rosalind W. Picard

Signature of Department Head: ________________________________ DATE__________

Print Full Name: ____

Please return this application with 3 photocopies to:

Leigh Firn M.D.
COHUES Chairman
E-23-230
253-6787

The Research Method

1 Summary of the potential of affective computing

Emotion was identified by Donald Norman in 1981 as one of the twelve major challenges for cognitive science. In this proposal we have argued that emotions can no longer be considered a *luxury* when studying essential rational cognitive processes. Instead, recent neurological evidence indicates they are necessary not only in human creativity and intelligence, but also in rational human thinking and decision-making. We have suggested that if computers will ever interact naturally and intelligently with humans, then they need the ability to at least recognize and express affect.

Affective computing is a new area of research, with recent results primarily in the recognition and synthesis of facial expression, and the synthesis of voice inflection. However, these results are just the tip of the iceberg—a variety of physiological measurements are available that would yield clues to one's hidden affective state. Moreover, these states do not need to be universal in their expression for a computer to recognize them. We have developed some possible models for the state identification, treating affect recognition as a dynamic pattern recognition problem. This proposal will hopefully provide needed research to discover which of these tools, coupled with which measurements, both of the person and their environment, give reliable indicators of affect for an individual in a learning situation.

2 The Research Plan

In this work, inspired by the particular application of human↔machine interaction and the potential use that human-computer interfaces can make of knowledge regarding the affective state of a user, we investigate the problem of sensing and recognizing typical affective experiences that arise in this setting. In particular, through the design of experimental conditions for data gathering, we approach the problem of detecting, among other states, frustration, exhilaration, confusion, and interest in human↔computer interfaces. By first sensing human biophysiological correlates of internal affective states, we proceed to stochastically model the biological time series with Hidden Markov Models (HMM) to obtain user-dependent recognition systems that learn affective patterns from a set of training data. Labeling criteria to classify the data are discussed, and generalization of the results to a set of unobserved data is evaluated. Final recognition results will be reported under two conditions, for the entire data set, and only for those subjects with sufficient experimental data.

A variation on the HMM could also be used to incorporate affective feedback. The recent Partially Observable Markov Decision Processes are set up to give a 'reward' associated with executing a particular action in a given state. These models permit observations at each state, which are actions; hence, they could incorporate not only autonomic measures, but also observations of your behavior.

2.1 Modeling an Affective System

Consider what happens when you attempt to recognize someone's emotion. First, your senses detect low-level signals-motion around their mouth and eyes, perhaps a hand gesture, a pitch change In their voice and, of course, verbal cues such as the words they are using. Signals are any detectable changes that carry information or a message. Sounds, gestures, and facial expressions are signals that are observable by natural human senses, while blood pressure, hormone levels, and neurotransmitter levels require special sensing equipment. Second, patterns of signals can be combined to provide more reliable recognition. A combination of clenched hands and raised arm movements may be an angry gesture; a particular pattern of features extracted from an electromyogram, a skin conductivity sensor, and an acoustic pitch waveform, may indicate a state of distress. This medium-level representation of patterns can often be used to make a decision about what emotion is present. At no point, however, do you directly observe the underlying emotional state. All that can be observed is a complex pattern of voluntary and involuntary signals, in physical and behavioral forms.

Not only do you perceive expressive signals from a person, but you also perceive non-expressive signals from the environment which indicate where you are, who this is, how comfortable the weather is, and so forth. These signals indicate the context, such as the fact that people are in an office setting, or that it is final exam season. The observer may notice that the weather is oppressive and reason that it could impact moods. Or, the context might be recognized as a situation where a person is expecting some exciting news. With contextual information, the observer proceeds not only to analyze low-level signals and patterns from the environment and from the person who is expressing an emotion, but also to reason in a high-level way regarding what behavior is typical of this situation, and what higher-level goals are at work.

The process of trying to recognize an emotion is usually thought to involve a transformation from signal to symbol, from low-level physical phenomena to high-level abstract concepts. However, because reasoning about the situation can modify the kinds of observations that are made, information can be considered to flow not just from the low-level inputs to the high-level concepts, but also from the high level to the low level. Suppose that in reasoning about a situation you expect that somebody will be in a bad mood; in that case, your high-level expectation can cause your low-level perception to be biased in a negative way, so that you are more likely to perceive a weak or ambiguous expression as being negative. The recognition of emotions is therefore not merely bottom-up, from signals to symbols, but also top-down, in that higher level symbols can influence the way that signals are processed.

High-level reasoning and low-level signals also cooperate in the generation of emotional expression. Suppose that an actor wishes to portray a character that feels hatred. He might begin by thinking, "I want to show hatred" and then proceed to synthesize low-level signals that communicate hatred, changing his posture, behavior, voice, and face, to reflect this emotional state. The whole process has started as a symbol-a cognitive goal to show hatred-and has ended with the generation of expressive signals, so that the audience can recognize his character's hatred. The process of trying to express an emotion is usually considered to involve a transformation from symbol to signal, from high-level concepts to low-level modulation of expressions and behaviors.

However, we have left one important piece out of both of the above descriptions: the emotional state of the system, which is either recognizing or expressing the emotion. In humans, this distinction is blurred because all humans have emotional states that automatically influence recognition and expression. But in computers, this distinction needs to be made explicit because a computer can be built with only a subset of these abilities. To recognize an emotion involves perception. But we know that human perception is biased by human emotion: an observer's own emotions influence both his low-level perceptual processes and his high-level cognitive processes. An observer will tend to perceive an ambiguous stimulus as being positive or negative, whichever is congruent with his mood. The emotional state of a human also influences her emotional expression. If the actress thinks, "Show hatred," then she may also begin to feel hatred. It can also be the case that simply posturing her muscles to accurately communicate expressions of hatred can provide bodily feedback to cause her to actually feel the emotion she is expressing. In these cases, the emotional state, if there is one, interacts with both the recognition and expression of an emotion, with both cognitive and physical processes, and with both high-level reasoning and low-level signal processing. More commonly, a person will find that an emotional state simply arises in response to perceiving or reasoning about some events, and expression of that state occurs mostly involuntarily. Figure 2 summarizes these interactions.

When a computer tries to represent emotions and their expression, it may use convenient levels of abstraction-from a low-level representation of a signal such as a waveform of heart rate or a motion sequence of muscular movements, to a high-level interpretation such as the sentence "He looks sad." At no point in this process does the computer have to use the same mechanisms used by humans; it might go about recognition, expression, and synthesis of emotions in an entirely different way. However, to the extent that we can understand the way humans do these things, we will have a better idea how to give these abilities to computers. Furthermore, to the extent that we imitate human mechanisms in computers, we have a better chance of debugging the computers when they behave in a peculiar way, and we stand to benefit because the ways in which they behave are likely to be close to ways that humans behave, making it easier for us to interact with them.

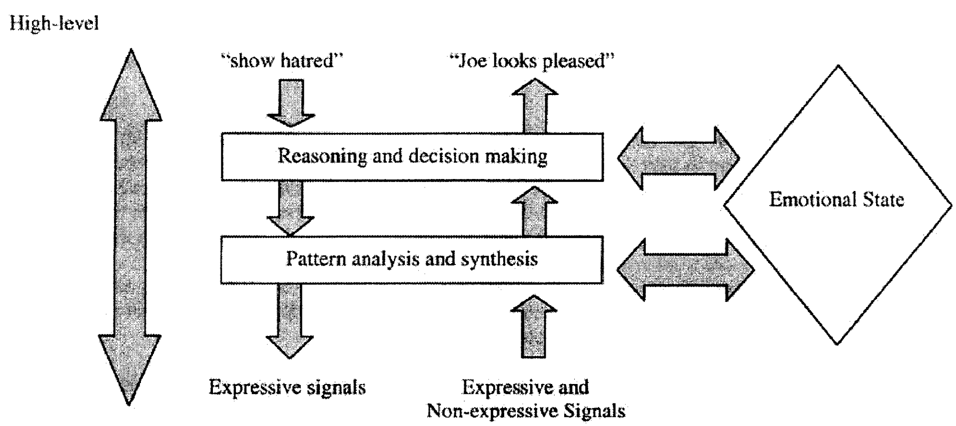


Figure 2: Information flows from high to low in a system that can express, recognize, and 'have' emotions.

2.2 Example of Experimental Design

Research in the psycho-physiological community has proposed that affective states are mapped onto physiological states that we may be able to measure. In order to collect data carrying affective content, an experimental situation has been designed so that users, engaged in a computer task, will experience various emotional incidents while his or her physiological signals are being collected and recorded for further processing.

The main objective of this section is to develop models and techniques, which we can apply in real time to track physiological signals and make inferences about the level of arousal of a subject. We envision this project being a useful building block that can be integrated into a computer that uses this information to adapt itself to -the needs of the user. This more ambitious idea goes beyond the present scope of this thesis but is a future research topic in this area.

Human physiology behaves like a complex dynamical system in which several factors, both external and internal, shape the outcome. In approximating such a system, we are interested in modeling its dynamical nature and, given that knowledge of all the independent variables that affect the system is limited. We want to approach the problem in a stochastic framework that will help us model the uncertainty and variability that arise over time. A class of models that has received much attention in the research community over past years to model complex dynamic phenomena of a stochastic nature is the class of Hidden Markov Models (HMM). HMMs have been widely used for modeling speech and gesture, and are currently an important building block of speech recognition systems. Motivated by their flexibility in modeling a wide class of problems, we decided to study the feasibility of using HMMs to model physiological patterns that are believed to correlate with different affective states.

Any future affect recognition system would be presented with an enormous amount of data with broad scope. Our initial approach to developing experiments for affective data collection was to create one relatively straightforward experiment, and to control possible extraneous variables as much as possible. It was decided to focus on a "frustration" task; an experimental paradigm which induced negative arousal in the subject, and roughly corresponded to the kind of hardware or software problems that users encounter in everyday interaction with computers.

In this experiment, we brought subjects into the lab under the pretense that their task is to participate in a vision-oriented computer game. The game consisted of a series of puzzles, and the task was to click the mouse on the correct answer at the bottom of the screen to advance the screen to the next puzzle. Subjects received $10 for their participation, but the game was also a competition; the individual who received the best overall score and speed at the end of the data collection was awarded a $100 prize. Thus, an incentive was created to increase subjects' desire to play quickly and receive a good score.

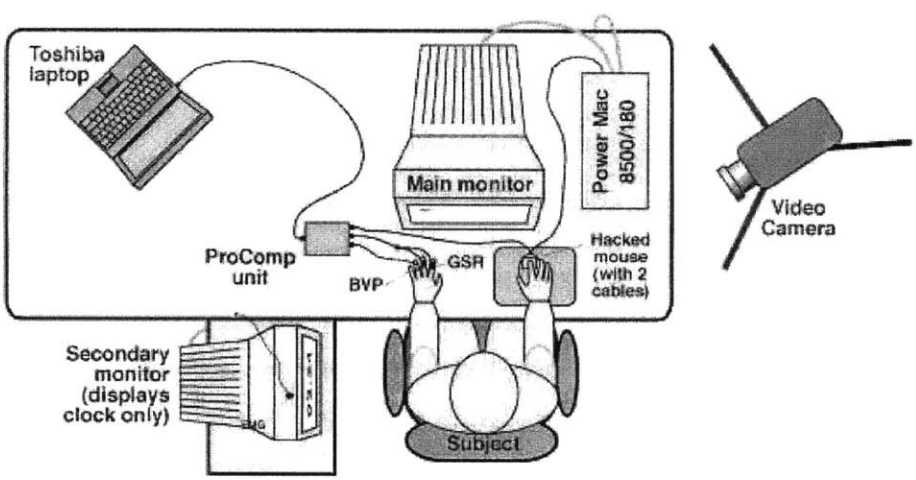


Figure 3: A schematic of the frustration experiment design. Two separate systems, the sensing system (run by the Toshiba laptop, top left) and the game system (run by the Power Macintosh, next to the main monitor), are tightly synchronized via a hacked mouse that sends a signal to both systems each time the mouse button is clicked.

2.2.1 Hidden Markov Models—Markov Processes

Stochastic processes "with memory" are processes in which the present value of the process depends on the recent history undergone by the process. Let us consider a discrete process generated by a random variable s which at time t takes a value on a finite set $S = \{S_1, S_2, \ldots, S_N\}$. Such a discrete stochastic process is said to be J^{th} order Markov if it satisfies the Markov property; that is, the conditional probability of the process given all past and present values depends only on the j most recent values:

$$P(S_t|S_{t-1},S_{t-2},\ldots S_{t-j}, S_{t-2},S_{t-j-1},\ldots) = P(S_t|S_{t-1}, S_{t-2},\ldots,S_{t-j})$$

When j = 1, we obtain a first order Markov process in which the value of the process at times t depends only on the value at time t - 1. In this case, the process is completely characterized by its first order state transition probabilities:

$$a_{ij} = Pr(S_t = S_i \mid S_{t-1} + S_j)$$

Consider a dynamic system with a discrete finite state space. At time t, this system finds itself in one of N states and takes on a value generated by a state-dependent probabilistic distribution. An HMM is a model for a dynamic system in which a discrete Markov process is used to describe the dynamic properties or state evolution of the system, and state dependent probability distributions (discrete, continuous, or mixed) axe used to model the observable outputs of the system, For this reason, HMMs are also known as doubly stochastic processes since there axe two level of random processes in the underlying model: one which remains hidden describing the state occurrences, and another one at each state modeling the observable outputs of the system. The general structure of an HMM is shown in Fig. 4. The circles of the diagram indicate different states, and the arrows indicate probabilistic transitions between states. The squares indicate observable outputs from the HMM. Notice that there axe two non-emitting states in this diagram. These states are reserved for the initial and final state of the model and allow the HMM to generate observations according to its own dynamics while ensuring that the initial and final states axe always visited. The functionality of the non-emitting states becomes clear if we want to build composite models in which several single models axe concatenated to model sequences which do not contain a single class, but rather several classes (as might be the case for a speech fragment containing several words, or a video sequence containing several facial expressions).

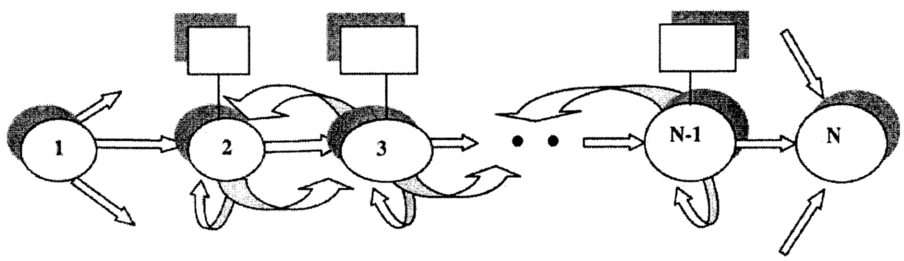


Figure 4: HMM with initial and final non-emitting states

For a first order HMM, the process of generating a dynamic system which follows the model consists of generating a state sequence according to the transition probabilities and then sampling from the output distribution associated with the state visited at time *t*. The problem of interest, however, usually consists of inferring the underlying model from a set of sequences, which are assumed to have been generated by a common model. To this inference problem, we turn next.

2.2.2 Estimation

We will formally define the parameters describing the model above (figure 4). Consider that a set of M observation sequences $\{X^m\}^M_{m=1}$ is available, and let:

T_m be the length of the m^{th} observation sequence $X^m = x^m_1, x^m_2, \ldots, x^m_{Tm}$

N be the number of states in the model

$S = \{s\}$, $s_t = i$ at time t, and $i = 1,2,\ldots, N$

$\pi = \{\pi_i \mid \pi_i = Pr(s_1 = i)\}$, initial state probabilities

$A = \{a_{ij}\}$ where $a_{ij} = Pr(S_t = S_i \mid S_{t-1} + S_j)$

$B = \{f_i(x)\}$ where $f_i(x)$ is the probability density function associated with state i. In the most general case, we will model density functions with mixtures of K Gaussian-s such that:

$$f_i(x) = \sum_{k=1}^{K} c_{ik} N(x, \mu_{ik}, \Sigma_{ik}) \quad \text{with} \quad \Sigma_{ik} = 1, 2, \ldots, N$$

For a given HMM of order N, then the set of parameters $\theta = \{A, B, \pi\}$ define the model completely.
Given a set of observation sequences, there are three (3) basic inference problems we need to address: the probability of an observation under the parameters of the model, how to modify the parameters of the model so as to maximize the probability of the data set, and lastly, what is the sequence of hidden states under the current model for the observation in question. We will address each one of these in the next sections.

We will calculate the Forward-Backward Algorithm, The Baum-Welch Re-estimation for a Single Model, The Embedded Baum-Welch Re-estimation Algorithm, and Recognition: The Viterbi Algorithm based upon previous research in this area by Fernandez.

2.3 Establishing a Ground Truth

We wish to treat this problem as a classification problem and determine whether we can characterize and predict the instants of arousal from a set of observed physiological signals. Before proceeding to do this, a ground truth needs to be established in order to determine a standard by which to evaluate the classifications. This is a non-trivial problem, which deserves careful consideration since the class categorizations we shall use to label the data have only been induced, not firmly established. In other words, there is an uncertainty associated with the class to which the data belongs. There is, for instance, a possibility that a stimulus failed to induce a high arousal response, and conversely, that a subject showed a high arousal response in the absence of the controlled stimulus due to another uncontrolled stimulus. In the classical recognition problem a set of data is used for learning the properties of the model under the different classes to recognize. The classification of this training data is usually fixed, and this knowledge is then used to derive the properties of the separate classes. We do not wish to abandon this framework entirely and will adopt a deterministic rule to label the training examples. However, establishing a proper labeling for the training data is one of the aspects of this problem, which should be adaptive and subject to further discussion.

Our only degree of belief about what class the data belongs to is given by the onset of the pre-controlled stimuli during the course of the experiment. A rather intuitive approach to define the classes is to consider the response following a stimulus as representative of a "frustration" episode. How we establish the temporal segmentation following a stimulus deserves some attention. The time window we use to capture this response has to be wide enough to allow a latency period, as well as the true physiological response due to the stimulus. The latency period consists of the time lag, which elapses between the onset of the stimulus and the start of the physiological change due to the stimulus. Some authors have established that for galvanic skin response this delay can be as much as 3 seconds. Following the onset of the stimulus, we allow a dormant period of *l* seconds to pass before we start assigning the labels; then we window the following 10 seconds of data as representative samples of the class we want to model as 'frustratiore' (F). In order to transition out of this class, since the model boundaries are not known with precision, we allow another dormant period (of 5 seconds) without any classification, and then consider the rest of the signal up until the next stimulus to correspond to the class of "non-frustration" (T). If the remaining set of samples is less than a minimum number of samples required to assign a label (3 seconds in these simulations), then a label is not assigned to this region. If the time windows used on two adjacent stimuli overlapped (the stimuli were spaced out by less than 10 seconds,) then the two resulting segments of data labeled as F would be merged together.

The chosen labels may be viewed as corresponding to positive and negative examples of the phenomenon we want to model. The reader should bear in mind, however, that this is a simplified mnemonic and modeling device and not an argument for what the true state of the physiology is since we can safely assume that human physiology exhibits much widely complex modes of behavior. The labeled regions roughly correspond to areas in which we have a higher degree of confidence about the class induced, whereas the unlabeled regions represent "don't-care" regions where our knowledge of the transition between affective states is too poor to include in the ground truth.

2.4 Stimulus Habituation and Anticipation

The procedure described above is the basis for establishing a ground truth, which we used for the data analysis. There are several other variations which could result by simply adjusting some of the temporal parameters we used to define the labeling. How the variation of these parameters affects the overall classification performance is an open area of research. There is another aspect of the experiment that may condition the classification of the data, namely the user's expectation of a stimulus after a certain habituation period has elapsed. We decided to investigate whether there can be any effect on the performance of the models by redefining the ground truth in a way that accounts for a user's expectation of a stimulus. To do this, we adjoined to the set of actual stimuli a set of "virtual" stimuli where a user might have expected them. The idea would be, for instance, that if a user becomes aware that after T seconds from the first incident the mouse failed to work again, he might also anticipate a third failure approximately T seconds following the second failure (unless the actual third mouse failure happens first). Based on this, we used the following simple algorithm to augment the set of original stimuli. Let $\{t_i\}^N_{i=1}$ denote the set of time instants of the *N* original stimuli. Then:

$T = t_2 - t_1$
for $i = 3$ to N
    if $T < t_i - t_{i-1}$
        Insert a stimulus at $t = t_{i-1} + T$
    $T = t_i - t_{i-1}$

With the new set of augmented stimuli, we could then redefine the labels assigned to the data by following the procedure outlined in the previous section. The analysis of the data (model training and testing) can then be carried out independently according to each of the ground truths established in this section.

2.5 Feature Extraction

The biosignals collected during the experimental sessions described above consisted of galvanic skin response and blood volume pressure. From these signals we need to obtain features that bear some relevance to the recognition problem at hand, namely a set of features that might have correlates with internal affective states. This is one of the most important research problems that exist in this area: the mappings between affective states and physiological states is still an area which is being investigated at large in the psychophysiology community. In deciding on a feature set, we must account for classical measures of affective states (i.e. level of arousal as registered in a GSR signal, heart acceleration, etc), while bearing in mind that we can also allow the models we are using to exploit more complex dynamic patterns that might not have received so much attention in other studies.

2.6 Classification Results and Discussion

The experiment described above was evaluated on 24 subjects. On 21 of the 24, the computer's recognition accuracy (using the HMM approach) was significantly greater than random as measured against the ground truth. However, the recognition accuracy was still not 100%, suggesting a number of things: the features and signals used are not optimal, the ground truth may not be perfect, and the states are probably more complex than the two states assumed in this initial model of "low-arousal" for when things were going smoothly vs. "high-arousal" for when things were not going smoothly (i.e., the mouse didn't respond). Recent experiments with two additional physiological signals (for a total of four) and with eight emotional states deliberately expressed by the user, have yielded emotion classification accuracy of 81% on a single subject's data gathered over many weeks [33]. In the latter case, we focused on measuring emotional expression of a single user over weeks, as opposed to traditional emotion studies that focus on lots of users over a few minutes or an hour. We found that the day-to-day variation in expression of a single emotion was often significantly greater than the same-day variation among different emotions, indicating that long-term studies will be crucial in developing robust affect recognition systems.

2.7 Beyond Physiology

Although most of our efforts to recognize emotional information from the user have centered on physiological signals, we do not want to have to require students to be wired up to physiological sensors. Skilled humans can assess emotional signals without directly reading physiological changes, and we believe computers can be given similar abilities. Hence, we are developing a variety of means of emotional communication. We have already developed tools for emotional expression via dialogue boxes with radio buttons and via sensors in eyeglasses that detect muscle movements indicative of confusion or interest. We have also developed skin-surface sensors of gestural information and video-based analysis tools for other forms of human behavior. We propose to develop additional computer-vision and computer-audition means of looking at and listening to people in an unobtrusive way.

In every means of sensing, we believe users should be in control of their emotional communication, so we design with this value in mind. We think it's important to educate users about the choices available among the different sensing possibilities. For example, to sense a facial expression, the computer could use a camera (sensitive to lighting, currently restricts head movement, relative small set of emotion categories recognizable, can "see" identity as well as expression, but nothing touches user's face) or the computer could use a set of sensors designed to capture only the facial expression (not sensitive to lighting, no restriction on head movement, more exact on broader set of emotional expressions, private in not detecting identity, but presently requires potentially obtrusive contact with the face). We propose to develop a variety of means of sensing emotional communication, so that users can choose the form that is most comfortable and respectful of their needs.

MIT Committee On the Use of Humans as Experimental Subjects

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 Massachusetts Avenue
Cambridge, Massachusetts 02139
Building E32-335
(617) 253-6787

To:	Rosalind Picard E15-001
From:	Leigh Firn, Chair COUHES
Date:	11/17/2005
Committee Action:	Renewal
COUHES Protocol #:	0403000140
Study Title:	Detecting, Interpreting and Attending to Emotions of Students
Expiration Date:	11/17/2006

The above-referenced protocol was given renewed approval following Expedited Review by the Committee on the Use of Humans as Experimental Subjects (COUHES).

It is the Principal Investigator's responsibility to obtain review and continued approval before the expiration date. You may not continue any research activity beyond the expiration date without approval by COUHES. Failure to renew your study before the expiration date will result in termination of the study and suspension of related research grants.

Adverse Events: Any serious or unexpected adverse event must be reported to COUHES within 48 hours. All other adverse events should be reported in writing within 10 working days.

Amendments: Any changes to the protocol that impact human subjects, including changes in experimental design, equipment, personnel or funding, must be approved by COUHES before they can be initiated.

Prospecitve new study personnel must, where applicable, complete training in human subjects research and in the HIPAA Privacy Rule before participating in the study.

COUHES should be notified when your study is completed. You must maintain a research file for at least 3 years after completion of the study. This file should include all correspondence with COUHES, original signed consent forms, and study data.

cc: Tom Duff

ASSENT TO PARTICIPATE IN RESEARCH

Detecting, Interpreting and Attending to the Emotions of Students

1. My name is Winslow Burleson.

2. We are asking you to take part in a research study because we are trying to understand how emotions and learning work together while using a computer and what responses to emotions are beneficial.

3. If you agree to be in this study you will be playing a computer game. You will sit in a pressure sensitive chair that will record your movements. You will use a pressure sensitive mouse that will record your mouse activity. You will wear a skin conductivity sensor on the palm of your hand that will record signals from your skin. This experiment will take about 35-45 minutes. You will receive a $10 gift certificate to Amazon.com for your participation in this study.

4. There are no risks expected from your participation in this study.

5. You may learn something and you may enjoy the study.

6. Please talk this over with your parents before you decide whether or not to participate. We will also ask your parents to give their permission for you to take part in this study. But even if your parents say "yes" you can still decide not to do this.

7. If you don't want to be in this study, you don't have to participate. Remember, being in this study is up to you and no one will be upset if you don't want to participate or even if you change your mind later and want to stop.

8. You can ask any questions that you have about the study now. If you have a question later that you didn't think of now, you can call me 1-617-308-5875. You can also call the Chairman of the Committee on the Use of Humans as Experimental Subjects at M.I.T. at 1-617-253 6787 if you feel you have been treated unfairly.

9. Signing your name at the bottom means that you agree to be in this study. You and your parents will be given a copy of this form after you have signed it.

______________________________ Date ________________

Name of Subject

Additionally, please read the following paragraph: Videotapes and/or audiotapes of your participation may be collected. This data will be used for experimental purposes only, and after the data collection is over, they will be 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 information, such as 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. At any time during or after the experiment you may request that your tapes be destroyed.

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

______________________________ Date ________________

Name of Subject

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

NOV 17 2006

Detecting, Interpreting and Attending to the Emotions of Students

You are asked permission for your child to participate in a research study conducted by Professor Rosalind Picard, Winslow Burleson and Devin Neal from the Media Lab at the Massachusetts Institute of Technology (M.I.T.). These results will contribute to a dissertation. Your child was selected as a possible participant in this study because we are seeking children from the general public. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

- **PURPOSE OF THE STUDY**

The study will research the interplay of emotions and learning so that we can more fully understand how to vary the learning environment in response to the transient emotions of the learner so as to optimize the learning process. We have developed tools and methods to recognize and assess emotional expressions of the learner, such as boredom, interest, confusion, and frustration. But we need to field test theses beliefs and theories.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask your child to do the following things:

Your child will play a computer game and we will be video taping them doing this. In addition to the video camera your child will be sitting on a pressure sensitive chair that will record movements, using a pressure sensitive mouse that will record activity and wearing a skin conductivity sensor on their palm that will record skin conductivity. These devices will not, in any way, restrain your child; he or she merely sits on the chair, uses the mouse and wears the skin conductivity sensor. While your child is playing the computer game the experimenter may assist them by providing suggestions, clues or hints to help your child solve the on-screen puzzle; but beyond that your child will be sitting and running a computer game. This activity will last 35-45 minutes.

- **POTENTIAL RISKS AND DISCOMFORTS**

There are no anticipated risks or discomforts.

- **POTENTIAL BENEFITS**

Your child may learn something about them themselves and his/her emotions and may enjoy the study. The study will help to develop a theory of learning and the importance of emotion to learning.

- **PAYMENT FOR PARTICIPATION**

There will be a $10 gift certificate to Amazon.com received for participation in this study.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact Winslow Burleson at 1-617-308-5875 (229 Vassar St., Cambridge, MA 02139); Rosalind Picard at 617-253-0611 (20 Ames St. E15-020G, Cambridge MA, 02139). You can also call the Chairman of the Committee on the Use of Humans as Experimental Subjects at M.I.T. at 1-617-253 6787 if you feel you have been treated unfairly.

- **EMERGENCY CARE AND COMPENSATION FOR INJURY**

In the unlikely event of physical injury resulting from participation in this research you may receive medical treatment from the M.I.T. Medical Department, including emergency treatment and follow-up care as needed. Your insurance carrier may be billed for the cost of such treatment. M.I.T. does not provide any other form of compensation for injury. Moreover, in either providing or making such medical care available it does not imply the injury is the fault of the investigator. Further information may be obtained by calling the MIT Insurance and Legal Affairs Office at 1-617-253 2822.

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E32-335, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

- CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.

Videotapes and/or audiotapes of your child's participation may be collected. This data will be used for experimental purposes only, and after the data collection is over, they will be 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 information, such as 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. At any time during or after the experiment you may request to review or edit the tapes and/or request that your tapes be destroyed.

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

____________________________________ __________________

Name of Subject Date

Name of Legal Representative (if applicable)

____________________________________ ______________

Signature of Subject or Legal Representative Date

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____________________________________ __________________

Name of Subject Date

____

Name of Legal Representative (if applicable)

____________________________________ ______________

Signature of Subject or Legal Representative Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

____________________________________ ______________

Signature of Investigator Date