Creating Believable Emotional Virtual Characters

by

Neesha Desai

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Computing Science
University of Alberta

Abstract

Believable, realistic video game character behaviour continues to lag behind the improvements in graphics, stories and game play in video games. In this dissertation we focus on the use of two techniques, emotional gaits and emotional incidents, as a way to add easily identifiable, non-verbal, and non-facial emotion to background game characters, thereby increasing the believability of these characters. Emotional gaits refers to the body posture, hand/arm positioning, walk and walking speed of the characters. An emotional incident is an emotion-specific interaction between characters or between characters and props within the game world. The selection and implementation of the techniques was designed to be easily scaled to large numbers of characters and require a minimal number of additional animations.

These techniques (emotional gaits and emotional incidents) were analyzed through six different user studies. The examination focused on three aspects: 1) the ease of emotion identification when the behaviour was isolated, 2) whether the gender of the participants and characters affected the results, and 3) emotion identification when observed during normal game play. The results show that participants were able to accurately identify the emotions, that the combination of both emotional gaits and emotional incidents was best overall (but some emotions could be equally achieved with only one), that there were some small differences based on participant gender, and that participants could easily and quickly learn to identify the character emotions when observed within a game world.

Preface

All user studies described in this dissertation received ethics approval from the Research Ethics Office at the University of Alberta.

- Study 1 - Identifying Character Emotion in Video Games *Pro00025530*, Chapter 4
- Study 2 - Gender perceptions of video game character emotion *Pro00030705*, Chapter 5
- Study 3 - Emotional Characters and Game Play *Pro00038529*, Chapter 7
- Study 3.5 - Believable Emotional Characters *Pro00038930*, Chapter 7
- Study 4 - Changing Characters Emotion *Pro00042256*, Chapter 7
- Study 5 - Guess Who - Video Game Character Study *Pro00046111*, Chapter 8
- Study 6 - Identifying Character Emotion in Project Spark *Pro00050746*, Chapter 9

Key results from Chapter 4 have appeared in a published paper:

Neesha Desai, and Duane Szafron. Enhancing the Believability of Character Behaviors Using Non-Verbal Cues, *Proceedings of the Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, Stanford, USA, October, 2012, 130-135.

Acknowledgements

Grad school is not a solitary experience. It requires many supporters along the way.

First up, I would like to thank my family. My parents, Nancy and Pete, and my sisters, Raani and Soni (and their families)have been there for me the entire way. They listened to my rants, cooked me meals and worked to cheer me up when times were tough.

Next, my supervisor (Duane Szafron) and the Believe research group (both past and present members) were always available for bouncing ideas off, testing out experiments and so much more.

I have been lucky to have made great friendships within the department but also within the startup community in Edmonton. The people who have climbed with me, shared a beer, or helped launch Alieo, have all helped keep me sane during the highs and lows of research.

Finally, I would be remiss if I did not thank the Natural Sciences and Engineering Research Council of Canada, Alberta Innovates Technology Futures, and the GRAND-NCE, all of whom helped fund my research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Video games have become an integral part of the daily lives of many people. As of 2014, the Entertainment Software Association noted that, in 59% of American households, people were playing video games (on consoles, phones or computers), with an average of two gamers per household [3]. While we generally associate video games with young children, less than 1/3 of game players are under 18 years old (29%) and the average gamer is now 31. The gap between males and females has continued to shrink, with the percent of female gamers at 48% in 2014. Over $15 *billion* was spent exclusively on video game content (ignoring consoles and accessories) in 2013.

Over the past few decades, video games have vastly improved in the areas of graphics, stories and game play mechanics. Therefore, it is not surprising that the quality of the graphics and the interesting storylines are two of the top reasons consumers give for buying games [3]. However, character behaviour has not kept pace, and the static and repetitive behaviour of game characters remains a problem. This can be seen in Mass Effect 2 [14]: no matter when the player walks through Commander Sheppard's ship, the non-player characters are in the same spot, performing the same set of repeated actions on loop and rarely acknowledging the player's (Commander Sheppard) presence. In Oblivion: The Elder Scrolls [11], the frequency in which non-player characters would be having a conversation about mud crabs became a running joke, as can be seen in the large collection of YouTube clips available on the topic. One fan of Skyrim [12] was so irritated by the frequency and timing of the non-player character greetings (they would introduce themselves over the top of an existing conversation) that he created a mod that removes them completely [17]. And in Hitman [73], even when the player has armed their character and there are dead characters on the ground, other non-player characters seem oblivious to this as they calmly walk across the bodies (see Figure 1.1 for a screenshot from a YouTube clip by b33k3rz [5].

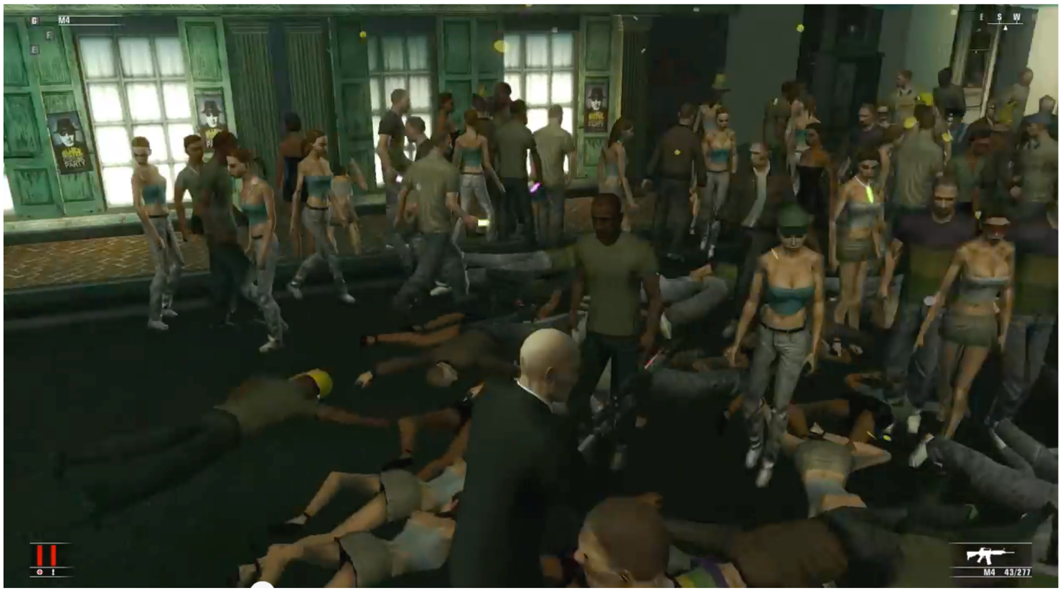

Figure 1.1: Screenshot from a YouTube clip by b33k3rz [5] showing non-player characters calmly walking across dead bodies as the player (the hitman) holds his gun.

The complex video game worlds we see in games like Skyrim and Mass Effect 3 [15] rely on the fact that they have hundreds of background characters (Skyrim has over 500 *named* characters [76]). These characters are needed so that the towns and space stations do not appear empty and deserted as the player explores the game world. Many of these games are open world, in that players do not have to follow a specific route or path through the game, allowing players to explore and interact at their own discretion and pace. This allows for diverse and complicated stories to emerge. These stories often include many of these background characters. However, unless a character is involved in a specific quest or game event, they rarely have any character-specific behaviour that sets them apart from other game characters. These are the characters that are often found standing in the same spot, doing the same thing, every time a player approaches them.

The lack of quality and the repetitive behaviour of these characters can distract from the overall experience when playing. However, creating complicated and realistic behaviour can be expensive. It requires many person-hours to custom-code each character. Spending the time coding these characters, when they may only be seen once or twice for a very short period of time, can feel like a waste of resources for a game company. Instead, it is easier to make scripts that can be shared between multiple characters. This, however, inevitably results in characters with near-identical performances, as they act and react according to the same script.

When specific attention is paid to characters, it is generally (and reasonably) focused on the non-player characters that have the most direct interaction with the player. Much of this attention has been spent scripting dialogue and cut scenes, such that they often play like a movie. In these scenes, the characters may suddenly "come alive" as they act more life-like, with changing body posture and gestural reactions to given events. For the brief time these characters are "alive", players have a better chance at relating to and identifying with them. It is during these cut scenes and dialogue interactions that the majority of character emotion is shown.

Character emotion can be incredibly powerful. This is well-known to writers as a way to have their audience connect with their characters [52]. It can be equally powerful in games, helping to pull the player into the storyline and getting them invested in the outcome. There are multiple ways to convey emotion in games, including (but not limited to): character facial animations, level design, background music, and dialogue (especially when voice acting is used). When these are combined into a cut-scene it can be especially powerful, as the result is often closer to a polished movie than the regular actions see during game-play. However, these methods are not always successful. For example, a character may have his or her back turned to the camera or be too far away for a player to accurately recognize a facial animation. Music can be great at conveying an overall emotion for a given scene, it usually cannot convey the different emotions multiple characters may be experiencing in the same scene. As for dialogue, not all characters engage in dialogue or have voice actors. We know from reading email that text does not always accurately convey the emotion that the writer intended.

There has been a strong effort in recent years to develop more efficient methods of creating complex character behaviour that does not appear immediately identical when shared among characters. For example, Zhao and Szafron have used Hidden Markov Models to generate dynamic behaviour [89] and Kraayenbrink *et al.* have conducted research on behaviour in crowds [47]. However, when it comes to emotion, most effort has been focused on facial animations during conversations. This means that, when talking to a character you may be able to determine that they are angry, but once the conversation is over they return to appearing neutral, a noticeable disconnect.

Georgios Yannakakis posited that there are four big areas of research within video game AI: player experience modelling, procedural content generation, data mining of game information, and improving non-player character behaviours and capabilities [86]. This research falls in the final category, specifically focused on enabling background virtual characters to display emotional cues using their non-verbal non-facial behaviours that can be recognized by the player.

There were two specific goals for this research. The first goal was to develop techniques that were able to produce a minimum accuracy of 70% in participant identification of the emotions. To achieve this goal, we created two techniques (emotional gaits and emotional incidents) that are further detailed in Chapter 2. The combination of these two techniques was able to create recognizable emotions that participants were able to accurately identify (88.6%). We hypothesized that these emotional characters would produce more believable characters.

The second goal was to demonstrate that game characters using two kinds of emotional cues are significantly more believable (at a 95% confidence level) than similar characters without emotional cues in a game setting. In Chapter 8, we present an affirmative result ($p = 0.03$) for a small interactive story-based game.

The rest of this chapter details the eight contributions that are found in this dissertation. Chapter 2 provides a more detailed problem formulation and explains our two techniques (emotional gaits and emotional incidents). Chapter 3 provides an overview of the related work in this area. Chapters 4-9 detail the various user studies we ran to validate our techniques and investigate their limitations. Finally, Chapter 10 provides a summary of the work along with suggestions for future work.

1.1 Contributions

The research presented in this dissertation contains eight contributions. They are discussed in more detail in the following chapters.

Contribution 1. *Commercial game characters do not have the flexibility needed in animations to create relatable believable everyday characters needed for many serious games.*

This research originally began as an attempt to support the creation of serious games to help patients suffering from chronic depression. More motivation and information about this contribution can be found in Appendix A, which contains a short paper we wrote on the topic [22]. That research goal was transformed into more basic research after it was discovered that the biggest hurdle in developing the game was the lack of believable *normal*[1.a] characters in video games.

Contribution 2. *Players can recognize non-verbal, non-facial emotion in game scenes.*

The first step of the research was to discover methods to convey identifiable emotion that didn't require facial animations or dialogue/voice-acting. Our solution was to use a combination of two techniques: emotional gaits and emotional incidents. Initial studies examined whether or not participants were able to accurately identify emotions from the clues provided by these two techniques (Chapter 4). Later studies examined putting these characters into a more game-like setting (Chapter 7, 8).

Contribution 3. *Gender does play a role, but what the role is needs more study.*

Based on the set-up of our original study (Chapter 4), it was clear that extra work was needed in order to determine whether emotion identification success depends on player or character gender.[1.b] The results were analyzed in terms of participant gender as well as character gender. While there were some significant results in the first study (Chapter 5), those results were not confirmed in a later study (Chapter 8). These results indicate that while gender *likely* play a role, we are unable to accurately conclude at this time what that role is.

Contribution 4. *Participant filtering is required.*

The studies we ran were largely conducted with student participants from the University of Alberta Psychology 104/105 classes. In return for participating, students received course credit. However, while interacting with the participants, it became clear that not all were interested in participating fully; some were more interested in finishing as quickly as possible. Based on this, it became important to include a strategy to detect participants who were not fully engaged, allowing for their results (the outliers) to be removed from experiments (Chapters 7 and 8).

[1.a]By normal, we are referring to characters that look like everyday people that could take place in the here and now. No fantasy outfits or extremely sculpted/muscled bodies.

[1.b]In this dissertation, the word gender refers to each participant's choice to identify as male or female.

Contribution 5. *Emotional gaits and emotional incidents do not allow for participants to recognize a character is changing emotions because the cues happen over a period of time.*

While this research focuses on recognizing emotion as displayed by virtual characters, it was not designed for recognizing *changing* emotions. In order to recognize changing emotions, it is important for participants to be able to anticipate the change and then see key information that the change is occurring. We ran one experiment (Chapter 7) based on changing emotions and confirmed that while our strategy is successful at conveying static emotions, it did not provide the anticipation and timely key reaction cues needed to signal to participants that a character was changing their emotional state (Chapter 7).

Contribution 6. *Emotional characters are more believable.*

After placing a mix of emotional and neutral characters into a game, we found that participants who spent a larger percentage of their time interacting with emotional characters found those characters to be more believable than those who spent a larger percentage of their time interacting with characters who did not display emotional cues (Chapter 8).

Contribution 7. *Players can learn to quickly recognize emotion in games.*

In the Guess Who study (Chapter 8), participants played two short games. They were then tasked with identifying the emotional state of a mix of neutral and emotional characters. Even though the participants had spent very little time playing the previous games, and had only been specifically tasked with identifying at most two different emotions in the earlier games, they were able to to accurately identify 70% of the characters' emotions (Chapter 8).

Contribution 8. *Adding emotional gaits and incidents is relatively inexpensive.*

Adding more animations and code to a game can seem overwhelming, especially when games are on tight budgets and deadlines. However, it was shown in the Guess Who experiment (Chapter 8), that only a few gaits and animations were needed to create a wide variety of emotional incidents and successfully convey emotion to players. A final experiment was run using Project Spark [54], which showed that it is possible to replicate almost the entire initial experiment (minus the gaits) within a short amount of time (Chapter 9), using a tool that did not require artistic or programming expertise.

Chapter 2

Creating Recognizable Emotions

Our main goal was to create believable characters that convey recognizable emotion through non-verbal and non-facial cues. But what *is* a believable character? And, how do we create them? In a paper on the role of emotion in believable agents, Joseph Bates [8] starts by stating:

> There is a notion in the Arts of "believable character." It does not mean an honest or reliable character, but one that provides the illusion of life, and thus permits the audience's suspension of disbelief.

How to define *what* counts as the illusion of life is tricky and is where believability often becomes muddled as people aim for slightly different goals. Fabien Tenc *et al.* [75] asks, "does a believable character have to give the illusion of life or have to give the illusion that they are controlled by a player?" With many video games existing in completely fictional worlds of magic and fantasy, the question of believability can become even more confusing, because how do we *know* what *is* or *is not* believable? Mark Riedl and R. Michael Young [69] created the following definition:

> Character believability refers to the numerous elements that allow a character to achieve the 'illusion of life', including but not limited to personality, emotion, intentionality, and physiology and physiological movement.

When creating believable characters, especially with many of the more open-world style games that exist today, it is important that these characters remain believable over time. It is common for character-character interactions to exist for hours and days (as represented by game time). As Andrew Ortony [62] stated:

> Believability entails not only that emotions, motivations, and actions fit together in a meaningful and intelligible way at the local (moment-to-moment) level, but also that they cohere at a more global level – across different kinds of situations, and over quite long time periods.

When it comes to creating believable characters, researchers have focused on many aspects, including crowd control [47], daily schedules for characters [90] and alibi generation [74].

We were focused on creating recognizable emotion, expressed through character movement and interactions with the game world, to create our believable characters. Animators have long used non-verbal behaviour to convey emotion. This can be seen in any Disney or Pixar animated movie. The Pixar short film *Luxo Jr.* [66] , about a young lamp's interaction with a rubber ball, is able to convey the lamp's joy and sadness even though the lamp does not have a face or the ability to speak. In the stop-motion short *Conflict* by Gary Bardin [37] from 1983, we start to feel for a collection of matches, even though they are faceless objects, because they appear to have feelings and emotions. Ken Perlin has also done research in expressing emotion. One example of his results is the simple triangular prism Polly in *Polly's World* [65]. Using this simple construction, and a circular path, Polly can answer "yes" or "no", walk dejectedly, swagger and more. Polly appears to have emotion, even without a face. But, as Perlin has pointed out, a believable character is not the same thing as realism [45].

All of these examples are able to get us to feel for the characters, even though none of them have faces, or even human bodies. They are just a few examples that show that highly expensive facial animations are not required to create believable emotional characters.

Emotion in video games has been most commonly presented through dialogue (voiced or written). It is becoming more common, with games such as the Mass Effect series, to use animations to express non-verbal emotional cues through body language during cutscenes. The non-verbal expressions of emotion are most commonly shown through facial animations.

The work in this dissertation focused on adding non-verbal non-facial emotion to characters as one method to increasing character believability. Joseph Bates states "Emotion is one of the primary means to achieve this believability, this illusion of life, because it helps us know that characters really care about what happens in the world, that they truly have desires" [8]. In the *Illusion of Life*, Thomas and Johnston [77] specified three requirements for successfully expressing emotion. First, that the emotional state should be clearly understood. Second, that emotions may need to be exaggerated in order to be understood. Third, that emotions that influence a character's decision process should be clear from the character's actions. While we avoided gross exaggeration of emotional cues, we slightly exaggerated some emotions so that players could clearly pick up on the emotional cues (e.g., the afraid gait).

Tinwell *et al.* found that it was very important that behavioural fidelity matches a character's human-like appearance [78]. The game world and characters we used for our study were designed to look like they could be taking place *right now.* There is no fantasy or science-fiction element to our world. Knowing this and because our research was motivated by deficiencies in believable character behaviours for serious games, we were very interested in not only making believable characters, but in making *realistic* characters. We wanted our characters to be identifiable and relatable to everyday people, performing actions and reactions that could be seen in everyday life. The second requirement from the *Illusion of Life* was that believable characters may need to exaggerate. We were hoping to avoid this as much as possible, as exaggeration is often not realistic (eyes do not *actually* pop out of one's head when surprised). While we resisted exaggeration as much as possible, there were situations where we were forced to exaggerate in order to make the emotions easily identifiable.

2.1 Constraints

Because this research was focused on all background virtual characters in video games (including characters that were incidental to the plot), there were some important constraints on creating viable techniques for displaying recognizable emotions for our particular research.

Constraint 1. *Do not assume the player will see any specific part of a character (such as the face).*

This meant that it was important that the techniques did not rely on using a single specific area of the body, such as the face. We had to assume that what we were designing could be viewed from many angles. For example, the face could be too small to identify features or turned away from the player character or that part of the body may be blocked from view. With this constraint, we decided to focus on techniques that used the entire body of the character, giving us a larger canvas.

Constraint 2. *Cannot assume the player will interact with, talk to or overhear the character.*

Removing the necessity of talking to or overhearing the character removed the expensive requirement of creating dialogue for the individual characters. Creating individual dialogue lines, and potentially using voice actors, adds a lot of overhead and cost to developing a background character, especially when the dialogue must be created to convey emotion. However, removing direct interaction with the player meant the emotion-conveying techniques needed to work when a player was only able to view the character from a distance.

Constraint 3. *Allow the character to interact with the game world.*

This constraint was to provide characters with objects that they could interact with and react to. Since direct interaction with the player had been removed, there needed to be another basic mechanism that could occur to convey emotion.

Constraint 4. *Try and keep reactions realistic by avoiding exaggeration as much as possible.*

The goal was to create emotional clues for the player that were as realistic as possible, knowing that some exaggeration would be necessary. For example, in order to show fear a startle response would be necessary. Because a small startle may not be noticeable, the character may need to have a stronger reaction than a regular person might. However, the goal was not to get to a point where the character would perform movements that were not humanly possible (such as eyes popping out of the head) or cartoonish.

2.2 Problem

The overall problem is low believability of background characters in games. Our solution was to create a technique that adds non-facial, non-verbal emotional cues that increase character believability. For this dissertation, we have decomposed the evaluation of our problem solution into three criteria:

1. Does our technique(s) produce easily recognized emotion?
2. Are the emotional characters that use this technique more believable than neutral characters?
3. Can we implement our technique using only a small number of animations and scripts so that the amount of extra effort required is small compared to the effort required to create existing character behaviours?

For the first evaluation, we chose to compare our technique's emotion recognition accuracy to the results obtained by Roether *et al.* [71] because we based our emotional gaits on theirs. Roether *et al.* were able to achieve an accuracy of 70.3% or higher for each of the four emotions we were exploring. Therefore, we will consider our techniques a success if they are able to exceed 70% accuracy.

The second evaluation was performed once we added our characters into a game. We created a game with a mix of emotional characters and neutral characters (they use the neutral gait and do not interact with the game world). We asked the participants to rate the believability of the characters. We tested the hypothesis that the participants who interacted with more emotional characters would find the characters significantly more believable than the participants who interacted with less emotional characters.

Finally, we knew that techniques that are too costly to implement (either in terms of actual dollars or person-hours) would not be feasible solutions. We therefore limited our techniques to those that minimized the amount of additional animations and code. This required us to limit our animations to those that shared animation components across characters and emotions.

2.3 Techniques

We discovered two different techniques that, together, passed the three evaluation criteria. The first technique focuses on the character gait while the second technique focuses on how characters interact with the game world. The resulting techniques produced non-verbal, non-facial behavioural cues based on a character's assigned emotion. The following sections detail the two solutions.

2.3.1 Emotional Gaits

For the first technique, the initial idea was to create different walking animations to reflect each of the four emotions we chose (happiness, sadness, anger and fear). These animations were based on the models developed by Roether *et al.* [71]. The gait animations varied by emotion on three separate axes. First, the walking speed varied across all the animations, with sadness the slowest and anger the fastest. Second, the upper body posture of the character varied. The angry character leaned forward slightly, the happy character was quite upright, the sad character was hunched forward and the startled character was leaning slightly back. Third, the arm/hand positioning differed. The angry character's hands were clenched in fists, the happy character's hands swung at his/her sides, the sad character's arms barely moved, and the startled character had arms slightly out to the sides and forward, with the hands open and fingers splayed. Examples of the four emotional gaits can be seen in Figure 2.1.

Figure 2.1: The four gait animations - happy, sad, angry, afraid.

As these animations were being created, it was determined that the characters would all need an idle animation that was specific to their emotion, so that they would visually appear to have a consistent emotional state even when stopped. At this point, it became clear that the emotional *gaits* were not strictly gaits, but gaits combined with body posturing.

A neutral gait animation was also created for testing purposes, resulting in five different gait animations: neutral, angry, happy, sad, and afraid.

2.3.2 Emotional Incidents

The second technique was designed to enable characters to appear more believable as it required the characters to be aware of and interact with the game world. Each day, the interactions we have with the world around us are influenced by our current emotional state. For example, when happy, we are more likely to smile, wave and interact with others. The goal with this second technique was to provide game-world interactions that characters could choose to participate in, such that players would ascribe emotions and motivations to these characters based on their interactions with the game world.

We called this second technique emotional incidents. Each emotional incident is made up of some combination of the three sub types: emotional paths, emotional props, and emotional characters.

Emotional Paths

Imagine you are walking down a sidewalk towards a destination that is a few blocks away. If you are in a good neighbourhood, it is unlikely that you will change your course, continuing to head straight to your destination. If it is a questionable neighbourhood late at night and you can see some people loitering ahead, you may decide to cross the street before passing the people, even though your final destination is on your current side. The path that you take in these scenarios is highly influenced by both the atmosphere and environment of the scenario, but also by your current emotional state. In the second scenario, the more scared you are, the more likely you will be to find an alternate path past the loitering people. On the other hand, if instead you are angry, you may not even notice the people, or you may be looking for a fight, and more likely to continue on your straight path.

Emotional paths encapsulate this idea, that a character's current emotional state will influence the path the character will take between two points. For example, the afraid characters will try to maintain a buffer of space between themselves and any obstacles, allowing them to see potential threats and prevent surprises. Figure 2.2 shows the path an afraid character might take to avoid getting too close to someone sitting on a bench. While happy characters may not be very influenced by obstacles in their path, they are more likely to delay following a path or change course if they see a character or prop they want to interact with. Table 2.1 provides a brief description of how a character's path may be influenced by each emotion.

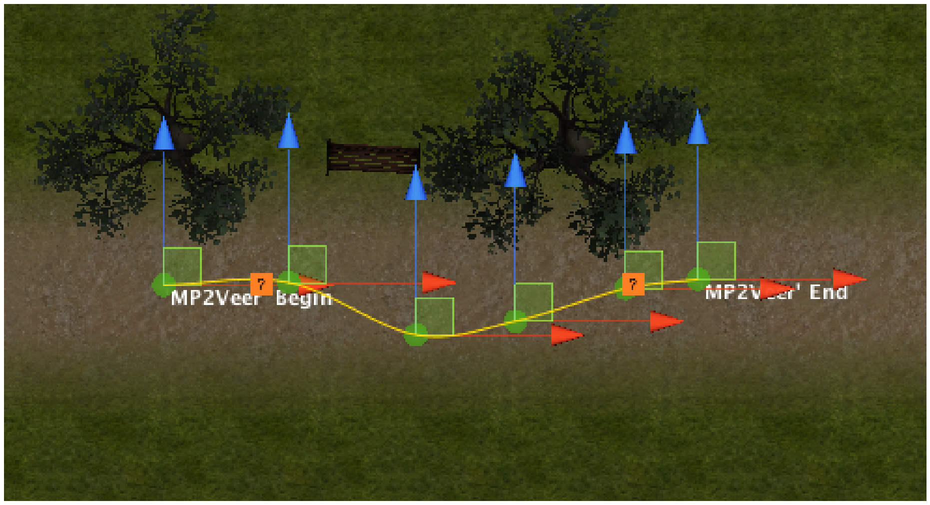

Figure 2.2: An example path for an afraid character for passing another character on a bench. The character will move away from the bench briefly in order to pass the second character before returning to the middle of the path.

Emotion	**Response**
Happy	Move to talk with another character
Angry	Direct path around obstacle
Sad	Ignore obstacles
Afraid	Buffer of space around obstacles

Table 2.1: Possible reactions for emotional path.

Emotional Props

Imagine you are walking down a sidewalk. What do you do if you notice a small rock in front of you? If you are a child, you may kick at it, whether you are happy or angry. If you are an adult, you are more likely to kick at it if you are angry or frustrated. A rock on a path is one example of an emotional prop that a character can interact with. Like emotional paths, interactions between a character and an emotional prop are influenced by the character's emotional state. Table 2.2 shows possible character reactions based on emotion.

Emotion	Response
Happy	Interact with prop
Angry	Move prop
Sad	Ignore prop
Afraid	Avoid prop

Table 2.2: Possible reactions for emotional prop.

Emotional Characters

You are still walking down the sidewalk, except this time, you see a friend on the other side of the street. What do you do? Here, the possible interactions become more complicated than those with emotional props and paths, because both characters involved in the interaction can have emotional states that influence the interaction. If both you and your friend are happy, you may wave at each other in acknowledgement, followed by moving towards each other in order to talk. However, what if you are happy and your friend is angry? In this case, you may still wave at your friend, only to find that your friend does not wave back, as she hasn't noticed you, too busy fuming as she storms along her path. Table 2.3 shows some example reactions based on another character's emotion.

Video games can also add additional levels of information that need to be taken into consideration when deciding on a correct emotional response. For example, it is common in role playing games for characters to belong to factions or groups. Characters from competing factions should probably not display a happy interaction towards each other, even if they are both currently happy.

Emotion	Response
Happy	Mimic positive gestures (wave if they wave)
Angry	Hostile acknowledgement
Sad	Ignore / no acknowledgement
Afraid	Avoid

Table 2.3: Possible reactions for emotional characters.

Summary

The three subtypes work together to produce more nuanced emotional incidents. For example, an afraid character may change paths in order to avoid interacting with another character, while a happy character may use a prop in an interaction with another happy character. Table 2.4 lists a series of props or character emotions that could trigger an emotional incident and the possible responses based on specific emotions.

Note that not all props have a corresponding emotional response for all four of the basic emotions. For example, the book only has a response for happy characters. If, instead, a sad character encountered the book, no emotional incident would trigger, and the book would be ignored. This list is not exhaustive of all emotional incidents that could be created, but was designed from the set of props we had and our test environment (a park). It is also important to note that this list is heavily weighted towards happy incidents. This is more indicative of our set of props and environment than anything else. Should our test scene have been in a back alley late at night, it is likely we would have had many examples of afraid incidents and very few happy incidents.

Our choice to use techniques such as avoidance to emphasize some emotions (e.g. afraid) is not meant to imply that this is the *only* reasonable response for afraid characters. It can be easily argued that a character who is afraid should be moving very quickly and *not* avoiding anything, as the character attempts to leave the frightening environment. This relates to personality, as characters will react differently to each emotion based on their individual personalities. However, to satisfy our constraint of using limited resources to create a single test environment, we focused on creating a single set of consistent reactions for each emotion.

2.4 Conclusion

This chapter introduced the two techniques that were designed and evaluated throughout the rest of the dissertation. These techniques, emotional gaits and emotional incidents, are used by characters in order to express visual cues about their emotional states. We have also presented the three main research questions this dissertation answers:

1. Does our technique(s) produce easily recognized emotion?
2. Are the emotional characters that use this technique more believable than neutral characters?
3. Can we implement our technique using only a small number of animations and scripts so that the amount of extra effort required is small compared to the effort required to create existing character behaviours?

Prop/Character	**Emotion**	**Reaction**
rock	happy	stop and kick
	angry	kick while passing
book	happy	pick up
cup	happy	pick up
newspaper	happy	pick up
	angry	kick on ground
garbage can	happy	throw out newspaper/cup
picnic table	happy	put down newspaper/cup/book
bench	happy	sit down
	sad	sit down if no one else is on bench
	afraid	sit down if no one else is on bench
two on Bench	both happy	sit and talk
	afraid & other	afraid character gets up and leaves
	sad & other	sad character waits a few seconds before leaving
two characters	both happy	wave, possibly stop and talk, possibly exchange book
	one happy	waves at other
	one sad	looks briefly at other, then ignores
	one afraid	moves away from other character
	one angry	speeds up and looks at other character while passing
ball	happy	stop, kick ball back to boys
	angry	kick ball away
	sad	ignore ball, keep walking
	afraid	pause and show startle animation

Table 2.4: The emotional incidents in the park.

Chapter 3

Related Work

This chapter details related work on emergent game play (Section 3.1), gender in games (Section 3.2), the benefits of playing games (Section 3.3) and emotions (Section 3.4),

3.1 Emergent Game Play

Emergent game play happens when interesting non-scripted encounters and reactions occur between objects and/or characters in a game. At the Games Developer Conference in 2004, Randy Smith and Harvey Smith gave a talk on emergent game play [72]. One example they gave of emergent game play is of a candle falling off a table into an oil puddle and the oil puddle igniting. This reaction is something we expect in real-life, but often not in games. Emergent game play involves setting up props such as the candle and oil puddle to be able to produce new actions/reactions. Here, the candle would be identified as a source for igniting, and an oil puddle as something that can be ignited. By identifying props throughout a game with labels/abilities, combinations of objects can interact without explicit designer intent.

In a sense, our final implementation of emotional incidents (detailed in Chapter 6), is very similar to emergent game play. But, instead of having props collide and create new interactions such as setting a puddle of oil on fire, in our system the *characters* collide with props and other characters to create new interactions. It works since characters know that they can *kick* when angry, and props such as rocks and garbage know that they *can* be kicked. Thus, when an angry character collides with a rock, this information surfaces to form an emotional incident response where the character *kicks* the rock.

Bailey and Katchabaw's framework for emergent psychosocial behaviour included a *stimulus-repsonse* system [6]. When something occurs (such as seeing an non-player character perform an action) a non-player character responds accordingly. These systems (stimulus-response) are often used to add emergent game play into games. Ion Storm used a form of stimulus-response to add emergent game play to their games, such as Thief [26] and Deus Ex [42]. Another notable example of a stimulus-response system is the Sims [31]. In the Sims, items advertise what needs they satisfy, and in return the Sims are attracted to certain items based on their current needs [33].

Paola Rizzo *et al.* were interested in creating personality-driven believable characters [70]. They approached personalities as "clusters of stable goals." Each believable agent has a cluster of goals that they are trying to achieve and a set of actions that they can do. The actions assist/achieve different goals and have different effects on the world. By setting up their environment this way, the believable agents are able to perform different sets of actions depending on their goal, allowing for varied behaviour. This structure also resembles the later stimulus-response systems mentioned above.

3.2 Gender

We were interested in making sure that our resulting techniques would be gender neutral. Gender and video games has a long history, but the number of women playing games is increasing. In 2005, the Entertainment Software Association (ESA) reported that 55% of gamers were male and 43% female (no explanation as to the missing 2%) [2]. By 2014, the ESA reported that females had risen to 48% of the market [3]. Even though the gap in genders has been steadily decreasing, many AAA titles are still geared towards males, both in the content (often seen in the lack of playable female characters) and marketing.

And yet, seeing themselves represented within a medium (games, books, etc.) has long been identified as a way that people connect with the medium. An examination back in 1998 by Tracey Dietz [24] noted that:

> The video games that are being played by today's youth present an overwhelmingly traditional and negative portrayal of women and that the development of gender identities and expectations among youngsters may be affected by these portrayals.

In 2002, a study by Berrin Beasley and Tracy Collins Stanley found in an analysis of 47 games that only one in four characters were female [9]. By 2006, when James Ivori [43] did an overview on gender representation within reviews of games, he found that little had changed: games still heavily favoured males. And he raised an interesting question:

> Though female video game characters appear to be underrepresented overall, as active characters, and as playable characters, they are proportionally more likely than males to be portrayed in a sexualized fashion. These results beg the question: If this is the nature of female video game characters, should their number be increased after all?

In 2006, Tilo Hartmann and Christoph Klimmt [40] studied the characteristics of video games that females dislike. They found the three main characteristics were the lack of meaningful interaction, the violent game play, and the gender stereotyping of characters. However, since this study was produced, it can be argued that there are now many games that start to address some of these concerns. Many recent games, such as Mass Effect 3 [15] and Skyrim [12] include much more complex and deeper story lines than earlier games, and players are often given a lot of choice as to *the type* of character they play and thus the actual game play. While there is still a significant amount of gender stereotyping in games [43], some game developers are acknowledging the problem and even including player input in design decisions [25]. One example is BioWare's involvement of the Mass Effect fan community when re-designing the female version of Commander Sheppard for Mass Effect 3 [25]. This was not a 'shinning' success, as it quickly devolved into more of a beauty contest than anything else, but at least they were acknowledging the problem.

3.3 Benefits of Game Play

Playing games has some positive benefits. For example, Jing Feng *et al.* used video games in a study on spatial cognition with non-gamers [32]. They found that while both genders improved, it almost eliminated the often noted differences between males and females. However, if games are only being promoted to one gender, the other gender (in this case females) often miss out on the benefits. Skills such as spatial cognition are widely used in mathematics and sciences, where females continue to trail in numbers [60].

Another study by Chandramallika Basak *et al.* examined the use of real-time strategy games as a way to slow cognitive decline in adults [7]. After 23.5 hours of training (done in 1.5 hour blocks) the participants showed an increase in executive control function (4 of 5 tests) and in one area of spatial control (mental rotation of objects).

Exergames, or games that require players to move (exercise) as part of the game play, are becoming more popular with the inclusion of motion controllers (like the Wiimote) and motion-capture devices (like the Kinect and Playstation Move). Yue Gao and Regan Mandryk found that a casual exergame can provide cognitive benefits after only 10 minutes of play [36].

More recently, with the growth of serious games and/or edutainment, many games that focus on training or teaching players are now being produced. These include games such as Foldit [19], where by playing a game, players are actually helping scientists make new conclusions and discoveries (e.g. new proteins in Foldit). Another example is Darfur is Dying [59], in which players experience life as a refugee.

3.4 Emotions

With regards to emotions, there is much disagreement among researchers about models, techniques and result implications [27, 63]. Many researchers have focused on trying to identify what are termed *basic* emotions. A basic emotion is an emotion that is consistent and identifiable across cultures [27]. However, which emotions should be included within this group is still under debate, as researchers range from including as few as four to as many as nine emotions [27, 63]. The six most commonly agreed upon emotions are happiness, surprise, fear, sadness, anger and disgust [27]. For this dissertation, we have focused on four emotions: happiness, sadness, fear and anger.

Paul Ekman, in 1992 [27], argued that there are nine characteristics used to identify the various emotion families (which generally correspond to the basic emotions):

1. distinctive universal signals;
2. presence in other primates;
3. distinctive physiology;
4. distinctive universals in antecedent events;
5. coherence among emotional response;

6. quick onset;
7. brief duration;
8. automatic appraisal;
9. unbidden occurrence.

The distinctive universal signals discussed by Paul Ekman include emblematic facial expressions such as the wink, facial conversational signals and hand gestures. These signals "are movements that have precise symbolic meaning and are socially learned and culturally variable" [28]. While there has been work done on identifying key components of facial expressions and hand gestures as they relate to emotion, we were unable to find a common set of emotional incidents that people may perform.

3.4.1 Emotions versus Moods versus Personality

Many of the characteristics identified by Paul Ekman fit with the generally accepted understanding in psychology that emotions consist of a brief response and that they are automatic [27]. Moods, on the other hand, last longer and are considered to be highly correlated with specific emotions (such as apprehensiveness with fear). Personality is generally considered "to be a coherent pattern of kinds of behaviours and interactions with the environment" [70].

These definitions beg the question: is our research actually on emotions or moods or personality? If we accept the premise that emotions are fleeting and moods are sustained, then our research is primarily about moods and moody behaviour rather than emotions. Even though we are interested in interactions with the environment, we are not delving into personality, as all our characters will react similarly based on emotions. Personality could be used to add another degree of variation to the reactions. In this dissertation we use the term *emotion*, rather than mood, as the word emotion will be more clearly understood by those within computing science and within the general population.

3.4.2 Emotion as represented through non-verbal cues

Examining whether emotion can be recognized from non-verbal cues can be traced back decades [83, 38], but it has most commonly been tested with human actors. Judith Hall [38] examined previous studies looking at gender effects when identifying emotion through non-verbal (visual and/or auditory) cues. She found that females statistically outperform males. However, she also found that there was little to no effect caused by the gender of the character being labeled, which means any effects found were on the side of the people doing the labelling.

Anthony Atkinson *et al.* showed that exaggerating whole-body movements allowed for increased recognition of emotions, as well as increasing the perceived emotional intensity [4]. This supported our decision to slightly exaggerate some of our emotional gaits when the gaits were not initially easily recognized. They also confirmed that basic emotions are readily recognized from human-body movements, which Harald Wallbott earlier proposed [82]. Wallet found that body movements and postures, to a degree, are specific for at least some emotions (e.g. crossing arms in front of chest for pride) and that body movement is indicative of the intensity of the emotion.

On the neural side, Beatrice de Gelder showed that emotional body-language is rapidly recognized in the brain, and, if available, compared against the perceived emotion of the facial expression and mismatches are noticed [21]. The emotional body-language can increase the intensity of a facial emotion (e.g. an angry person with clenched fists) and can also provide extra information (such as indicators as to where the anger is being directed). When it comes to emotional body-language, people use relatively simple perceptual cues to determine the complex motives and intentions of others. Raymond Mar *et al.* found that people have a stronger reaction to these cues as displayed by human actors than by virtual characters [49]. This further supports our decision to use exaggerated movements for some of our responses.

More recently, Magy Seif El-Nasr and Huaxin Wei [29] looked at creating non-verbal behaviour models. They had animators create an office scene containing two characters discussing one character's job search. The animators were asked to make a series of scenes. In each scene, each character was provided a short description, such as *to impress someone* or *to be thought normal*. They found that for some of the descriptions, the animators expressed them differently depending on which actor they were applying it to. This suggests that, for the animators, the character's identity or perceived importance affected how they thought that character would express a given state-of-mind. In our research, our characters all shared the same level of importance (they were all random people at the park). However, had our games taken place in a setting with more defined roles between characters, the emotional incidents and frequency would likely have needed to take into account the character's roles in addition to their assigned emotion.

In 2009, Claire Roether *et al.* [71] conducted a study in which they developed five different walking animations to represent four basic emotions (happy, sad, angry, and afraid) as well as neutral. These animations were placed on a simple human-shaped mannequin with no identifying features (see Figure 3.1). Part of the study was to attempt to isolate features that are important to emotion expression. Some of the ones they found included head inclination (typical for sadness) and bending at the elbow (associated with anger). They also changed the speed of the gait. One part of their user study had participants view animations and rate them as happy, sad, angry or afraid (no option for none), and their accuracy results for the four emotions were 75.1% for happy, 89.8% for sad, 70.3% for angry and 77.1% for afraid. The results of their user study showed that people could generally identify emotion based on gait.

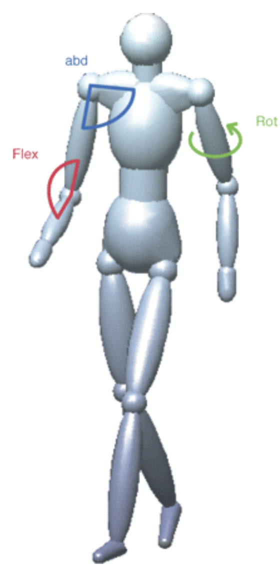

Figure 3.1: Model used by Roether *et al.* for examining emotional gaits. Image from [71].

According to Marco Pasch and Ronald Poppe, the use of the simple mannequin should not greatly influence the perception of emotion [64]. They did, however, note that some emotions appeared to be better perceived from more abstract forms than others. However, as the highest accuracy rate they achieved from any of their examples was 62% (with 4 of the 6 emotions producing a top result of less than 40%), these results may change with more accurate representations of the emotions. Rachel McDonnell *et al.* applied motion capture data to a series of virtual characters (low and high resolution humans, cartoon, mannequin and zombie) [53]. They also found that the overall ability of participants to identify the character's emotion was mostly independent from the character's form.

Joann Montepare *et al.* also ran a user study trying to identify what specific characteristics of gait help with emotion identification [58]. Some of the characteristics they looked at were arm swings, stride length, and standing up straight. However, because they recorded five actors walking from a top down view, the resulting videos made it very difficult to find effects for posture. They believe this is due to the inability to detect head angle in their videos.

Kenji Amaya *et al.* ran a study looking at creating emotional movements by affecting speed and spatial amplitude of movements [1]. They point out that humans are very sensitive to the motion of others, but only in the sense that humans "can easily detect erroneous movement ("it simply doesn't look right"), although [humans] often find it much more difficult to isolate the factor which causes the movement to look incorrect." They used motion capture data to try and identify important features of emotion and focused on speed and spatial amplitude. They then ran a second study applying those features (changes to speed and amplitude) to a neutral animation. Their results matched closely to the original motion capture data. However, because the type of changes in speed and amplitude were *fitted* to motion captured data, there is the question of how accurate their results would be if they didn't have the initial motion capture data to work with.

3.4.3 Theories of Emotion

There are multiple theories of emotion that have been proposed over the years. There have been four main types of emotional models (appraisal, dimensional, anatomical and rational) that have been extended and combined over the years (Figure 3.2). Two of the more common models are the OCC Theory of Emotion and EMA. These two are especially common when adding emotion to characters or agents, as they can be easily implemented as computer models.

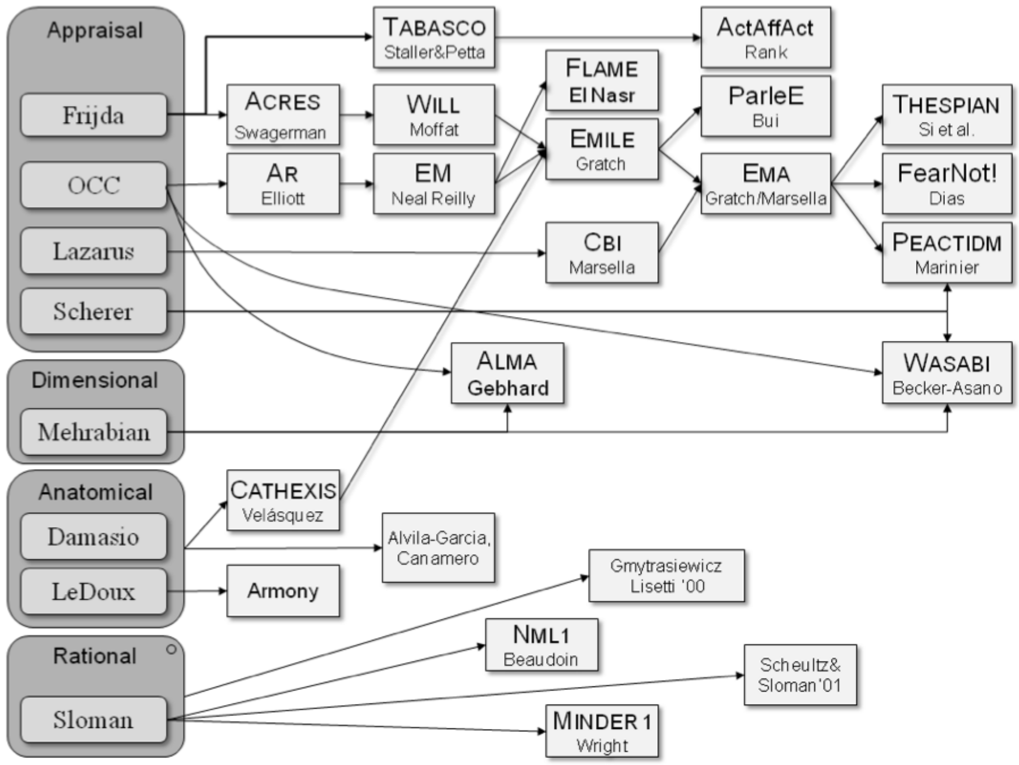

Figure 3.2: Computational models of emotions and many of the extensions. Image from [51].

OCC

The OCC theory was designed specifically for computer implementation by Andrew Ortony, Gerald Clore and Allan Collins [61]. It is an appraisal theory where emotions are valence (good or bad) reactions to an event. A character appraises an event, another agent, or object in terms of goals, norms (the standards the character expects others to adhere to), and tastes (the character's own attitudes or feelings on an event) and uses this information to determine the corresponding emotion. For example, a character appraising a disliked object, would generate a hate-based response, which would be translated into an actual emotion based on their current emotional state.

The original theory contained 22 types of emotions, but later the authors presented a way to collapse them into five positive and five negative specializations based on two types of affect reactions (generalized good or bad feelings) [62]. These reactions can be further simplified into goal-based, standards-based and taste-based. Emotions can result from a specific specialization being achieved (joy because there is the possibility of something good happening, which is goal-based) or from a combination (anger because something bad might happen that is against the standards/norms of a character, which is combination of standards and goals).

Andrew Ortony suggests using personality of characters as a way to ensure consistency and coherence when expressing a given emotion [62]. This could also be used as a mechanism to enable characters with different personalities to react differently when expressing the same emotion. Someone with a more constrained personality is less likely to be seen jumping for joy compared to someone who is more outgoing.

OCC has been used and expanded upon by others, such as in FLAME (Fuzzy Logic Adaptive Model of Emotions) by Magy Seif El-Nasr *et al.* [30] and by Christina Conati and Xiaoming Zhou to track students' emotion in educational games [18].

EMA

In EMA (EMotion and Adaptation), emotions influence how a virtual character judges events and the resulting actions [50]. EMA appraises the situation and then maps from a set of domain-independent features to individual appraisal variables. The resulting set of appraisal variables is combined to determine an emotional state that influences the character's behaviour. The EMA framework can be used to implement a system where characters can respond to sudden changes in emotional state. For example, EMA can be used to determine how a character should react if the bench they are sitting on suddenly catches on fire. This can be very important in order to create noticeable emotional change between characters, because, as Joseph Bates noted, these changes require broadcasting in advance that something is going to occur (allowing for anticipation), so that viewers can prepare to see it [8]. Otherwise, with realistic timing, events can happen too quickly for players to notice.

João Dias, Ana Paiva, Sandy Louchart *et al.* created *FearNot!*, a game in which the player is a spectator in regards to a bullying incident, and then is consulted on what the bullied character should do next [48, 23]. They included an emotion model they called FAtiMA, which is based on EMA. It uses an affective appraisal system and incorporates emotion into the reasoning process. It also includes a planner (allowing the character to act intelligently while being able to replan around unexpected events).

Our solution for creating recognizable emotional characters, as detailed in this dissertation, is focused on characters maintaining a sustained emotional state. They do not acknowledge or react to events and therefore do not need to appraise them in order to assess how they should change their emotional state and reactions. All of the emotion responses we have developed are designed to show a character's *current* emotional state. Our research is focused on whether players can identify virtual character emotion in the context of an active game and whether this identified emotion enhances believability. Recognizing changes in emotion is an important topic of future study and would require an emotional model. Since our research is about displaying emotion, not its evolution, our research is independent of the emotion model.

3.4.4 Emotion as seen in Animation and Games

When games attempt to apply non-verbal emotion to background characters, the results are often quite powerful. In Call of Duty Modern Warfare 2 [84], there is a mission called *No Russian* where the players must storm through an airport in Russia and kill all citizens they encounter. The objective of the mission was considered offensive to many, but it was the animations applied to the citizens that made it painful for some to even view. When the citizens try to run away, they perform a half-crouched run that is used to emphasize their fear. A citizen who has been shot (but not fatally) may perform a painful looking half crawl/half limp. There are multiple videos posted to Youtube of gameplay during this mission (a good example was posted by xCheezbrgr [85] starting at about 1:40). However, as powerful as the animations are, with more careful watching of the video, it becomes clear that there are only a few animations available to the citizens, and that they often perform a chosen animation as a group (such as the crouched run), not individually.

Dragon Age Origins [13] uses a toolset that includes the ability to choose and specify an emotion for each dialogue line in a conversation. The emotion assigned to a dialogue line is used to determine the head and facial expressions for a character, with some variation. However, adding emotion to a character only applies during the conversation/cutscene and does not translate over to regular game play behaviour.

In Skyrim [12], characters do perform non-verbal behaviour actions during conversations and cutscenes. For example, a character may cross his arms in response to a dialogue line.

However, these actions appear to be limited and not used very often. Characters are much more likely to turn their head to ‘look at’ the current character speaking than to perform a more complex action.

L.A. Noire [34] is probably one of the most emotion-laden commercial games. As the story involves investigating a series of cases as a LAPD detective, it is integral to the story line to be able to pick up on emotional cues. In order to add a higher level of emotion to the gameplay, most of the characters in the game were animated using a combination of advanced facial animation and motion capture from actual actors. While this worked extremely well in creating believable emotional characters, the techniques used are very expensive and not feasible for the potentially hundreds of background characters in a game.

Marjorie Zielke *et al.* have been building a First-Person Cultural Trainer to be used to train military members on how to deal with cultural differences [92, 91]. The current game focuses on Iraq and Afghanistan. The game can create a wide variety of characters consistent with the culture and uses non-verbal communication via posture and expression. Emotion is also conveyed via spoken dialogue. The characters use a traffic/errand system which models what the characters should do over the course of a normal day. If you follow a character they will appear to shop, eat, perform chores and go home to sleep. However, it is not clear how expressive the characters are while completing these tasks. Do they perform them differently depending on their emotional state? And if so, can these differences be observed by the players?

How important is emotion in video games? The answer is largely dependent on the actual game. Many commercial games that would likely benefit from emotion (such as Skyrim) have been extremely successful without it. However, other commercial games, like L.A. Noire, would likely *not* have been as successful as they have been without emotion. The current process for adding emotion to many characters in a game is often expensive and time consuming or results in multiple characters performing the same action as a group. These problems inhibit game designers from including believable emotion in large game worlds. However, many games that fall into the serious games or training games genres depend heavily on players identifying and learning from cues given by the on-screen characters. For these games, the potential costs (inability to connect with or teach their audience) of *not* adding emotion in their games is much higher.

Chapter 4

Identifying Emotions

Once we had decided on emotional gaits and emotional incidents as the two techniques that we would use to convey emotional cues, it was important to determine how well the techniques worked. We created an experiment to evaluate the techniques on their own as well as combined together.

We were interested in answering three questions:

1. Can participants identify each emotion with a minimum of 70% accuracy?
2. Do our emotional incidents convey accurate representations of the emotions they are supposed to represent?
3. Is one technique (emotional gaits or emotional incidents) better at conveying emotional cues?

In the studies by Roether *et al.* [71], they were able to achieve accuracy rates for the four emotions ranging from 70.3% to 89.8%. We designed our emotional gaits based on their final set and therefore our definition of success was to achieve a minimum value of 70% accuracy with our techniques.

There are no lists of specific incidents that are known to represent specific emotions. Because of this, when creating our emotional incidents, we had to develop our own list based on the objects and characters in our game and through discussion with others. With question two, we are interested in trying to confirm that our emotional incidents are valid representations of the emotions they represent.

Finally, while we expected the combination technique (gaits plus incidents) to produce the best results, we were interested in confirming whether or not this was true or if a single technique could achieve similar or better results.

In order to answer these questions, we created a series of scenes (video clips of in-game screens) to showcase the gaits, incidents and combined gait/incident pairing and recruited participants to watch and label each scene with the emotion perceived.

4.1 Experimental Design

To test our two techniques (emotional gaits and emotional incidents) we designed an experiment to test player recall and precision in identifying the correct emotion. For the experiment, we tested four emotions (happiness, sadness, anger, and fear) and a neutral response. This resulted in five gait animations (one for each emotion and a neutral one). We also picked two emotional incidents. The first emotional incident involves the test character passing a second character who is sitting on a bench and waving. The second emotional incident has a small child kicking a soccer ball towards the test character (Figure 4.1). The emotion-specific responses to the emotional incidents are listed in Table 4.1.

Figure 4.1: Screenshot from a scenario in the experiment.

Since one goal was to investigate how well the two techniques work independently and when combined, we designed the experiment such that players would see the techniques on their own (either just the gait or just the emotional incidents) as well as together. This resulted in 13 different scenes for the participant to watch, as detailed in Table 4.2. The

	Character on bench	Kid with ball
Happy	waves looks at character	kick ball to kid
Sad	brief glance, then ignore	pause, look at ball, walk straight
Angry	head turns to face other character, speeds up to pass	kick ball away
Afraid	slight startle, veers away from bench	startle at ball, then walk around it

Table 4.1: Emotional incident responses

first five scenes consist of only the gait animations and no emotional incidents (similar to the experiment done by Roether *et al.* [71]) and tests whether or not participants could identify emotion from gait alone. The next four scenes (6 - 9) had the character use the neutral walk but perform emotion specific emotional incidents and are used to test whether players could identify emotion based on the emotional incident response. The final four scenes combined the appropriate emotional gait with the emotional incidents.

Scene	Gait	Incidents	Emotion for Incidents
1	Neutral	No	-
2	Happy	No	-
3	Sad	No	-
4	Afraid	No	-
5	Angry	No	-
6	Neutral	Yes	Happy
7	Neutral	Yes	Sad
8	Neutral	Yes	Afraid
9	Neutral	Yes	Angry
10	Happy	Yes	Happy
11	Sad	Yes	Sad
12	Afraid	Yes	Afraid
13	Angry	Yes	Angry

Table 4.2: The thirteen scenes.

The 13 scenes were presented in a random order to each participant in order to reduce any bias. Participants viewed each scene in its entirety before being presented with the label options. We had two orders for the labels ("happy, sad, angry, afraid, none of these" and "angry, afraid, happy, sad, none of these") to prevent participant bias towards a specific response, and each participant was randomly assigned one of them at the start. When presented with the label options, participants also had the chance to re-watch the scene, although this option was rarely used (2.6% of the time). Once a label was chosen and submitted, the participants were unable to change that decision.

Participants were informed at the start of the experiment that there were no facial animations and that they could use a label as many times as they wanted. The participants did not know there were three types of scenes and given the fact that they had 13 scenes and five labels, they could not predetermine how many times each label was supposed to be used. After labelling the thirteen scenes we asked for basic demographic information as well as any comments on the experiment (see Section 4.5).

In the development of the experiment, we discovered a couple of our emotional-incident scenes contained emotional cues that we had not clearly defined as being a part of the gait or as part of the emotional incident. The first was the head tilt of our sad characters. We decided that the head tilting down (to look towards the ground) should be considered part of the gait. This meant we needed to make sure that when the sad character was performing the neutral gait (in the incident-only scene), that their head was up, until they were specifically involved in an emotional incident (at which point the head could be tilted down). The second change was the hand/arm positioning of the afraid characters. When the afraid character performs the neutral gait and becomes involved in an incident, there is no immediate cue to the viewer that an emotional incident has started. We decided to have the afraid character briefly show the hand/arm positioning of the afraid gait at the start of the emotional incident (simulating the character having a short startle reaction).

4.1.1 Participants

There were ninety participants, all undergraduate students taking a first year psychology class in the Spring 2012 semester at the University of Alberta. This group was made up of 63 females and 27 males and had an average age of 20.7 (mode 18, range 17 - 44) and average year of study of 2.1 (mode 1, range 1 - 5). In return for participating, they received 2% credit in their psychology course. Forty-four percent self-identified as gamers (play video games at least once a week).

4.1.2 Experiment Limitations

The results from this experiment must be placed into the context of the limitations of the experiment. Our participant pool is very homogenous in age, education and likely cultural background and therefore cannot be extended to the general population. Also, the majority of our participants do not play video games on a regular basis and are younger than the average gamer. These facts suggest that our conclusions should not be blindly applied to the general game-playing community.

4.2 Results

4.2.1 Confusion Matrices

In order to analyze our results, we created three confusion matrices: gait only, incidents only, and combination (gaits plus incidents). Confusion matrices are a statistical tool used to evaluate classifiers [79]. They measure recall, precision and accuracy. In machine learning, a classifier is a product of machine learning; however in our experiment the participants are the classifiers. The three confusion matrices are presented in Tables 4.3, 4.4 and 4.5.

Each row in the confusion matrix represents a particular scene, and the row label (happy, sad, angry, afraid) indicates the emotion that the character was meant to portray in the scene. The five columns (happy, sad, angry, afraid, or none) are the labels a participant could use to classify a scene. *RSum* is the sum of the row: the number of results for that scene. *Recall* is the ratio of how many participants correctly identified the scene divided by the RSum. The number of correctly identified emotions are on the diagonal, where the row and column labels match, and are shown in bold. *PSum* is the sum of a column: the number of times participants used that label over all scenes. Finally, *Precision* is the ratio of how often participants used a label correctly divided by the PSum.

For example, in Table 4.3, for the row Happy, 42 participants correctly identified the scene as happy. However, five identified it as angry, one as afraid and 42 as neutral. On the other hand, the column Happy shows how many participants over all four scenes labeled any scene as happy. Once again, there were the 42 who used the Happy label on the correct scene. However, there was also one participant who identified the sad scene as happy, four participants who identified the angry scene as happy and one participant who identified the afraid scene as happy. Finally, the recall value for the happy scene was 0.467 and the precision was 0.875.

The overall precision and recall values for the table are in the bottom right corner. Overall recall is the total number of correctly identified emotions by all participants divided by the total number of participant/scene combinations (90×4). Overall precision is the total number of correctly identified emotions by all participants divided by the total number of participant/scene combinations not identified as neutral. In the case of Happy, this is $42/(42+1+4+1) = 0.875$.

The confusion matrices show that, generally, precision was quite high for each emotion in each matrix (the minimum precision was 0.73). This means that when a participant identified an emotion (non-neutral), there was at least a 73% chance that they identified the intended emotion. However, the recall values varied much more, even within individual confusion matrices.

4.3 Overall Results

4.3.1 Gait

The results of the gait animations were surprising as most were below the results obtained by Roether *et al.* [71] where they achieved a minimum accuracy of 70.3, which we expected to exceed. Because we had observed the animations so frequently during their design and the implementation of the experiment, the identifying characteristics appeared obvious. It became clear as the results were computed that these differences were not as immediately noticeable when viewed by an untrained observer (the participants).

Gait	Happy	Sad	Angry	Afraid	None	RSum	Recall
Happy	**42**	0	5	1	42	90	0.467
Sad	1	**56**	6	12	15	90	0.622
Angry	4	1	**46**	1	38	90	0.511
Afraid	1	0	0	**79**	10	90	0.878
Neutral	15	5	2	5	**63**	90	0.700
PSum	48	57	57	93	105		
Precision	0.875	0.982	0.807	0.849		**0.875**	**0.619**

Table 4.3: Gait confusion matrix. The neutral results are not included in the PSum, individual precision, overall precision and overall recall calculations, to remain consistent and comparable to the incident and combination tables.

In all cases, the recall and precision were well above chance (20%) values. The precision results were good, with the lowest being 0.807 for Angry. The recall values were more interesting, with the range from 0.467 to 0.878. From these results, it appears that the emotional cues for some of the emotional gait were much more apparent than others. The afraid gait was surprisingly good at conveying the emotion (much more so than the afraid incidents, and equivalent to the combination of gait and incident). However, the participants appeared to be confused when it came to the happy gait, labelling it just as frequently as neutral as they did as happy; they were much more accurate in correctly identifying the neutral gait (0.700 versus 0.467).

4.3.2 Incident

Incident	Happy	Sad	Angry	Afraid	None	RSum	Recall
Happy	**72**	1	1	1	15	90	0.800
Sad	0	**65**	2	2	21	90	0.722
Angry	7	5	**47**	2	29	90	0.522
Afraid	2	11	16	**32**	29	90	0.356
PSum	81	82	66	37	94		
Precision	0.889	0.793	0.712	0.865		**0.812**	**0.600**

Table 4.4: Incident confusion matrix.

The results from the incident scenes were markedly different from the gait scenes. On one hand, the happy cues in the incident were understood much more clearly than the happy gait as seen in the recall results: 0.800 versus 0.467. However, on the other hand, the afraid incident cues, while surpassing chance (20%), were by far the worst recall rates we saw (between all techniques) at 0.356. Looking at the frequency and variety of labels used for the afraid scene, it was clear that many of the participants were just confused.

These results gave us mixed feedback on our emotional incidents. The results for both happy and sad exceeded the 70% accuracy we were aiming for, suggesting that they did a reasonable job of conveying emotional cues for happy and sad. However, our angry and afraid results were below 70% accuracy (52.2% and 35.6% respectively). This suggests that the emotional incidents we chose were not strongly conveying emotional cues for angry and afraid.

4.3.3 Combination - Gait plus Incident

The combination of both techniques increased the precision and recall values for almost every emotion. The anecdotes shared after the experiment (Section 4.5) helped us understand these results. According to the anecdotes, each participant was observing and using different emotional cues to determine what they believed to be the correct label. If the main identifying cue they used was based on the gait, then either the gait only or combination (gait plus incident) scenes were easiest to identify.

Combination	Happy	Sad	Angry	Afraid	None	RSum	Recall
Happy	**83**	0	1	2	4	90	0.922
Sad	0	**79**	4	5	2	90	0.878
Angry	5	1	**78**	2	4	90	0.867
Afraid	1	0	2	**79**	8	90	0.878
PSum	89	80	85	88	18		
Precision	0.933	0.988	0.918	0.898		**0.933**	**0.886**

Table 4.5: Combination (gait plus incident) confusion matrix.

Interestingly, fear is the only emotion that did not benefit from the combined techniques. Instead, the recall value for fear is the same as the recall value when using only the gait. This suggests that even when the combined technique was used, the participants were relying primarily on the gait information in order to label fear.

The combination technique was the only technique that was able to exceed 70% accuracy for all emotions, leading to the only technique that was successful overall in achieving our threshold for success.

4.3.4 Comparison of Techniques

Bootstrapping [80] is a method for resampling data that produces new data sets. It creates a new dataset by randomly choosing from the original set with replacement. In our case, we had 90 rows of data, each representing a unique individual. Using bootstrapping, we created 1000 new datasets of equal size. Each row in the 'new' datasets was a row from the original dataset. The datasets were created by using sampling with replacement, which means that a new dataset may contain multiple copies of an individual row from the original.

We ran an ANOVA to test whether there were significant interaction effects on the results by the emotion and techniques. The summary is in Table 4.6. The table shows that while both emotion and technique influenced the results, there was an additional effect caused by the combination of emotion and technique. These results can be seen in the earlier confusion matrices, as participants were better able to identify some emotions (e.g. afraid) by a single technique (e.g. gait) than by others.

The individual results were then compared for statistical significance at 95% confidence ($\alpha = 0.05$). The p-values are found in Table 4.7 The p-values were calculated following the techniques detailed by Hardin and Shumway [39] for comparing confusion matrices using bootstrapping.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Emotion	4	1.89	0.471	2.644	**0.0323**
Technique	2	18.40	9.201	51.602	**<2e-16**
Emotion:Technique	6	18.31	3.052	17.114	**<2e-16**
Residuals	1157	206.30	0.178		

Table 4.6: ANOVA summary for effects caused by the emotion and technique.

	Gaits vs Incidents	Gaits vs Combination	Incidents vs Combination
Happy recall	<0.001	<0.001	<0.001
Sad recall	0.129	<0.001	<0.001
Angry recall	0.428	<0.001	<0.001
Afraid recall	<0.001	0.131	<0.001
Overall recall	0.102	<0.001	<0.001
Happy precision	0.274	0.012	<0.001
Sad precision	<0.001	–	<0.001
Angry precision	0.039	<0.001	<0.001
Afraid precision	0.112	0.236	0.218
Overall precision	<0.001	<0.001	<0.001

Table 4.7: The p-values comparing gaits vs incidents vs combination for all results and precision. The results were only accurate up to 3 decimal places.

From these results, the three techniques, gait only, incident only and combined were ranked according to their effectiveness. The resulting ranks are shown in Table 4.8. At first glance, the confusion matrices implied that the combination technique (gaits plus incidents) was much better than either technique on its own. However, the rankings and comparisons show that there are many cases where an individual technique is as useful (or possibly better) than the combination, especially if the goal is to have higher precision.

	Prefer Recall	Prefer Precision
Overall:	Combination >> Gait ≈ Incident	
Happy:	Combination > Incident > Gait	Combination ≈ Incident ≈ Gait
Sad:	Combination ≈ Incident > Gait	Combination ≈ Gait > Incident
Angry:	Combination > Incident ≈ Gait	Combination ≈ Incident > Gait
Afraid:	Combination > Gait > Incident	Incident > Combination > Gait

Table 4.8: Ranking of the three techniques: gait, incident and combination (gait plus incident).

4.4 Which technique(s) to use?

From the results, it is not immediately clear which technique one should use so that participants can most accurately identify the emotions of these characters. This is because the best method will depend on what matters more in the game: precision or recall. The relative importance of precision and recall may be highly dependent on the type of game being created. Table 4.8 shows the ranking of the three methods depending on whether one values recall or precision.

According to the table, the combination (gaits plus incidents) technique produced the best overall results for recall (and for most individual emotions). There was only one emotion, sadness, where the incident scores were equivalent to the combination scores. However, for precision measures, three of the four emotions were able to produce equally good results from an individual technique (gait or incident).

Gait, incident and combination present different costs to the game designers. Gait will require a minimum of four gait animations per character type to represent each of the four emotions. However, if all characters always perform one of the four gait animations, game players may learn to ignore the emotion cues, so a neutral gait is probably necessary to maintain the impact of the emotional gaits. On the other hand, using incidents requires a larger set of animations to implement the actions and reactions needed. It also requires identifying game objects for use within the incidents. However, in a game, the incidents would likely happen randomly (whenever the conditions allow for an incident) and for finite periods of time (unlike gaits, which are continuous), allowing them to create a more unpredictable atmosphere. This unpredictability and decreased frequency make it less likely that game players will become so used to seeing the incidents. Finally, using both the gait and the incidents together requires creating both sets of animations. Overall, the combined method produced the strongest results, but the cost of implementation suggests that game designers should focus on the best technique for each emotion.

4.4.1 Use in Entertainment Games

Entertainment games could use emotion to enhance player experience. However, emotion is still rarely seen on background characters. We believe that most entertainment game designers will be more concerned about precision than recall. Since current games rarely provide emotional cues, most designers will probably be less concerned about a player missing a cue than wrongly interpreting a cue. That is, a designer will want to know that if a player *does* ascribe an emotion to a character that it was the intended emotion.

4.4.2 Use in Serious Games

On the other hand, for serious or training games, we believe the opposite is true: recall matters more than precision. A high recall value would mean that when players recognize that a character *is* emotional, they are identifying the correct emotion in relation to that character. If an emotion is being displayed in a serious game, it is usually for a very specific reason and it is important for the player to identify (recall) it correctly.

However, at the expense of a high recall value, players may ascribe an emotion that the designer was not trying to display (low precision). This could cause a player to mistakenly identify a neutral character as happy.

4.5 Anecdotes

Anecdotal accounts from the study participants indicate that participants focused on very different characteristics and identified different patterns and observations as important for classification. For example, in the happy interaction, the character kicks the soccer ball back to the child, while the angry character kicks the soccer ball away from the child. Some participants noticed this and determined that it was emotionally relevant, while others thought the difference had to do with the character not being accurate when kicking the ball.

A second example is based on the arm and hand positioning of the character. An angry character walks with clenched fists, while the afraid character has their hands slightly in front of them, with splayed fingers. Again some participants used hand information to identify all four emotions, while other participants ignored or did not notice this cue. Figure 3.1 in Chapter 3 shows the different hand and arm positions.

When the angry character passes the man on the bench, she turns to face him and speeds up. Some participants remarked that the "angry character glares at the man", which is notable, as the player only sees the back of the character, and cannot see the character's face. This strongly suggests that participants were not only identifying the emotion, but were mentally ascribing extra non-verbal behaviours to the character which were consistent with the perceived emotion.

4.6 Conclusion

In this experiment, we were interested in answering three questions:

1. Can participants identify each emotion with a minimum of 70% accuracy?
2. Do our emotional incidents convey accurate representations of the emotions they are supposed to represent?
3. Is one technique (emotional gaits or emotional incidents) better at conveying emotional cues?

From the results, we were able to exceed a minimum accuracy of 70% when using the combination technique. The individual techniques (gaits or incidents) were able to exceed this value for some emotions (afraid for gait, happy and sad for incidents), but not all. Two of the emotional incidents (angry and afraid) were below the 70% accuracy we were aiming for. This suggests that they were not conveying strong enough emotional cues from the incident's information.

While not all techniques worked well for all emotions, the experiment did confirm that participants can accurately and precisely identify emotional cues from non-verbal, non-facial behaviour as expressed through combinations of emotional gaits and emotional incidents. We found that the individual techniques could produce equally strong results on their own for individual emotions, but that using the combination technique would produce the strongest overall results.

Chapter 5

Gender Differences

During the examination of the results from the first study, we started to investigate whether or not the participant's gender[5.a] had any influence on the results. As the ratio of game players continues to shift towards a 50:50 male-female ratio, including game features that discourage either males or females from playing games will become more costly. In addition, since the major motivation for this research came from the serious game world, it is even more important that gender issues are considered.

It quickly became apparent that there were two design issues in the initial study when it came to gender differences. First, the character that the participants were asked to study was female, while the characters who interacted with her were male. As it is possible that the gender of the character will influence a participant's perception of an emotional cue, having participants only observe one gender may mean that our results cannot be extended to characters of both genders. Second, the attire of the character, while not provocative (especially when compared to most female video game characters), could be toned down to be less revealing so as to not potentially influence the results.

We designed a second experiment to specifically answer questions about gender differences and took into account both of these issues.

[5.a]In this chapter the term gender refers to both the self-identified gender of the participant, given two choices: male and female, as well as biological gender of the characters as male or female, represented in game using body shape.

5.1 Experimental Design

The first step in creating our new experiment was to develop two character models, male and female, with similarly-styled basic everyday clothing. For these models we settled on a simple plain t-shirt and jeans. As we didn't want the characters to look *identical*, there were two versions of the T-shirt, light grey and light tan. The colour of the T-shirt was randomly determined for each character at the start of the experiment (Figure 5.1).

Figure 5.1: The two models used, displaying the two different T-shirt colours used.

In the first study, we had thirteen different scenes for the players to watch. However, as we were more interested in whether or not there were noticeable differences affected by gender over whether there were differences caused by techniques, we did not feel the need to repeat all thirteen scenes. Instead, we decided to focus on the combination scenes, as the previous study had already shown that the combination technique produced the best overall behaviours. However, since we now had two characters of different gender, we wanted the participants to see each emotion performed by each character. This resulted in eight different scenes: four emotions (happy, sad, angry, afraid) by two characters. *Both characters (male and female) were using the exact same scripts and animations within the scene.* This was done so that any differences noted could not be attributed to differences in animation or reaction.

Similar to the first study, the participants were presented with the scenes in a random order. They were required to watch the entire scene to completion and then were presented with five possible labels (happy, sad, angry, afraid, none of these) and the option to re-watch the scene (which was used 3.1% of the time). We re-used the two emotional incidents from the first study. The first incident involves the character walking past another character who is sitting on the bench. The seated character waves at the walking character. The second incident has a small child kicking a soccer ball towards the walking character. Again, the study focuses on identifying the emotion of the walking character. While the small child (a boy) remains the same between scenes, the character that is seated on the bench is the opposite gender of the one who is walking. When the female character walks the male character sits on the bench and vice versa. The responses to the emotional incidents based on emotion are listed in Table 4.1.

5.1.1 Participants

The participants for our study were undergraduate students taking a first year psychology class at the University of Alberta during the Spring 2012, Fall 2012 and Winter 2013 semesters. They were between the ages of 17 and 31 (mean of 19.7) with 81 females and 81 males. Forty-five percent of the participants reported playing video games at least weekly (24% of females and 67% of males). Sixty-nine percent of participants reported playing video games at least monthly (51% of females and 86% of males).

5.1.2 Experiment Limitations

Again, for this experiment our participant pool was limited. They do not represent the general population or even the game-playing community, so care should be taken in applying the study results to a broader community.

5.1.3 Statistical Techniques

In order to analyze the data, we used the same two statistical techniques we used in the previous study - confusion matrices and bootstrapping. For more information on the techniques, refer to the descriptions in Section 4.2.1 for confusion matrices and Section 4.3.4 for bootstrapping.

5.2 Results

Our previous study showed that participants were able to accurately and precisely identify emotions using emotional gaits and emotional incidents as shown by a female character. In this study we examined the ability of participants to identify emotion of both female *and* male characters, with an additional focus on the gender of the participants. There were two questions we were trying to answer:

1. Does a participant's gender affect their precision and recall of a character's emotion?
2. Does a character's gender affect the participant's precision and recall of the character's emotion?

Because the characters were using the same animations and scripting for the emotional incidents, we expected that there would be no noticeable differences between the two characters. However, based on the earlier work by Judith Hall [38], we *were* expecting overall that the female participants would outperform the male participants.

We ran an ANOVA to determine if the results were influenced by emotion, participant gender, character gender, whether the participant was a gamer (played at least once a week), consistency (did they label both the male and female character with the same label for the same emotional cues) or any combination. The results of the ANOVA indicate that emotion, participant gender, game playing and consistency affect the results (see Appendix B). As well, there is an interaction effect between emotion and participant gender and another between emotion and consistency.

5.2.1 Female vs Male Participants

First let us examine the results of female versus male participants. The confusion matrices are shown in Tables 5.1 and 5.2 for the female and male participants based on the raw data.

Both male and female participants had precision values of at least 89.8%.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**150**	1	1	0	10	162	0.926
Sad	0	**148**	3	6	5	162	0.914
Angry	9	3	**132**	3	15	162	0.815
Afraid	1	0	4	**142**	15	162	0.877
P Sum	160	152	140	151			
Precision	0.938	0.974	0.943	0.940		0.949	0.883

Table 5.1: Overall confusion matrix for female participants.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**152**	1	0	1	8	162	0.938
Sad	0	**150**	6	1	5	162	0.926
Angry	9	8	**127**	6	12	162	0.784
Afraid	4	8	2	**117**	31	162	0.722
P Sum	165	167	135	125			
Precision	0.921	0.898	0.941	0.936		0.924	0.843

Table 5.2: Overall confusion matrix for male participants.

The recall results were compared for statistical significance at 95% confidence ($\alpha = 0.05$) and the p-values are listed in Table 5.3. The p-values were, again, calculated by following the technique detailed by Hardin and Shumway [39] for comparing confusion matrices using bootstrapping.

The most immediately noticeable results in the confusion matrices are the differences in the recall values for angry and afraid. The recall rates for both males and females were quite a bit lower for angry than for happy and sad, although there was no significant difference between genders (Table 5.3 - 0.345). For the afraid results, while the females had a slight drop compared to happy and sad, they were still quite high at 87.7%. However, the males had a much bigger drop, down to 72.2%, which is far below any of the other results. The difference between males and females was significant (Table 5.3 - 0.003).

Our hypothesis was that the female participants would overall outperform the male participants. However, this was not confirmed by our results, as overall the p-value was 0.081 (Table 5.3) and the only difference noted was between their ability to identify afraid characters. This suggests that our techniques are relatively (there is the one significant result) gender neutral.

	Happy	Sad	Angry	Afraid	Avg
Male Participants vs Female Participants	0.400	0.334	0.345	**0.003**	0.081
Male Characters vs Female Characters	0.384	0.312	0.512	0.399	0.377
Male Participants - Male vs Female Characters	0.156	0.373	**0.008**	0.365	0.334
Female Participants - Male vs Female Characters	**0.043**	0.160	0.296	0.432	0.313

Table 5.3: P-values comparing recall.

5.2.2 Female vs Male Characters

When comparing female versus male characters, we broke down the results into three groups. First, we compared the results from all participants (Tables 5.4 and 5.5), then only those by male participants (Tables 5.6 and 5.7) and finally only those by female participants (Tables 5.8 and 5.9). The results are summarized in Table 5.3.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**152**	2	0	1	7	162	0.938
Sad	0	**150**	4	3	5	162	0.926
Angry	11	7	**124**	4	16	162	0.765
Afraid	1	4	1	**130**	26	162	0.802
P Sum	164	163	129	138			
Precision	0.927	0.920	0.961	0.942		0.938	0.858

Table 5.4: Confusion matrix for **all** participants on **male** characters.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**150**	0	1	0	11	162	0.926
Sad	0	**148**	5	4	5	162	0.914
Angry	7	4	**135**	5	11	162	0.833
Afraid	0	4	0	**59**	18	162	0.796
P Sum	161	156	146	138			
Precision	0.932	0.949	0.925	0.935		0.938	0.858

Table 5.5: Confusion matrix for **all** participants on **female** characters.

As shown in Table 5.3, we did not find statistical significance for any of the overall comparisons (at the 95% confidence level), as expected from the ANOVA results. This suggests that there is no perceived difference in the labelling of male and female characters, which matches with what Hall found in her study of previous literature [38] and confirmed our initial hyphothesis.

Although the ANOVA results indicated there were no interaction effects between character gender, participant gender and emotion, we compared these results and found two apparent differences. Specifically, males appear to be better at identifying angry female characters compared to angry male characters and females appear to be better at identi-

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**75**	1	0	1	4	81	0.926
Sad	0	**75**	3	0	3	81	0.926
Angry	6	5	**59**	3	8	81	0.728
Afraid	0	4	0	**59**	18	81	0.728
P Sum	81	85	62	63			
Precision	0.926	0.882	0.952	0.937		0.924	0.827

Table 5.6: Confusion matrix for **male** participants on **male** characters.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**77**	0	0	0	4	81	0.951
Sad	0	**75**	3	1	2	81	0.926
Angry	3	3	**68**	3	4	81	0.840
Afraid	4	4	2	**58**	13	81	0.716
P Sum	84	82	73	62			
Precision	0.917	0.915	0.932	0.935		0.925	0.858

Table 5.7: Confusion matrix for **male** participants on **female** characters.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**77**	1	0	0	3	81	0.951
Sad	0	**75**	1	3	2	81	0.926
Angry	5	2	**65**	1	8	81	0.802
Afraid	1	0	1	**71**	8	81	0.877
P Sum	83	78	67	75			
Precision	0.928	0.962	0.970	0.947		0.952	0.889

Table 5.8: Confusion matrix for **female** participants on **male** characters.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**73**	0	1	0	7	81	0.901
Sad	0	**73**	2	3	3	81	0.901
Angry	4	1	**67**	2	7	81	0.827
Afraid	0	0	3	**71**	7	81	0.877
P Sum	77	74	73	76			
Precision	0.948	0.986	0.918	0.934		0.947	0.877

Table 5.9: Confusion matrix for **female** participants on **female** characters.

fying happy male characters compared to happy female characters. These results suggest that a more detailed study is needed to understand whether or not there is any real effect of character gender on the results.

5.2.3 Consistency

Fifty-six percent of participants made at least one ‘error’ in their identification. However, even more interesting was that within the 56%, approximately 81% of them made a ‘gender inconsistent’ error. A participant was considered ‘gender *consistent*’ if they gave the same label to both the female and male characters when the characters performed the same emotion. It did not matter if the participants were *correct*, just that the same emotion was selected for both the male and female characters. Table 5.10 provides an example of the definitions of correct and gender consistent.

Character	Happy	Sad	Angry	Afraid
Male	Happy	Happy	None	Afraid
Female	Happy	Sad	None	Afraid
Correct	both	female	neither	both
Consistent	yes	no	yes	yes

Table 5.10: Definitions of correct and gender consistent.

	Males (%)	Females (%)	P-value
Overall	82.4	84.3	0.255
Happy	92.6	90.2	0.237
Sad	92.7	88.9	0.176
Angry	72.9	74.1	0.393
Afraid	71.5	83.8	**0.024**

Table 5.11: Average consistency by participant gender and p-value.

This raised the question: were males or females more consistent? The overall result was insignificant ($p = 0.255$). However, as when comparing male and female recall, we broke these results down even further and looked at the individual emotions to see if there were any differences between the genders. Here there was one significant result. Male participants were significantly ($p=0.024$) *less* consistent when it came to identifying afraid characters compared to female participants. The p-values for all of the emotions (as well as overall) are shown in Table 5.11. This result is not entirely surprising, given that the males had a 72.2% recall in identifying afraid characters compared to females at 87.7%. However, we must take into account that our representation of afraid had issues back in our original experiment in Chapter 4. This result may be tied directly back to our representation.

5.2.4 Gamers vs Non-Gamers

We did wonder if the differences we were seeing could be due to other factors besides differences in player gender. However, our participant pool was homogenous. Everyone was about the same age (min 17, max 31, mean 19.7, mode 18), working on their undergraduate degree, all taking a first year psychology class, and all admitted to the same university. While cultural information on the students was not collected, the percentage of international students would likely be roughly equivalent between the male and female groups. This left gamers and non-gamers as an identifiable factor (from the demographics that we collected) that could also be influencing the results. Was it possible, that by playing more games, participants were being taught to ignore character animations as indicators of potential emotional cues?

	Gamers	Non-Gamers	Total
Males	54	27	81
Females	19	62	81
Totals	73	89	162

Table 5.12: Gamers versus Non-Gamers.

We divided our participants into two groups - gamers and non-gamers. Gamers were defined as those who indicated that they play video games a minimum of once a week. Table 5.12 details how many males and females fell into each category. Again, we created confusion matrices for these two groups (Tables 5.13 and 5.14) and used bootstrapping for analysis. However, although the ANOVA indicated there may be an effect, we found no statistically significant differences between gamers and non-gamers.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**141**	1	0	0	4	146	0.966
Sad	0	**137**	5	1	3	146	0.938
Angry	6	8	**116**	5	11	146	0.795
Afraid	2	8	2	**115**	19	146	0.788
P Sum	149	154	123	121			
Precision	0.946	0.890	0.943	0.950		0.932	0.872

Table 5.13: Confusion matrix for gamers.

	Happy	Sad	Angry	Afraid	None	R Sum	Recall
Happy	**161**	1	1	1	14	178	0.904
Sad	0	**161**	4	6	7	178	0.904
Angry	12	3	**143**	4	16	178	0.803
Afraid	3	0	4	**144**	27	178	0.809
P Sum	176	165	152	155			
Precision	0.915	0.976	0.941	0.929		0.940	0.855

Table 5.14: Confusion matrix for non-gamers.

5.3 Conclusion

We had two hypotheses for this experiment. First, that the female participants would outperform males and, second, that there would be no differences between male and female characters. In the end, we found one significant gender difference between the participants. While the female participants did not outperform the male participants overall, they did outperform the male participants on identifying afraid characters.

For our second hypothesis, the ANOVA results indicate that there were not significant differences between character genders. However, we did find two potential differences (male participants' ability identify angry male versus female characters and female participants' ability to identify male versus female happy characters) that suggest that our techniques may not be completely gender neutral and that more investigation is needed.

However, we were able to once again achieve recall rates above 70%. The lowest value was 71.6% for male participants on identifying afraid female characters (Table 5.7). Our previous study (Chapter 4) indicated that participants were most likely to stumble in identifying fear. From designing that study, we know that small tweaks in an emotional incident can cause significant changes in participants' ability to correctly identify that emotion. We believe that further refining of these emotional incidents would likely be able to overcome any of the noted differences.

Chapter 6

Creating self-determined Emotion Behaviour

Once we had validated that participants were able to accurately identify emotions from our emotional gaits and incidents, the next step was to put these characters into an actual game. Our original study was done using a virtual world with very controlled conditions. Every participant saw exactly the same incidents and gaits. However, in an actual game this would not be the case. An important purpose of this chapter is to provide dissertation readers, designers who wish to use this research, and other researchers who wish to extend this research, with a blueprint. This blueprint explains how we added emotional behaviours to a game and provides a description of the amount of effort required.

Creating a system for emotional incidents has two main components. First, we need to figure out how to determine when an incident should happen. Second, we need our incidents to be location independent, allowing them to occur wherever they are triggered.

However, it is not as simple as adding arbitrary emotional incidents at random. Each emotional incident is an interaction between game props, characters and emotions; if any piece is missing, the incident will be unable to occur. Also, because the goal of the emotional incidents is to convey important emotional cues, the frequency of the incidents must be balanced between too rare where they may never be seen and too frequent where they may be ignored.

One of our overall goals is to use the emotional incidents to create more believable characters. We want to produce the *Eliza Effect*, where people see behaviours that are not explicitly coded/designed by the creator [75], such as the study participant who "saw" a character glare at another character, even though the character's face was not visible. We were hoping that by watching characters perform emotion-specific interactions, players would develop their own reasons about why characters act the way they do, leading to perceived emergent behaviour [6]. In order to produce the Eliza Effect, from the literature, we found two characteristics to be the most important. First, the characters needed to be consistent ("stay in character") [62]. Second, there needed to be some level of unpredictability to their actions [75]. The chapter provides details on the goals of consistency and unpredictability, before detailing how the final system was implemented (and how these goals were met).

6.1 Consistency

A consistent character is a character that, when observed, stays within the expected behaviour bounds for that character. For example, a guard character is expected to be guarding *something*. For this character to stay consistent, the guard should continue to guard that item and react should someone attempt to steal it. An inconsistent guard might allow strangers to walk away with the item being guarded.

In order to make sure our characters stayed emotionally consistent, we chose to make all of the incidents emotion-specific. This meant that when a character decides to perform an incident, they could only choose from the reactions that were tied to their current emotional state; i.e. a happy character would never suddenly respond with an angry reaction. However, some incidents were actually shared among multiple emotions. For example, happy, sad, and afraid characters may all choose to sit down on a bench. However, these responses were not identical. In our scenes, happy characters are always willing to sit on benches while sad and afraid characters will only sit if no one else is already sitting, and an afraid character will immediately vacate the bench should another character sit on it.

6.2 Unpredictability

Predictability is needed so that players can figure out the rules of the game and how to play within them. Too much predictability can lead to a very boring game as everything starts to feel predetermined. When it comes to character behaviour, too much predictability leads to the perception that the behaviour is artificial. Even though we are generally creatures of habit (e.g. we have our morning routines and head to work around the same time) there is still variation within that behaviour (we usually don't eat *exactly* the same thing at *exactly* the same time).

Unpredictability can initially feel like it is at odds with consistency. It seems intuitive that behaviour thats seem consistent would allow observers to predict exactly what a character will do. However, consistency and unpredictability can exist in harmony. While we want our participants to be able to predict *how* a character **may** react (which is consistency), we introduce unpredictability by preventing participants from knowing **if** the character will react and/or the specifics of the reaction. Ideally, characters need to balance between being *too* predictable and *too* unpredictable.

For our scenarios, we added a probability factor (see Section 6.9.1) that determines whether or not a character should become involved in a potential emotional incident. This means that every time an incident is *possible*, we decide to engage with some pre-set frequency. For example, a probability factor of 100% would result in a character sitting on every bench that is passed (which looks odd, especially if there are a couple of benches next to each other). By reducing that probability (we used 50%), it was unlikely that participants would observe this odd behaviour, resulting in behaviour where characters only sit half of the time.

6.3 Emotion Architecture

In order to implement our incidents, we created an emotion architecture. The emotion architecture is responsible for deciding when to execute an emotional incident, what incident to execute, and controlling the actual execution. Unlike in our previous experiments, the emotional incidents could no longer be hard coded for specific combinations of characters and props. Instead, the architecture needed to be designed to allow the characters and props to determine whether or not they should start an incident and how they should respond. This meant the code needed to be written so that it was not character or prop specific, and that any path or timing changes needed must be able to be determined during game play.

The emotion architecture we used is shown in Figure 6.1. Each emotional character and prop has a collider. The collider is an invisible object surrounding the character that can be used to detect collisions with other game objects. There are also five main script components in the architecture, each described in a separate section of this chapter:

1. Section 6.4 - A **basic behaviour** script that controls the character's default normal behaviour.
2. Section 6.5 - A **character incident trigger** script to determine if two characters have collided.
3. Section 6.6 - A script that contains the code for executing emotional incidents between characters (in our system, it is contained in the **basic behaviour** script).
4. Section 6.7 - An **animation controller** used to start, stop and blend animations together.
5. Section 6.8 - A **prop specific** script to catch collisions between an emotional prop and emotional character as well as execute the resulting incidents.

Many of the decisions we made regarding the architecture (such as the use of a second collider on characters) were highly influenced by the fact that we were using the Unity 3D game engine. The following sections are based on the implementation we used in Unity and would likely need to be modified for a different game engine.

Our system also required an extended library of character animations in order to implement our emotional incidents. This includes both gait animations (for the emotion-specific gaits) as well as reaction animations (such as waving and kicking). The animations we used are described in more detail in the the section on the animation controller.

We ran all of our experiments within a park that we designed in Unity. A top-down view of the park is shown in Figure 6.2. The park contains a variety of objects (such as picnic tables, benches, books, newspapers, coffee cups, garbage cans and a soccer ball) that the characters can interact with. Many of these props are tied to emotional incidents.

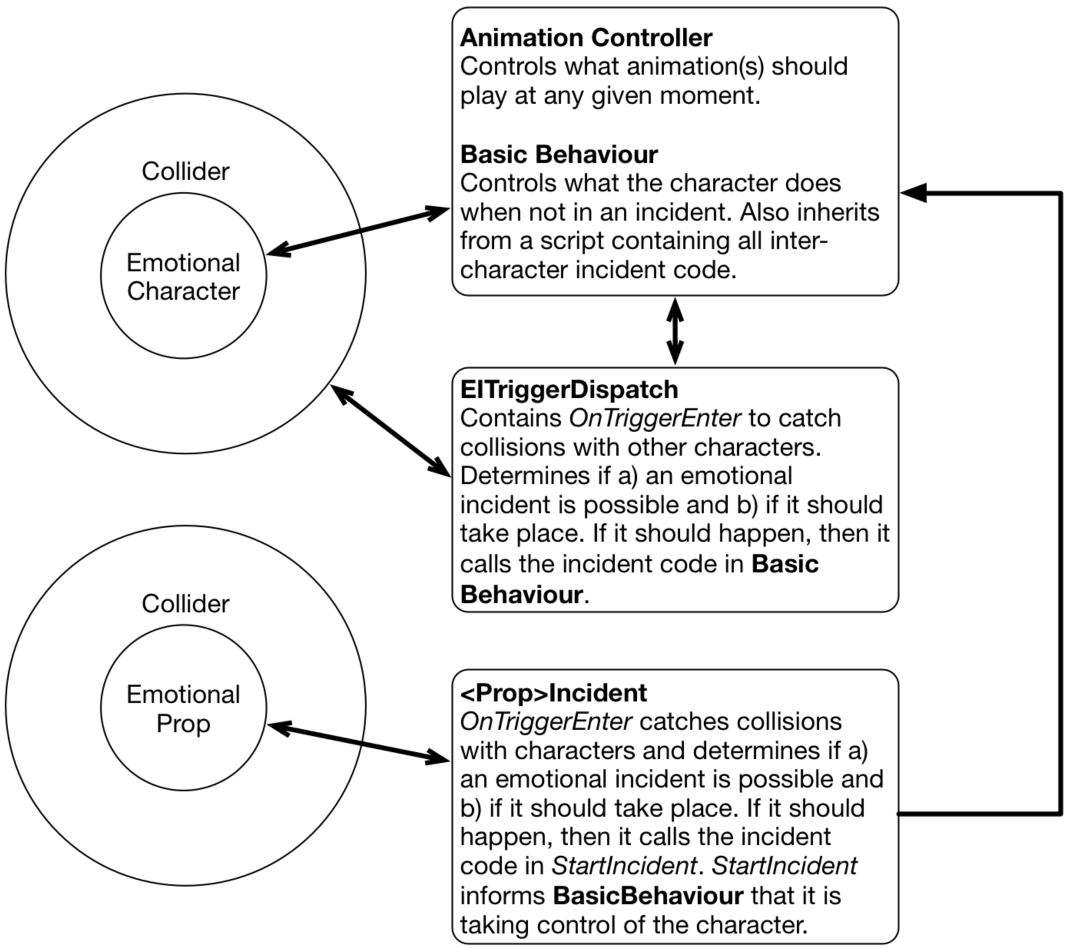

Figure 6.1: Organization for controlling emotion behaviour in the test scene.

Figure 6.2: The park where the game takes place.

6.4 Basic Character Behaviour

Each character in a game, emotional or not, will need to have some level of basic behaviour assigned. It could be as simple as they stand in the same place forever to a more complicated behaviour that integrates a daily cycle (such as the character goes to work, eats out, then goes home). For our system, our characters shared the same basic behaviour; they would randomly walk the paths in the park. Because of this, it was simplest to just create a single script file that contained this basic behaviour and all of the code for character incidents. However, in a game that has a more diverse set of basic character behaviours, it would make sense to have the basic behaviour inherit from the character incident script. This would allow the characters to detail their own basic behaviour while still sharing the code for all the character incidents.

The basic behaviour needed to have a flag that could be set that could be set to indicate if the character was currently involved in an incident and would stop executing the basic behaviour. It also included a variable that stored the character's current emotion.

6.5 Triggering Character Incident

Each character in our system has two colliders attached. First, they have the ‘normal’ collider, representing their actual body to prevent them from walking through each other. Second, they have what amounts to a giant cylindrical collider (used as a trigger). The character is located at the centre of the cylinder, and the radius of the cylinder determines how far away other characters and props can be for the character to initially ‘detect’ them. By using a circular shape, the characters are able to equally detect objects in all horizontal directions from them.

We decided not to rely on a ‘visual perception’ type system that would have a character detect objects within “eyesight.” Instead, our approach is consistent with a more general perception system where a character could react to a noise behind them. Therefore, some of our incidents can be triggered by an object colliding from behind the character (for example, causing the character to do a shoulder check).

A script was assigned to this larger collider and was fired every time an object entered the collider. The script determined if the entering object was another character. If the answer was ‘yes’, then the script would check if the character was already in an incident (characters could only be involved in one incident at a time). If the script passed that check, the script would then determine if an incident was possible between the two characters. For an incident to be possible, three things were required:

- There was an incident that was possible given the emotions assigned to the two characters.
- There was an incident that was possible given the “in an incident” state of the two characters. (At most one character could be already engaged in an incident.)
- It passed a probability check.

It was not a requirement that both characters were free to interact as some interactions between characters should happen no matter what. For example, an afraid character that veers away from other characters should continue to do this even if the character(s) s/he is passing are otherwise engaged.

If the script passed through all the checks, it would then call the *basic behaviour* script and start the appropriate response. It would also pass a reference to the other character,so that the two basic behaviour scripts could work together in their response. Once the behaviour was finished, the character would have its “in incident” status revoked and the basic behaviour script would be told to resume ‘normal operations.’ This approach is a simplification of the behaviour architecture described in [20].

6.6 Character Incidents

As stated previously, the code for the inter-character incidents is stored in the basic behaviour script. Once all the steps for starting an incident are passed, the code for individual character incidents occurs. Depending on the incident, code would either be started on both characters (that would work in tandem), or a single character would become the dominate character for the incident and run code to control both characters throughout the incident. This was dependent on how highly integrated the incident needed to be.

For example, the scripts for an incident between a happy and angry character were able to be kept separate. The happy character waved briefly until the angry character passes her by. The angry character speeds up long enough to pass the happy character. While they both need to know the location of the other character, they do not need to time the wave in response to some action from the angry character.

A more intertwined incident is one between two happy characters. Here, they start by both waving at each other. As they get near, they stop to chat for a while. Finally, if one character is holding a ‘shareable’ object, that character may choose to offer it to the other character. All of this requires careful timing and positioning of the characters and thus is better controlled by a single script. Naturally the complexity of the interaction determines the degree of coordination necessary and such coordination is more expensive to create.

With any complex system, robustness is important as it is always possible in a game that there may be a glitch. Therefore all incidents had a timeout value assigned. If an incident does not finish before the timeout, scripts are terminated and the character goes back to behaving normally. Termination is also triggered if character A started responding to character B (who he thought was approaching) only to have character B turn away. Character A will only wait until the timeout before giving up on character B.

6.7 Animation Controller

The animation controller was responsible for enabling, disabling and blending animations together. Each incident was usually built from a combination of the individual reaction animations we had developed.

In order to implement our emotional gaits and emotional incidents, we created a large number of individual animations that could be combined together. The following briefly describes the animations that are included in our system. Not all of these animations were used in all of our experiments as many were only used if specific emotions were included in the experiment.

Emotional Gaits

Our first set of animations is used to create the emotional gaits. Each emotional gait (and neutral) were broken into two individual animations. First, there is an animation that controls the general motion of the animation (moving the legs). The second animation is tied to the character's arms. This was done so that it would be possible for characters, when standing idle (which is included in the following list), to continue to maintain the arm/hand configuration that was tied to their emotional state (angry characters would continue to have their hands in fists, afraid characters would have their hands splayed and away from their body).

- Neutral
- Neutral Arms
- Angry
- Angry Arms
- Sad
- Sad Arms
- Happy
- Happy Arms
- Afraid
- Afraid Arms
- Idle

Head Movements

Next, we added a set of head-movement animations. Unity has a script to control basic head movement, allowing characters to try and "follow" an object with their eyes/head while still not twisting past reasonable limits (no 180 degree head turns). However, there were some movements we wanted to display that cannot easily be done by visually tracking a target.

The first is the 'check over shoulder' movement. This is used if characters "feel" like they are being followed. Or, if a character is afraid, they may use this periodically no matter what, because their fear compels them. Next we have nod and shake head movements. As people generally use nodding for agreement and shaking for disagreement, these movements allow for characters to convey information when in a "discussion" that doesn't require any verbal dialogue. Finally, we include the 'Talk' animation in this system. As our characters had moveable jaws, when they were supposed to be conversing with another character (or the player), they would use the talk animation to appear more life-like. However, the talk animation did not change the regular emotionless expression that the characters all shared since we didn't want participants to rely on (or be able to collect information from) the facial animations.

- Check over shoulder
- Nod head
- Shake head
- Talk

Sitting

As we wanted to increase the number of incidents possible in the park, we decided to make use of the picnic tables and benches throughout, and the best way to use them was to allow characters to sit down. To start with, we needed three basic animations: one to allow characters to sit down, a second so characters could stand back up, and a third for them to use while staying seated (sitting idle).

We added three more animations to this combination. A nervous animation was created as we believed that afraid characters would not sit as calmly as the other characters. As the benches were created to accommodate two characters sitting at the same time, we decided to make it possible for two characters sitting together to interact. This led to the creation of two companion talk animations that characters could use while sitting. These animations combined a head look to the left/right, hand gestures and mouth movements, making the interaction appear more real. A seated character could also use the basic talk animation and did, but usually only if the other character was standing in front of them.

- Sit down
- Stand up
- Sit idle
- Sit nervous
- Talk to left
- Talk to right

Kicking

The path of the park was littered with small rocks that characters could kick as well as a couple of soccer balls. We created two kicks (normal and angry) and each was done twice so it could be done by the left or right leg. In the end, we ended up using the normal kick animation and speeding it up if we wanted to display more anger in the movement. The angry animations ended up being slightly too exaggerated for our needs.

- Kick left
- Kick right

Item Interactions

Throughout the park there were items (newspapers, books, a tablet, coffee cups) that characters could choose to pick up and interact with. The items were generally placed at two height levels: the top of the picnic table (which corresponded closely with the height of a character's bent arm), or on the ground. We therefore had two animations designed for picking up from these two different height levels.

Once an item was picked up, there was a basic hold idle animation, allowing a character to carry the item around. The character could also choose to interact with the item. First, they could look at the item (either briefly, or for a sustained period of time), point at the item, drink from the item (which only makes sense with cups), and mimic using a tablet.

When a character was finished with using an item they could choose to throw the item, or pass the item to either a table or another character. If the character was passing to a second character, they would use a combination of pass item down and pass item up, such that the character who gives up the item finishes with his hand down by his side and the character receiving the item finishes with her arm bent and ready for the hold idle animation.

- Pick up from ground
- Pick up from table
- Hold idle
- Look at holding item briefly
- Look at holding item idle
- Point at item
- Drink from item
- Use tablet
- Throw item
- Pass item down
- Pass item up

Other

We had two final animations to round out our system. The first was a wave animation. The wave animation could be set to use either the left or right arm or, if left alone, it would use both. The second animation was a startle animation, used by the afraid character. The startle animation works from an initial standing idle animation. The character takes a half step backwards while his arms came up slightly in front, before returning to back to the idle animation. Changing the speed on these animations made a noticeable difference in how friendly/afraid a character appeared. A short wave could come off as more curt than happy. And a longer startle animation (allowing the character to move further into his back step) showed a much greater level of surprise.

- Wave (can choose between left, right, or both arms)
- Startle

6.8 Prop Incidents

The final script was used to control incidents that involved a character and a prop. In this case, all the code was placed onto the individual props. A prop with emotional incident code could initiate an emotional incident. Props without this code would not be used by characters for an emotional incident. This meant that, should a prop have a specific purpose in the game (think of a lost tablet that the player is trying to find), by not putting emotional incident code onto the special prop other characters wouldn't accidentally pick it up and move it (or, in the case of our park, throw it out).

Each prop had a collider on it, larger than the initial prop. This was similar to the cylindrical collider used by the characters and allowed characters and props to 'notice' each other with enough time for an incident to be decided on, planned, and executed. Similar to the script on the cylindrical collider for characters, the script on the props would trigger when the collider was entered and would determine whether or not to start an incident. Again, the colliding object needed to be an emotional character who was not already involved in an incident, and whose emotional state had a possible interaction with the object. For example, there were specific incidents designed between angry or happy characters and a rock, but not for sad or afraid characters.

Once an incident was started, the prop would tell the *basic behaviour* script that the character was 'in incident' and that it was taking over control of the character. It would also mark itself as 'in incident', so that another character could not attempt to start an incident with it. The prop would then control the execution of the incident by controlling the character's movements. Once the incident was over, the prop would inform the *basic behaviour* script that the incident had ended and that it should resume control of the character.

6.9 Emotional Incidents in the Park

In the end, we had a large variety of animations that could be used in the park. They are all listed in Table 6.1, sorted by prop, then emotion. We did notice that, with our setting and props, it was much easier to identify and create 'happy' emotions than the other three emotions. However, this is not completely surprising as there is nothing inherently scary about the park (fear), no one is taking particularly annoying actions (anger) and as it's a nice sunny day, the only thing to make someone sad would be something internal going on before the character arrived in the park. If we had changed the setting, we could end up with the opposite problem. For example, if our games took place during the night in a dark alley, it would be much more natural for available incidents to favour the afraid emotion.

Object	Emotion	Reaction
Rock	happy	stop and kick
	angry	kick while passing
Book	happy	pick up
Cup	happy	pick up
Newspaper	happy	pick up
	angry	kick on ground
Garbage Can	happy	throw out newspaper/cup
Picnic Table	happy	put down newspaper/cup/book
Ball	happy	stop, kick ball back to kids
	angry	kick ball away
	sad	ignore ball, keep walking
	afraid	pause and show startle animation
Bench	happy	sit down
	sad	sit down if no one else is on bench
	afraid	sit down if no one else is on bench
Two on Bench	both happy	sit and talk
	afraid & other	afraid character gets up and leaves
	sad & other	sad character waits a few seconds before leaving
Two characters	both happy	wave, possibly stop and talk, possibly exchange item
	one happy	waves at other
	one sad	looks briefly at other, then ignores
	one afraid	backs away from other character
	both afraid	change direction and walk away from each other
	one angry	speeds up and looks at other character while passing
	all emotions	shoulder check if second character enters the collider for the first character from behind

Table 6.1: The emotional incidents in the park.

Beside the emotional incidents, there were some 'static' behaviours that we do not consider incidents that can occur throughout the park. First, if a character picks up the cup, they will periodically 'drink' from it (approximately every 20 to 30 seconds). Second, if the character is afraid, she will periodically check over her shoulder. This is different from the shoulder check incident as it is not triggered by another character coming close to the first. Third, anytime a character gets too close to the ducks, the ducks run away from the character.

Throughout the various experiments we tested in the park, this animation list was modified. Some were excluded because the game did not involve the specific emotion. Others changed because the combination of events became too difficult to time, such that it never actually occurred (this occurred most often with the ball).

6.9.1 Probability Factor

When we stared, we knew that we didn't want the incidents to occur *all* the time because this would lead to overly predictable behaviour that would feel artificial. We also witnessed this ourselves as we would watch a character sit on a bench, get up, walk farther down the path, and then sit down again, continuing this way throughout his entire exploration of the park. To support unpredictability, we added a probability factor to determine whether or not an incident *should* occur. We set most of the initial values around 80% so that they had a high likelihood of occurring, but were not guaranteed.

As we continued to develop the park, two issues occurred. First, the total number of characters in the park grew. Second, we started to mark some of the characters as neutral. These ended up causing some interesting side effects.

As the number of characters grew, there was a greater chance of inter-character emotional incidents occurring. As more of these occurred, the likelihood of an incident with a random prop (such as a rock on the ground) decreased. This was because characters were often already 'in incident' when they collided with the rock, making them ineligible to kick the rock (should their emotion also line up).

As we changed some of these characters to neutral, initially the number of incidents increased again because there weren't as many inter-character interactions. This soon changed because we realized it was odd that emotional characters would only respond to other emotional characters. There seemed to be "two worlds" existing in the park. The first world was full of these neutral ghost characters that did not appear to realize that there were other characters (or props) in the park. The was a second more "lively" world, where characters would interact but who were apparently unable to see the "ghost" characters. We therefore changed our system, allowing the emotional characters to react to the neutral characters, even if the neutral characters would not respond back. This allowed afraid characters to still try to avoid getting close to neutral characters, and angry characters to storm by neutral characters. However, this returned us to the situation where so many inter-character incidents were occurring that prop incidents rarely occurred.

In the end, we found that for *many* of the incidents a probability factor was not needed. However, we did leave it in for two situations: kicking the rock and sitting on benches. This was because these were the two most noticeable incidents that people would catch and both made the characters look unrealistic if performed repeatedly. A character can wave at all approaching characters and appear friendly. However, kicking every single rock does not appear "normal", except perhaps for kids.

6.10 Conclusion

The extra requirements for emotional gaits and incidents are relatively simple to implement. Any game involving 3D characters that wander a game world already needs an internal animation controller system, so ours would not be considered something 'new.' Similarly, all characters would need a basic behaviour script (or scripts) controlling what they should be doing in the game.

The are three main additions that are needed to support emotions. First, all emotional characters and props need an extra collider (or mechanism) to detect each other from far enough away so that, if an emotional incident occurs, there is enough time for the planning and execution. Second, these colliders need a script that allows them to evaluate an incoming object and determine whether or not an incident should be started. Third, generic code needs to be written such that it is not tied to individual characters, to execute the actual incidents.

In our system, the code for each incident was only written once. All rocks shared the same code for a rock incident. All characters shared the same code for executing inter-character incidents. Using the same code may initially feel like the characters and incidents are all going to appear identical. However, because incidents can occur anywhere, they appear more unique than they might otherwise appear. A character kicking a rock by the pond in our park appears different than a character kicking a rock over by a tree, even though the code running is the same. The setting and triggering of the events often means that characters approach incidents from different angles, increasing the perceived variability of the individual incidents.

Chapter 7

Experiments on adding emotional characters into games - Studies 3 and 4

Our emotional behaviour architecture was designed so that emotional characters could make their own decisions as to when and what emotional incidents to be involved with. Once implemented, we introduced the characters into the game. While we knew that study participants were able to accurately identify the emotions of these characters, they had done so in a very scripted and artificial environment. It was important to determine whether participants could still identify these character emotions when the chances of seeing emotional incidents was much lower.

This chapter details the first few experiments we ran where we tried to put these characters into games. While we didn't end up with the objective comparison measures we were hoping for, we learned many lessons from these experiments (detailed in Section 10) that enabled us to create a successful comparison experiment (Chapter 8). In addition to preparing us for a successful comparison experiment, these studies also provided qualitative evidence that players ascribe motivations to characters who present emotional cues in a game environment.

7.1 Return Tablet - Study 3

The first game we created tasked players with finding the owner of a lost electronic tablet so that they could return it. We wanted the participant to play a character who would have some motivation to wander around and observe the emotional characters. We created a simple game based on the classic 'search' quest, where the participants were tasked with searching for the tablet's owner. In the game, the only ability the participants had was to talk to characters and ask them one question, "Is this your iPad?"

7.1.1 Experimental Design

This experiment consisted of three stages: a tutorial, the actual 'game'/experiment, and, finally, questions about both the game and basic demographic information.

Tutorial

This was the first experiment that we ran where players needed to move and control a character. In the previous experiments, the scenes were all heavily scripted and, therefore, for the participants, it was similar to watching video clips. We also knew from the previous experiments that our participant pool was largely made up of non-gamers and that many would not have experience with controlling characters in a 3D game environment.

The tutorial required the player to control a character right in the park, as if it was the actual experiment. It started by explaining the control keys (WASD) for moving the character (see Figure 7.1) and then how to use the mouse to look around. Next, it asked players to track down a character (Fred) in the park. Once the player's avatar was close to the character, they were then taught how to talk to a character. After talking to Fred, they were then tasked to talk to another character in the park before interacting once with a prop (a rock) in the park.

At both the start and end of the tutorial, participants were told that it was very important that they were comfortable controlling their character, and that they could remain in the tutorial after it was finished for more practice. Twelve of the 85 participants (14%) did remain in the tutorial longer and they averaged an extra 51 additional seconds in the tutorial.

Figure 7.1: Screenshot during the tutorial phase of the Return iPad experiment.

Game

After participants had exited the tutorial, it was time for them to play the actual game. Before starting the game, they were told that they had found an iPad in the park and they were given the task to find its owner and return it. From the tutorial, they knew that they could talk to characters in the park.

There were 12 characters in the park who could be the potential owner of the iPad. The characters were grouped into pairs. Five of the pairs were assigned an initial emotion (2 happy, 1 each of sad, angry and afraid). In each pair, one character was labelled A, while the other was labelled B. In some versions of the game, the A 'group' would be assigned to show visual emotions while the B 'group' was neutral. For example, in Table 7.1, Steve is an A character and Mark is a B character. While they may both be assigned *happy* (as in the table), only one will be chosen to use the emotional gait and perform emotional incidents. Whether it is Steve or Mark (A or B) is determined randomly at the start of the game. The other character will use the neutral gait and have no emotional incidents.

The sixth pair is special, as one character is assigned to neutral and the other is assigned to be the character who lost the tablet. The character who lost the tablet starts off as sad, but he or she becomes happy when the tablet is returned. As we paired characters by gender and three emotions were only represented by a single pair, we created two sets of these pairs/groups so that there was a 50/50 chance that participants would see females vs males as each of these three emotions. This resulted in four variations of our experiment (two sets of emotion assignments and two groups for each set). Table 7.1 shows one such breakdown of the pairs and emotion assignments.

Emotion	Character A	Character B
Happy 1	Steve	Mark
Happy 2	Rebecca	Erin
Sad	Jenny	Donna
Angry	Jake	Fred
Afraid	Kate	Laura
Neutral		Jack
iPad	Andrew	

Table 7.1: One example of pairs and groups for the Return iPad game.

Each character also had two dialogue lines to use when the player asked them about the tablet. The dialogue lines were designed to be emotion-specific. So even though Steve may be assigned to *not* visually show emotion, he would still use the happy dialogue lines. The set of dialogue lines used are listed in Table 7.2. The characters would use their first dialogue line when the player first approached them. However, on any additional queries from the player, they used their second dialogue line. None of the dialogue lines provide any meaningful information in order to locate the actual owner. When the participant chose to speak to a character, most characters would stop and respond. However, the angry characters would give their dialogue line in orange text over their head as they continued walking (speeding up if they were supposed to be showing visual emotional cues). The choice of orange text was only to make it clearly visible against the game background.

The actual owner does not start off in the park. Instead, he or she appeared out of sight of the player once the player had talked to six (different) characters. This prevented the game from ending too early and forced the players to interact with more characters in the park. Once the player returned the tablet, they could choose to remain in the park or exit the game. Twenty-three participants (27%) did choose to remain in the park and did so for an extra 87 seconds on average.

Emotion	Order	Dialogue Line
Happy 1	1A	Oh. Glad it's not me who lost it.
	1B	Yikes, still haven't found the owner?
	2A	I thought I saw someone looking for something earlier. Maybe they lost it?
	2B	Try near the picnic table.
Happy 2	1A	I always have my smart phone on me, so I don't have a tablet.
	1B	It's such a nice day here. Good luck finding the owner.
	2A	Oh wow, I love mine. But it's at home.
	2B	Glad I left mine at home. I wouldn't want to lose it here.
Afraid	1A	Huh? A tablet. No... Not mine.
	1B	It's not mine.
	2A	No... No, it's not mine.
	2B	I told you, it's not mine.
Sad	1A	I wish I had one. But that's not mine.
	1B	Still not mine.
	2A	No, not mine.
	2B	Maybe someday I'll own one.
Neutral	1A	Sorry, I don't know who that belongs to.
	1B	Sorry, I still don't know who that belongs to.
		The other half is the one carrying the iPad with special lines

Table 7.2: Emotion specific dialogue lines used in the Return iPad game.

Questions

The final phase of the experiment was the question phase. This phase asked the players to remember the characters from the park and answer some questions about them. For each character, players were asked if they observed or interacted with the character. If they answered 'yes', we asked them to explain what they saw the character do, what emotion they thought the character was displaying, and why they picked that emotion. After going through all 12 characters, they were then asked about the characters by pair. The goal here was to determine if the emotional character was more noticeable/enjoyable than the neutral character. We then supplied a checklist of the possible interactions that could occur in the park and asked participants to check the ones that they saw during game play.

7.1.2 Participants

We had 85 participants from the psychology 104/105 classes at the University of Alberta during the Fall 2013 semester. There were 27 males and 58 females. They had an average age of 18.8 years (min 17, max 26). Twenty-one percent self-identified as game players (indicated they played at least once a week).

7.1.3 Results

On examining the results from this study, we noticed some key issues:

- Participants, on average, talked to 9.8 different characters. However, they *remembered* seeing or interacting with, on average, 7.2 characters.
- For the 7.2 characters they remembered, they indicated that they could not remember/did not notice the emotion for 2.5 of them.
- 31% of the time participants indicated they did not know or remember both characters when answering the comparison questions about the pairs of characters.
- 19.5% could not remember the character they returned the tablet to.

It soon became clear that players did not remember the characters well enough to objectively measure differences between the characters who displayed emotional cues and those that did not display emotional cues. However, looking through the comments that were submitted for the characters who display emotional cues, we did notice that, even if it was difficult to compare the individual results, the participants were noticing character emotion and ascribing reasons for the behaviour of these characters. Some of these comments are listed below. A full list of these comments is in Appendix C. Note that there were no facial expressions (happy eyes or smiles, etc.) in the game so participants who describe facial expressions are inferring them. These are examples of the Eliza Effect we hoped to elicit.

- He was walking around very slowly looking very sad.
- He was walking faster than other characters, a little paniced, he was the one who lost the ipad.
- Walking sulkly through park
- he sat on a bench by the lack. he was lonely.
- run around agrilly
- Engaged in friendly conversation when I talked to him. Appeared relaxed. Could not see his eyes, but from his conversation and his body language he still seemed happy.
- She was the happiest of the characters. She was nice and conversed in a very friendly way even though we had never met. She gave more than a yes or no answer the first time I talked to her, and the second time she furthered our conversation. Her clothes were also clean, nice, and colourful. Her face and eyes looked happy and friendly. She looked healthy.

- Jake was happy to be enjoying the outdoors on a nice and sunny day. He was just talking a stroll in the park.
- maybe sad because he sat lonely and maybe was reminising on things
- His shoulders looked slumped and his expression looked sad or depressed
- He was walking slowly and he was hunched. He was looking down and his eyes looked sad and droopy. He looked unkempt: He had stubble on his face and was dressed sloppily in a dirty looking shirt, like he had given up.

However, even for the non-emotional characters, there were a few times where emotion was ascribed to a character.

- He said something along the lines of good luck finding the owner"[sic]. Sounded cheery and smiled.
- I can't remember whether she was happy or afraid. She talked to me plesantly the first time, and gave me more than just a yes or no answer, which would indicate that she was happy. But she had a frightened look in her eyes and appeared tense, and I think when I talked to her the second time she seemed more distracted and afraid in what she said.
- walk around grumpily

7.1.4 Return Tablet version 2 - Study 3.5

We also ran a second version of this experiment that differed in two significant ways. First, the players played through the game twice. One game contained only emotional characters while the other contained only neutral characters (the order was randomly determined at the start). The second change was that we asked these participants to rate each game as to how enjoyable, how believable the characters were, and the amount of variety in the characters' actions and activities. They were also asked to chopse specifically between the two games for each question.

As we were analyzing the results from this experiment, we realized that the participants were not sharing a similar baseline for any of the categories. A participant who never played games was using a completely different scale for believable characters than a participant who had played many RPGs. Because of this discrepancy, we could draw no meaningful conclusions from the data.

7.2 Changing Emotions - Study 4

As we concluded the Return Tablet studies, we started thinking about another issue that could be also be explored. What would happen if we allowed game events (including actions taken by the player) to affect a character's emotional state? This is something that *should* happen in games, even if it seldom does. If a player burns down a town it only seems natural that the townspeople should be angry with the player.

We decided to try a simple model. In this model, each character had a vector of four values, where each one represented a different emotion (happy, sad, angry, afraid). The current emotion was determined by looking for the emotion that has the max value and determining if that value also passed a preset threshold. For example, each value was a number between 0 and 100 and the threshold for displaying an emotion was 50. So if a character had the following vector [0, 60,70,0] they would display anger. And a vector of [40,0,0,0] would elicit neutral behaviour, since the happy value (40) does not cross the threshold of 50.

Each emotional incident was then assigned an effect vector, where the value for each component emotion in the vector would be positive or negative. A positive value meant that a character who witnessed the emotional incident should increase the component of that emotion in their emotion vector, while a negative value decreased that particular emotion component in the witness's emotion vector. So an effect vector of [50,-50,-50,0] would make a witnessing character more happy, less sad, less angry and would not change afraid.

For this user study, we implemented several actions that the player could take. For example, the player could kick the ducks (which made characters angry and afraid) or pick up litter and place it in a garbage can (making characters happy). These actions were added so that the player could purposefully try to change a character's emotion by performing specific actions in front of a character.

7.2.1 Experimental Design

The user study consisted of a tutorial, two play-throughs of the game (one with emotional characters and one without) and questions. The tutorial ended up being quite lengthy, as we were trying to teach the players many of the actions that were available to them. However, as we have mentioned before, participants tended not to read the instructions, and we ended up with many participants getting stuck in the tutorial because they did not know what they were supposed to be doing (not having read the instruction).

In each game, participants were tasked with changing the emotion of 4 characters. They were told what the goal emotion was. As they played the game, they were not given any direct feedback when they accomplished each goal. The expectation was that participants would be able to track this change on the characters who were visually expressing their emotion and be unable to accurately track what was going on with the other characters.

After each game, the participants were asked a series of questions about their experience. We had learned from the Return Tablet game to be more careful about the phrasing of the questions and this time we used an assessment tool developed by Vermeulen *et al.* [81] to ask the participants about character believability, presence and agency. We also told the participants (after they completed the set of questions) how they had done in that game (how many characters they had correctly changed the emotions of).

7.2.2 Participants

We had 81 participants from the psychology 104/105 classes at the University of Alberta during the Fall 2013 semester. There were 24 males and 57 females. They had an average age of 19 years (min 17, max 28). Twenty-seven percent self-identified as game players (indicated they played at least once a week).

7.2.3 Results

The results were disappointing. There did not seem to be any differences between the two games. This was odd, as we had expected there to be a noticeable difference between characters that visually show emotion and those that do not. As we further analyzed the results, we realized what was happening. Our behaviour architecture allowed characters to display specific emotions so that participants can accurately identify them. However, it does not have any mechanism to indicate that a change in emotion is occurring. The participants' comments from both games indicated that many did no even notice that there *were* different types of characters between the two games.

- I was not sure if I was actually changing the moods of the other characters. Because of that it made the game less appealing.

- The emotional response is also limited because the responses are only visual, there is no change in action, facial features, or even sound to listen to changes in tone
- I found it difficult to identify often whether the charachter was angry or scared when I would push them or surprise them they would often say Ahh! which is indicitive to both, perhaps if I had heard the tone of their voices it would have been easier to identify their emotions
- It was harder to tell if i had completed the task of changing their emotions. So I was confused if I was doing it right or not. It was just better when you could easily tell what their emotions were.

The last comment above was particularly telling, because this participant had played the emotion version first and made this comment after playing the version without emotions.

We also had multiple participants refer to how the game was similar to life:

- it reminds me of sims and i like how you can directly influence their emotions by my actions
- It's a realistic game, where one can observe what's happening in the game in real life.
- I would play a similar game again because I enjoyed playing a part in the emotions of the characters on the screen. Ir was very life-like.

Finally, there were some who mentioned that they found it fun and satisfying.

- It was satifsying to see the impacts of my actions on the different characters in the game. I felt very much in control.
- my actions clearly affected the characters it was satisfying to find hidden interactions
- I would play this again because I could actually see the characters changing their moods as a result of my actions. This made it worthwhile to play.

Although our emotional architecture was successful in conveying the current emotions of characters, it was unsuited to conveying emotional change. Without key indicators of emotional change, participants had to rely on their instincts to determine if they thought a character had witnessed enough 'happy' incidents to turn the character happy.

7.3 Conclusion - Lessons Learned

Although the experiments described in this chapter were unsuccessful in objectively measuring differences between characters who did and did not display emotional cues, we learned enough to conduct the successful experiment described in Chapter 8. We also learned a lot about our participant group and what we could and should expect from them. This section details many of those lessons along with some mitigation strategies.

7.3.1 Tutorials

Depending on the game being tested, it is possible that the participant pool will contain many participants who are unfamiliar with playing games and controlling a virtual character. There may also be new skills that you expect/want the participant to learn. Participants will need a chance to learn (and practice) these skills ahead of time and a tutorial can be the right place for this.

Lesson 1. *Computers have taught people to just click "next" on dialog boxes and not read.*

If the instructions you are delivering are shared between a few dialog boxes (or pages), it is important to try to prevent the participant from just clicking 'next' without reading. For our experiments, when each new dialog box becomes active, there is a three second delay before the possible response buttons are enabled. While this does not guarantee that the participants will read it, it does prevent them from immediately pressing 'next'.

Lesson 2. *Unless it is absolutely vital, remove the tutorial from the actual game or simplify the game environment as much as possible in the tutorial.*

During the tutorial, it is important that participants are focused on the specific skill they are supposed to be developing. If there are unnecessary game elements (characters, props, buildings, etc.) the participants may get distracted from the task. Make the environment as simple as possible so that the focus remains on the skill.

Lesson 3. *Realize you cannot teach everything and only teach the bare minimum required.*

It is difficult to get participants to read even the shortest amount of text. The best way to guarantee they read what you wrote is to trim the text down as much as possible. If anything you are teaching is possible for participants to easily pick up (without compromising the task they will be required to do) leave it out of the tutorial.

Lesson 4. *Some participants do care - give them the tools so they can participate fully.*

Ensure that these participants have a chance to develop the skills they want and need to succeed. We discovered that giving these participants a choice to remain in the tutorial after it is completed was beneficial. If a participant was uneasy about their ability to control a character, it allowed them to get extra practice before entering the actual game. Allowing them to practice means that results of their gameplay will be less likely to be influenced by their lack of experience.

7.3.2 Game Play

During the experiment, make sure the process is as easy as possible for the participants. Build help right into the game wherever you think there may be a misunderstanding in the process.

Lesson 5. *Participants' memory is unreliable.*

From our experiments it was very clear that the participants had a very difficult time keeping track of our characters, even though we tried to make them look different from each other by varying hair style, hair colour, skin colour, type of clothing, clothing colour and gender. Even what we thought was distinctive information (one of our characters was carrying a fishing pole) was insufficient to prevent some players from confusing this character with others. Therefore, in our next experiment, we provided a list of the character names on screen during the experiment, allowing participants (if they so chose) to keep track of the characters they had and had not interacted with. Providing "in game" memory aids is beneficial.

Lesson 6. *Incorporate a mechanism to determine if participants are* actually *participating.*

Based on interaction with our participants as well as comments and submitted results, we knew that many of our participants were more concerned with getting the experiment over with versus actually participating in the experiment. It became clear that it is essential to include some mechanism to identify these participants.

7.3.3 Questions

Asking participants questions about their experience is an essential part of most studies.

Lesson 7. *Need to know* what *participants are comparing their answers to.*

Determine whether your question requires a base knowledge or definition. For example, we asked participants if the "characters were believable." We failed to provide a definition of believable, therefore, in return, we received results that could not be compared. It is important to provide participants with the exact knowledge you want them to use.

Lesson 8. *Disable questions until previous one has been answered.*

The same way that participants would just press next, it was easy for participants to not realize that they had skipped a question. While we would not enable the "next" button between screens until all the questions had been answered, we found it was also beneficial to only enable questions one at a time. This helped the participant focus on the specific question being asked and removed distractions.

Some of the above lessons may seem very straight forward and, in hindsight, many are. However, it is easy to overlook them when designing an experiment. From the above lessons, we were able to successfully design a game and experiment for objectively evaluating our emotional characters against neutral characters. The experiment is described in detail in Chapter 8. From that experiment, one of the most shocking results with respect to lessons learned was when we discovered that approximately 35% of our participants did not meet our bar of *actually* participating.

Chapter 8

Guess Who? - Study Five

Based on our experience from the previous studies we designed a new experiment to measure the impact of emotional character gaits and incidents in a game. The previous attempts inspired the following new design constraints:

- Players need to observe the characters in order to achieve the game objective.
- We need a mechanism to determine if participants are *actually participating*.
- We require a simple method for controlling the character and interacting with the game world.
- We require a tutorial whose completion would ensure that the players had mastered character control enough to move about and play the game successfully.

From these additional criteria, we created a game inspired by the popular board game *Guess Who* [41]. At the start of the game the players are informed that someone has been harassing the ducks in the park, and that they needed to figure out who it was by talking to the characters. The characters in the park were evenly divided between emotional and neutral and between male and female. The emotional characters would provide a clue to help eliminate potential suspects, while the neutral characters would just say something generic, but not helpful.

This experiment was designed to answer three questions:

1. Does the addition of emotional cues expressed through non-verbal, non-facial behaviour on virtual characters increase character believability in a game setting with many characters?

2. Will players learn to recognize emotions from emotional cues as they play?
3. If we only use a small number of extra scripts and animations to provide emotional cues, will they still be recognizable and learnable?

For question one, we expected that our emotional characters would be perceived as more believable than the neutral characters, as they would be observed acting and reacting to other characters and objects within the park. We knew that putting these characters into a game would likely decrease the accuracy of our participants, since they would no longer be guaranteed to observe exactly the emotional incidents and gaits we wanted them to see. However, we still measured success at identifying the emotions to be a minimum of 70% accuracy. In the game, how quickly participants were able to learn to identify our emotional cues would be necessary to be able to achieve this 70% accuracy. Finally, games need to provide characters with a minimum of a walking and/or idle animation. The better the character behaviour in games, the more likely characters will also have a minimum of some reaction animations and a behaviour script controlling their basic movements. Question three asks if the implementation we used, which contains only a small number of scripts and animations, would be sufficient to provide positive answers to the first two questions. If the answers to the first two questions were no, we would need to repeat the experiment using more scripts and/or animations.

8.1 Implementation

The new game takes place in the park we used previously, shown again Figure 8.1. The park contains a path that characters walk, deciding (randomly) at each intersection to go forward or to make a turn. They will never go backwards. The park is populated with a variety of objects that the characters can interact with, such as benches, picnic tables, newspapers, cups, rocks, and ducks.

In the implementation of our system of emotional gaits and emotional incidents for this experiment we used four gait animations and 12 reaction animations. We only used four gaits in this experiment instead of all five, as all the characters in the park were either neutral, sad, angry or afraid (there were no happy characters).

An emotional incident is a collaboration between props in the park, character emotions, and the reaction animations. For this experiment, there were nine different possible emo-

Figure 8.1: The park where the game takes place.

tional interactions. These included angry characters who kick soccer balls, sad characters who sit on benches, and afraid characters who shoulder check when another character approaches from behind.

Because the characters walk the park randomly, each game experience in the park is unique. The emotional interactions depend on which characters pass each other, whether a character is already engaged in another interaction, and whether the probability 'spin' causes the character to perform a given interaction. This means that a player may witness lots of interactions or very few. The chance of a player witnessing an emotional incident can be increased by adjusting the probability factors and increasing the number of emotional characters. In the end, for this experiment, we left the probability factors for each emotional incident at 100%, allowing every possible interaction to occur.

8.2 Experimental Design

What happens when characters who express emotion through emotional cues are placed into a game environment? Will players identify the expressed emotions? Will emotional characters be perceived as more believable than non-emotional characters?

In the end, our study included five activities. The experiment started with a tutorial, followed by two play throughs of the 'Guess Who' game, then a third 'game' where players must identify each character's assigned emotion and, finally, a brief survey. After each game segment, participants were asked to answer questions about the game. The participants were informed that the emotions of all characters were changed after each game.

8.2.1 Tutorial

A 'maze' tutorial (Figure 8.2) taught participants how to move their character using the mouse and 'WASD' keys. The goal of the tutorial was to locate a second character within the maze. After finding the character, they were taught how to interact with (talk to) a character. Once they finished, they were given the choice to spend extra time in the maze or to move on to the next part of the study.

The floor of the maze was coloured because, in the initial design, a player could end up in a loop if they chose to navigate by always turning left. By colouring the floor, participants were given a clue that they had already been to a specific area of the maze before.

8.2.2 Games 1 and 2 - Guess Who

In Guess Who, participants were tasked with identifying the character who had been harassing the ducks in the park. In order to uncover the identify of the perpetrator, participants were instructed to ask the characters in the game if they saw what happened. All of the characters who had an assigned emotion would provide a clue, while those who were neutral would just give a generic response (for example: "Someone was kicking ducks? What's wrong with people."). An example screenshot is shown in Figure 8.3. In order to identify the suspect (and receive enough information from the clues), the participant needed to talk to three emotional characters to collect three clues.

Participants were told to think carefully about who to ask:

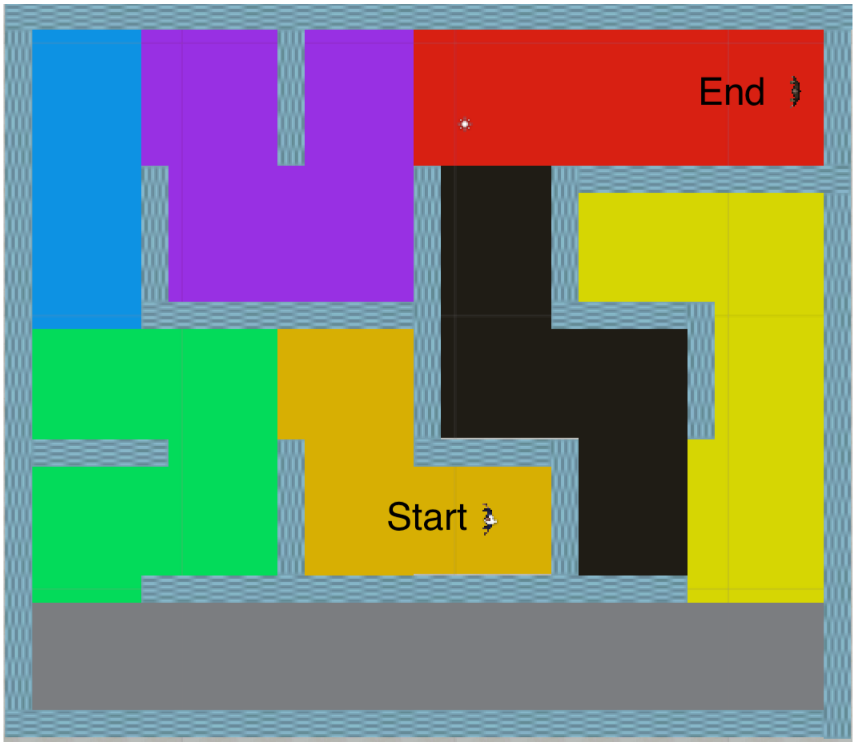

Figure 8.2: The tutorial maze with floor colors to aid memory during navigation.

Figure 8.3: Example of conversing with a character in game 1.

> "Not everyone saw what happened (or remembers), so use your questions carefully. You should question those who may have been affected by the incident."

We expected that the instruction "those who may have been affected" may result in players talking to the characters who exhibited negative emotions (afraid, sad and angry) as a result of witnessing the ducks being harassed. However, participants actually used five different strategies as described in the results section.

Once a participant had received three clues (which could have taken more than 3 interaction since some interactions did not provide a clue), the game would switch to the "accuse" mode. At this point, participants were no longer able to question the characters. Instead, they had to make their decision as to who to accuse.

As participants played, they were also shown their current 'score' in the game. The score was decreased every time the participant questioned a character that supplied no useful information to encourage them to think about who to question. If they accused the wrong character, their score was lowered completely to zero.

Each game consisted of three possible visual clues. A visual clue could be used to eliminate characters based on an easily observable characteristic. We had six visual cues, split into two groups of three. Group one consisted of short sleeves, wearing sunglasses, and black hair. Group two consisted of red shirt, holding a newspaper and blonde hair. During each game session, a player would receive two of these clues. Which clues they received was determined by who they talked to. Each clue eliminated four characters. And each group of four eliminated characters consisted of a neutral male, an emotional male, a neutral female and an emotional female. An example of a clue is: "I couldn't see their face, but I know they didn't have black hair."

Each emotional character had two visual clues that they could share. The first visual clue they would share would eliminate themselves as a suspect. However, if that clue had already been given, then they would use their back-up clue.

The final clue that a character received always specified the gender and emotion of the guilty party. For example "Yes, I saw him. He looked really sad." Even without the visual clues, it would be possible to identify the guilty party from this clue alone (there is only one character of each gender/emotion combination). When it was time to give the final clue, the game would choose a guilty party based on the clues that had been given and characters the player had talked to, such that the guilty party was not already eliminated, nor had she or he been talked to. The design and number of clues guaranteed that there was always an emotional character left that met this criteria.

Because of the structure of the game, we know that if a participant was actually *trying* to play the game, then after the first clue, they would have eight characters left, and after the second clue four characters left. At this point, the final four characters would consist of two males and two females (with each group consisting of one neutral and one emotional character). *No recognition of emotion was required to reach this point, since the first two clues always eliminated individuals using visual clues such as hair color, shirt color or whether they wore glasses.* The final clue would eliminate two of these characters based on gender. The participant was then left with two characters to choose between. Of the final two characters, one would be exhibiting emotional cues and the other would not. At this point, the participant was required to recognize the emotion to succeed in the game or they could pick randomly and be correct an expected 50% of the time.

8.2.3 Game Three - Identifying Emotion

Figure 8.4: Example of labelling characters in game 3.

The last interactive phase in the study was to play a third "game". In this "game", participants were tasked to find and identify the emotion of all 12 characters in the park. Figure 8.4 shows a screenshot from this game. Their choices were: Neutral/None, Happy, Sad, Angry, and Afraid. Participants were allowed an unlimited number of times to change their mind. Once all the characters had been labelled, and they were satisfied with their choices, they could exit the game. Participants were reminded before starting this game that the emotions of the characters had been changed since they played the previous game in the park.

8.3 Participants

We had 90 undergraduate students participate in return for course credit. Unfortunately, some students approach these studies with the single goal: "how quickly can I get through the study." Since we had included a mechanism to detect these participants, we used it. We removed the data from any participant who did not accuse one of the characters who was in the "final two" suspects. No identification of emotion was necessary to deduce the "final two". This allowed us to focus on the students who attempted to play the game as designed. After doing this, we were left with 58 valid participants.

The validated group contained 29 female and 29 male students with a mean age of 19.6 years old (min 18, max 29), who are primarily in first or second year (mean 1.7, min 1, max 5) and 43% of them are gamers (indicated they play video games at least once a week).

8.3.1 Experiment Limitations

Similarly to our earlier experiments, our group of participants cannot be generalized to the general population or even the game playing community.

8.4 Results

We refer to the first instance of the Guess Who game as game one, the second instance of the Guess Who game as game two and the emotion identification game as game three.

We were specifically interested in whether or not participants found our emotional characters more believable than the neutral characters. We expected the participants who interacted with more of our emotional characters than the neutral characters would find our characters more believable because the characters they were observing would likely be engaged in acting and reacting to the game environment (while the neutral characters just walked the path forever, never acknowledging other characters or game objects).

We ran an ANOVA to see if there were any effects (interaction or otherwise) on participants' ratings of believable characters (see Appendix D) by the participant's gender, frequency of game playing, winning the first game, or the number of characters talked to. The results suggest that the number of characters the participant talked to affected how believable the participants found the characters.

Because we had the data, we also ran ANOVAs on participants' ratings of presence, agency, and enjoyment (also in Appendix D). Both character believability and presence were also affected by an interaction effect of the participant's gender, how frequently they play games, and the number of characters they talked to.

8.4.1 Believable Characters

To measure character believability, we used the assessment tool developed by Vermeulen *et al.* [81]. It is a series of likert-scale questions that participants answered after game one. We modified the questions slightly by replacing "the environment" with "the game."

For character believability, we found that the number of characters a participant talked to was significant in relation to how believable they found the game characters. Specifically, participants who primarily interacted with characters who exhibited emotional cues found the characters more believable (p=0.03 in Table 8.1). Since each participant had to talk to exactly three emotional characters during the game (in order to receive their three clues), participants who only talked to three or four characters talked with at most one neutral character, whereas participants who talked to five or more characters talked to at least two neutral characters. In other words, for participants who talked to 3 or 4 characters, 75 to 100% were emotional characters. For participants who talked to 5 or more characters, at most 60% of the characters were emotional characters. By comparing the reported believability of these two groups of participants, we were asking whether interacting with characters who display emotion increases the believability of the game. From Table 8.1, the answer is yes.

	Believable Characters	Presence	Agency	Enjoyment
3 or 4	2.20	2.11	2.41	2.13
5+	1.73	2.23	2.60	2.03
T-Test	**0.03**	0.34	0.15	0.36

Table 8.1: Character believability, presence, agency, and enjoyment that participants experienced based on the number of characters they talked to. Values are likert scale averages from 0-4, with higher being better.

8.4.2 Playing Strategies

We asked the participants how they decided on which characters to ask for information. We found that they used five character selection strategies: close to the player; based on the clues; random; based on character behaviour/signs of emotion; and close to the pond and ducks. Figure 8.5 shows the percent of participants who used each strategy. The columns total more than 100%, as 31% of the players used a combination of at least two strategies. The unknown strategies resulted from two participants who answered with "yes" and "Leigh" (the name of one of the characters).

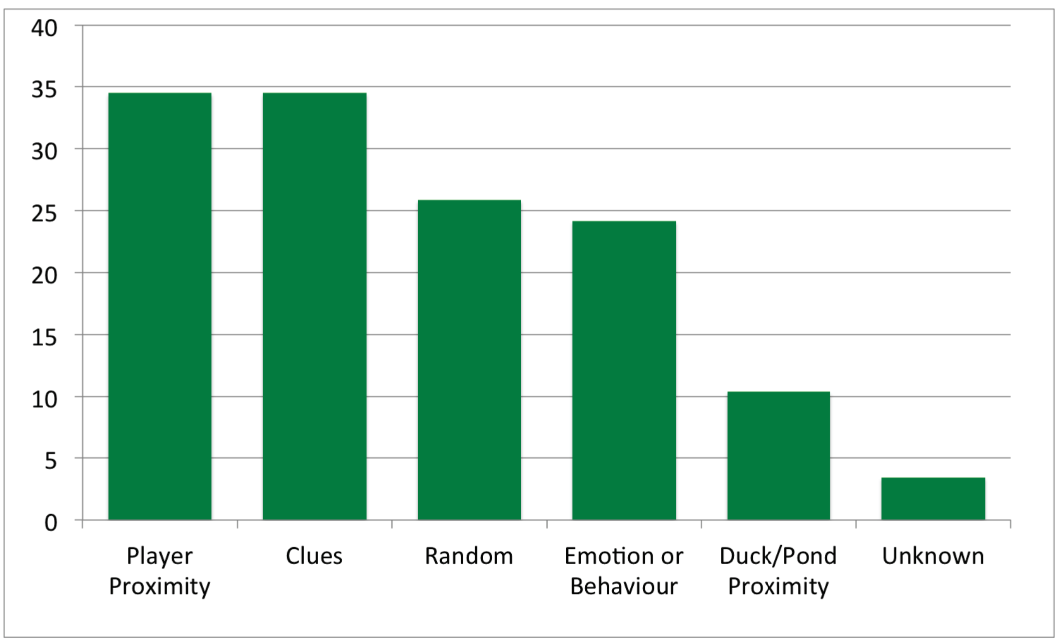

Figure 8.5: Strategies used by participants to decide who to talk to (in percentages). Note that the heights do not add to 100 as a third of the participants used more than one strategy.

We found that participants whose interaction strategy involved identifying emotional characters found the game characters more believable than participants who didn't pursue this strategy (p=0.012). While this may seem immediately obvious, these participants were not strictly the ones who talked to more emotional characters. In fact, one participant stated that they purposely "choos[e] people with neutral emotions" to question.

8.4.3 Learning

We wanted to measure whether or not participants improved their ability to identify emotions as they played the games. Figure 8.6 is a graph showing the percent of participants who were able to accurately identify character emotion in each game. In the first two games, participants only needed to identify the emotion correctly once, at the end of each game, to determine which of the final two characters to accuse. Since the players we have included in the data were attempting to play the game according to the clues, by the time they needed to use emotion to choose a character (clue 3), they had narrowed down the possible set of suspects to the correct final two characters. This meant participants had a 50% chance of guessing the correct answer. This 'chance component' is shown on the graph as the yellow (or bottom) part of each column. Figure 8.6 shows that while participants were only ∼2% above chance (50%) in game one, by the time they finished game two they had increased to almost 30% above chance (50%).

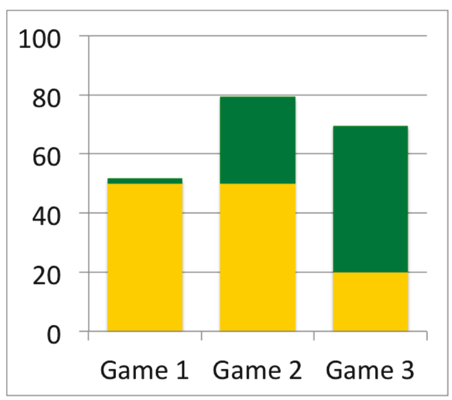

Figure 8.6: Percent of players who were accurate in identifying character emotion in each game.

For game three, participants were required to identify the emotion of all 12 characters. For the graph, we focused on their accuracy on the six emotional characters (the other six were neutral). Unlike in the first two games, players had to choose between five possible labels (neutral/none, happy, sad, angry, and afraid). This means if they were assigning the labels randomly they would have a 20% chance of being correct (thus the much smaller yellow "chance" part on the column). The results show they were identifying the emotion correctly ∼50% above chance (20%). These results illustrate that players learned to identify the characters' emotions as they played the games.

8.4.4 Time Spent Playing

Participants spent an average of 11.5 minutes playing the three games in the experiment (total average time was 21.5 minutes when including answering questions and the tutorial). However, it is clear from Figure 8.6 that it did not take participants long to pick up on the emotional cues that were presented. This suggests that, while there is likely to be a learning curve when emotional cues are added to a game, players will quickly adapt and be able to pick up on the cues.

8.4.5 Identifying Emotion

The third game was included so that we could determine the accuracy and precision of identification when characters were placed in a more realistic game context. In our original experiments, our average precision and recall rates were in the high 80s and 90s. However, it is important to remember that those tests were done in isolation, where participants were asked to watch a character in a specific situation, one at a time. Therefore, there were no distractions of other characters wandering around and performing their own interactions, and the specific interactions the participants saw were highly scripted (removing the possibility that participants would not see any interactions).

	Neutral	Sad	Angry	Afraid	Happy	RSum	Recall
Neutral	**189**	25	15	12	107	348	0.543
Sad	15	**83**	7	7	4	116	0.716
Angry	11	1	**70**	15	19	116	0.603
Afraid	12	5	6	**89**	4	116	0.767
PSum	227	114	98	123	134		
Precision	0.833	0.728	0.714	0.724		0.750	0.657

Table 8.2: Confusion Matrix for Game Three. Bottom right values are average precision followed by average accuracy.

The overall confusion matrix for the "in game" results are shown in Table 8.2, while Table 8.3 shows a comparison of recall rates when the results are divided by gender and gaming ability. While our results were not as high as our initial study, they were still well

	0verall		Female	Male	Gamer	Non-Gamer
	Precision	Recall	Recall		Recall	
Neutral	0.83	0.54	0.47	0.61	0.63	0.47
Sad	0.73	0.72	0.66	0.78	0.84	0.62
Angry	0.71	0.60	0.60	0.60	0.74	0.50
Afraid	0.72	0.77	0.72	0.81	0.80	0.74
Average	0.75	0.66	0.61	0.70	0.75	0.58

Table 8.3: Overall precision and recall results as well as recall results by gender and gamer status.

above chance (0.20). Again, we were aiming to exceed 70% recall rates in participants' identification. From the table, participants were able to surpass 70% when identifying sad and afraid characters (71.6% and 76.7%), but not angry characters (60.3%). Even in the more vibrant and chaotic game world filled with distractions, the precision and recall for identifying specific emotions are still high enough to be used by game designers.

An ANOVA analysis showed that players' ability to identify emotion was dependent on their gender, gaming experience and the time they took to play game three. The results for gender and gaming experience can be clearly seen in Table 8.3. When we looked at the time spent in game three, we found that participants who spent less than three minutes had an accuracy rate of 65%. On the other hand, participants who spent over six minutes had an accuracy rate of 73%.

Figure 8.7 provides a more detailed breakdown of the labels used by participants over the groups of characters. Each column represents the actual emotion of a group of characters (two characters for sad, angry and afraid, and six characters for neutral/none). The column itself is broken into groups representing the percentage of players who choose to use each of the five possible labels. Because there were five options, if participants were randomly assigning labels, we would expect a 20% accuracy rate (represented by the purple band at the bottom). As the purple band overlaps the correct emotion label in each of the four columns, it is easy to see that each label was used correctly much more than chance would suggest.

8.5 Conclusion

This study was designed to answer the following questions:

1. Does the addition of emotional cues expressed through non-verbal, non-facial be-

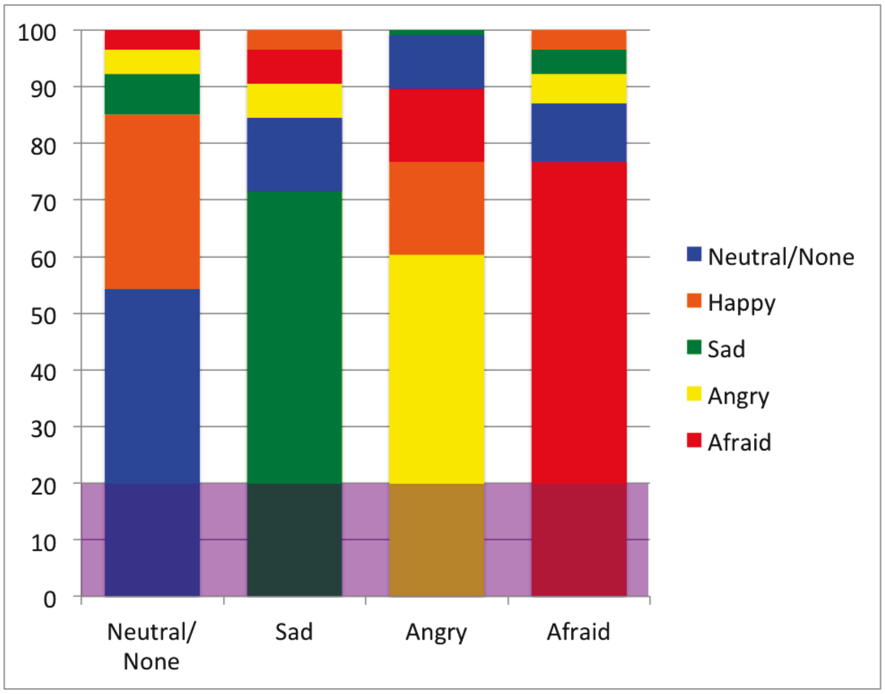

Figure 8.7: Percentage of players who used each label. X-axis is the correct character emotion. The band, across the bottom, represents the percentage expected if done by chance.

haviour on virtual characters increase character believability in a game setting with many characters?

2. Will players learn to recognize emotions from emotional cues as they play?
3. If we only use a small number of extra scripts and animations to provide emotional cues, will they still be recognizable and learnable?

For question one, we found that the answer was yes. The more characters a player talked to who displayed emotion, the more believable the participant found the game characters (p=0.03).

For question two, we found that learning occurred rapidly. Even though participants only spent an average of 11.5 minutes playing the games, participants increased their accuracy above chance as they played. They improved from ∼2% above chance (50%) in game one, to ∼30% above chance (50%) in game 2 and ∼50% above chance (20%) in game 3. The participants were able to achieve higher than 70% recall rates in identifying sad and afraid characters, but struggled with the angry characters (60.3%).

For question three, we obtained positive answers for questions 1 and 2 even though we implemented our entire experiment with the addition of only three emotion-specific gait animations (four if you include neutral, but games would need at least this one) and 12 reaction animations. Our results also show that emotional cues through emotional gaits and incidents are an effective way to convey information about a character's emotion. This shows that a low investment can produce measurable improvements in believability.

Finally, we share this response from one participant, which summarizes that these emotional cues can make a significant impact:

> It was surprisingly easy, I thought, to read emotions from the body language without any facial expressiveness, although they were pretty simple models I suppose. And yet, I can't really say I remember people using something as simple as slumped shoulders in any RPG I've recently played.

Chapter 9

Easy construction of animation and incidents - Study Six

This chapter asks and answers the question whether it is possible for non-experts to create the emotional animations and incidents that are necessary to use the techniques described in this dissertation. Specifically, the goal is to re-do the user study about identifying emotion based on isolated emotional gaits and incidents using non-professionally produced content. Creating games without professional experts is a hot issue in recent years, and there are many tools out there to aid budding game designers.

The affordances offered by these tools vary widely. Some tools allow the user to write 'actual' code, assisting by providing an expanded API for common tasks (like pygame). Several of the tools have created their own visual programming language, often using blocks or puzzle pieces that are snapped together. Some of the tools are all-inclusive, providing the art, game environment and coding environment all in one place. Some allow for 3D game development while others focus on 2D. To discover whether the emotional animations and incidents could be created by non-professional animators and programmers, we needed to first select a tool. We studied the landscape of available tools before selecting one to use in our study.

This variety in tools has provided a lot of choice to a new game designer who wants to make a new game. Last fall, Microsoft continued their foray into this market with the release of Project Spark, building off their earlier entry program called Kodu. Project Spark provides a rich stylized environment, full of 3D characters and objects [54]. Users can create their own sculpted terrain, place objects across the map, and quickly build a game by relying on the precoded behaviours included on many objects and creatures.

This chapter presents an overview of our experience working with Project Spark to recreate the user study on techniques to convey emotional cues through character actions. This study was designed to discover whether or not the relatively small effort that required professional animators could be replaced by the motion capture tools provided in Project Spark. While our resulting 'game' (scenes) did not require the use of all the components in Project Spark (such as the player controls), we did use most of them. So, in addition to a discussion of the results of our study, this Chapter also provides an early evaluation of the capabilities of Project Spark.

9.1 Related Work

There are many tools available that enable non-programmers to create their own games. They widely vary in how 'code'-like the language is, the difficulty in getting started, the quality of the graphics and animations available and the age-range they are designed for.

9.1.1 Kodu

Kodu was Microsoft's first big foray into game development tools for non-programmers [55]. Project Spark is the 2nd generation of Kodu, and uses a very similar coding style, although with many more options. 'Koding' is done by choosing tiles from a 'kode' wheel to answer two questions: When and Do. *When* specifies the game event that should occur (such as bumping an object) and Do specifies the action that should be taken (move toward an object). Figure 9.1 shows example 'kode'. While the graphics in Kodu are commercial level, they are also very stylized. And so while Kodu is designed for everyone, the graphics and game environment are more appealing to young children, and it has been used in code.org's *Hour of Code* initiative. Many of the tutorials and help material focus around curriculum ideas on how Kodu can be used to teach programming in a classroom.

Figure 9.1: Kodu [56].

9.1.2 Scratch

Scratch, from MIT, was designed specifically for children [57]. Coding is done by snapping together colour- and shape-coded puzzle pieces (see Figure 9.2). It provides a 2D environment and a few basic pieces of artwork, although most content needs to be generated by the designer. It is most popular with users between 8 and 16 [68]. One of their goals was to make the experience *social*, and so they included an easy way for users to share their games.

9.1.3 Alice

Alice, from CMU, is another tool designed for children [16]. However, Alice was designed with a goal of reaching out to girls and, unlike Scratch, includes a built-in 3D environment. Coding in Alice is done by selecting from a collection of pre-built methods (see Figure 9.3). Alice does not try to hide the code from the designer. All the 'pieces' used to write code in Alice look like they could be code from a 'real' programming language. And, in Alice 3.0 designers can switch to writing code in java, instead of their slightly simplified language. While Alice includes a large library of 3D objects for designers to use, the quality is still very behind commercial games, making it not appealing to older users.

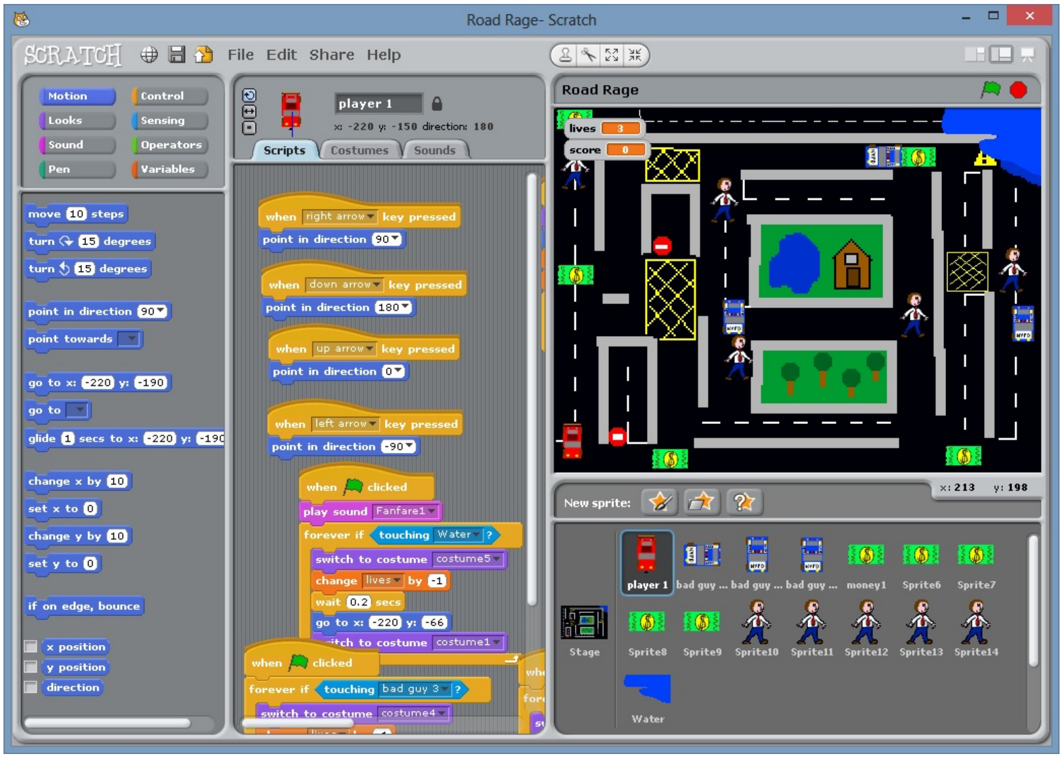

Figure 9.2: Scratch [67].

9.1.4 Kano

Kano, well known for its successful KickStarter campaign, is a build-your-own computer kit focused around a Raspberry Pi [46]. However, within their custom operating system, they have included a customized MineCraft game with a simplified coding language (see Figure 9.4). The coding language appears visually similar to the one in Scratch, focusing on puzzle pieces that snap together. However, they also allow the users to see the actual code generated by the code blocks. In order to have users slowly learn the tool, instead of diving into everything from the start, they've built a series of challenges that designers must go through to unlock additional code blocks.

9.1.5 ScriptEase

ScriptEase, from the University of Alberta, takes a different approach, by separating the coding experience from the design experience [10]. By doing this, ScriptEase allows users to choose their game engine (current choices include Neverwinter Nights and Unity), which allows users to produce games that are much closer to commercial quality in terms of

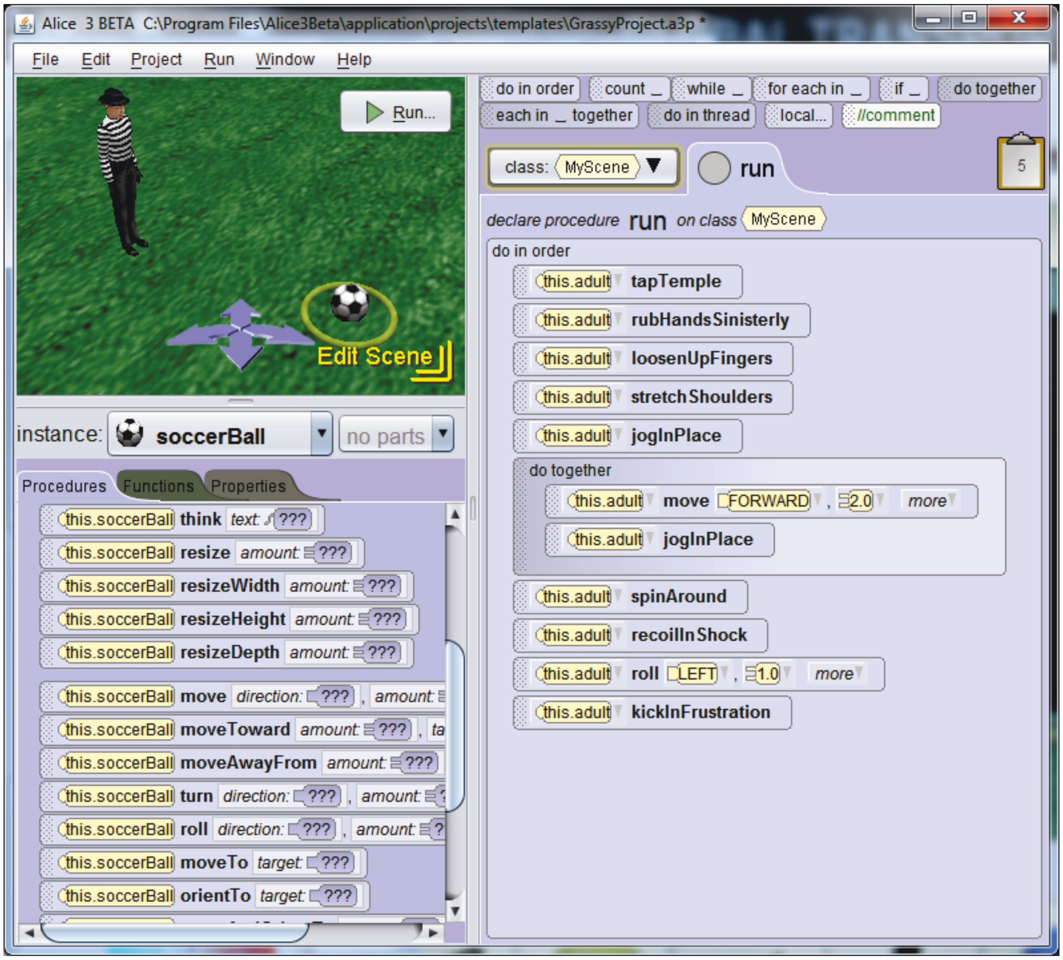

Figure 9.3: Alice 3.0 [44].

graphics. However, this also raises the difficulty level for new users (for example, while Unity can make high quality games, it is not a beginner-friendly tool). While ScriptEase uses drag and drop pieces of code, they have been worded such that they do not resemble a coding language (see Figure 9.5). Instead of building an 'if' statement, users build a 'Question'. Individual 'effects' (code pieces that change the state of the game in some way) are worded to resemble human readable sentences (for example, 'Set the <door>to <unlock/lock>' where <> indicates a place for variable).

9.1.6 Gamemaker

GameMaker Studio, by YoYo Games, produces commercial level games [87]. While it claims it is for everyone, from novices to experts, the tool's game design portion can be quite

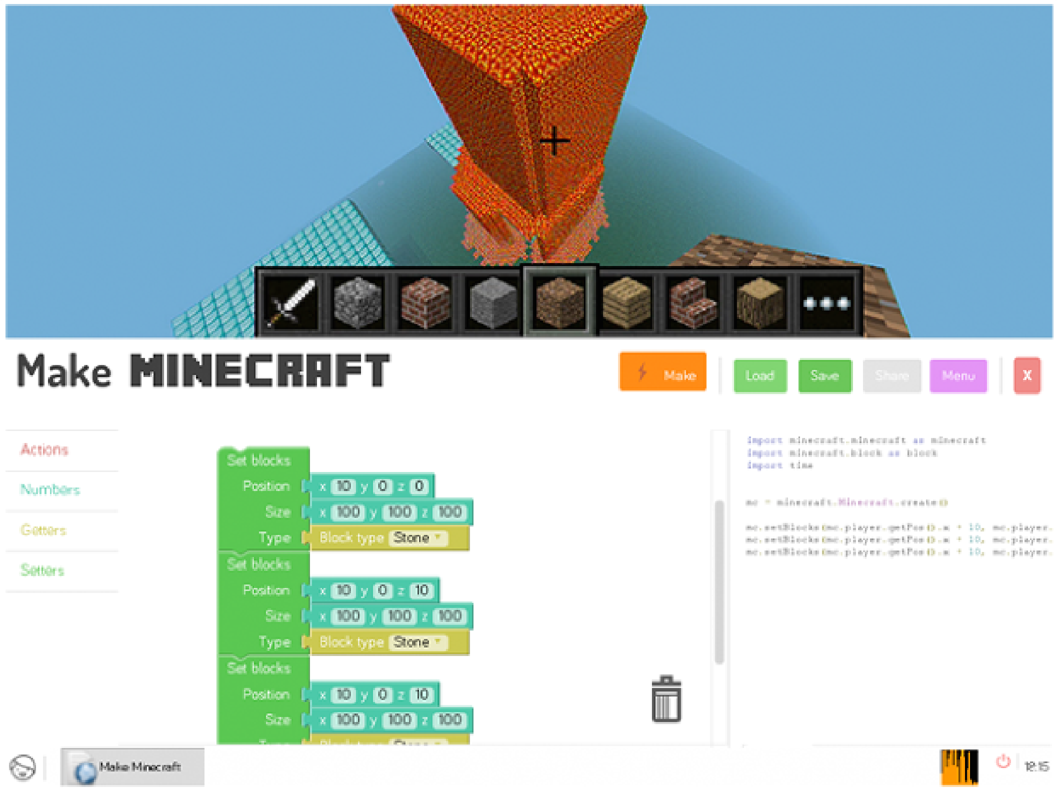

Figure 9.4: Coding Minecraft in Kano [35].

challenging to use and does not include much art (although there is an asset store where you can buy more). The actual coding is done by choosing from a series of visual menus that, similarly to ScriptEase, result in natural language looking sentences (see Figure 9.6). Users must select an event (such as a button being pressed on a keyboard) and then supply one or more actions to occur (such as moving in some direction). GameMaker is designed for making 2D games, although it is possible to simulate some level of 3D functionality.

9.2 Lessons Learned in Using Project Spark

The following section details our experience with using Project Spark to implement a series of short scenes for our experiment about identifying emotion from emotional animations and incidents. The scenes we created were not played, but watched as videos and so there are no player controls involved. However, the scenes involved additional characters who act and react to each other.

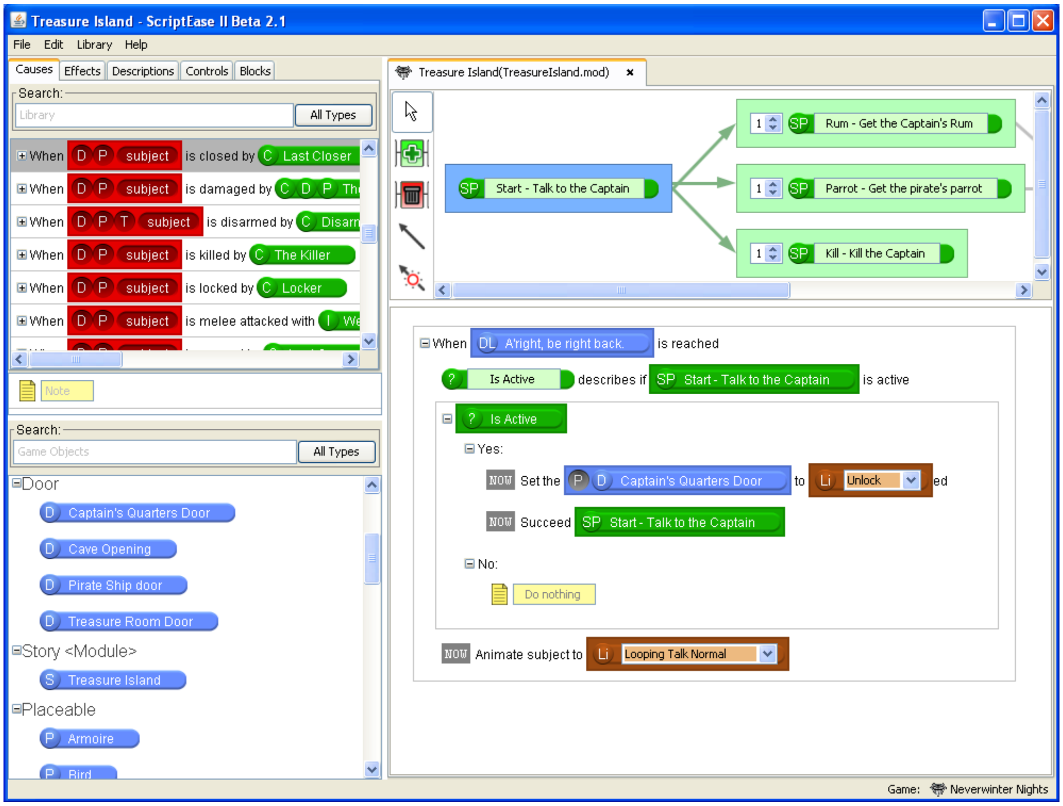

Figure 9.5: ScriptEase.

9.2.1 Building your map

In Project Spark, there are two ways to build the map where the game takes place: build it on your own starting with the "empty

grey land or have Project Spark create it randomly from a few set options. There are advantages and drawbacks to each method; the main one is the level of control you have. The ability to customize the created map is provided.

The main advantage of the on-demand map generation is that it creates your choice of one terrain (such as either mountains and valleys or hills), then paints it (with your choice of one texture - winter and ice or a green forest) and populates it with a variety of trees and creatures. The disadvantage is that if you are looking for a specific set-up (in our case, we needed an area of the map where a character can walk a straight path while being involved with the preset incidents), there's may be nothing suitable. The included creatures in your selected map can also introduce unexpected behaviours, as most objects in Project Spark come with built-in behaviours.

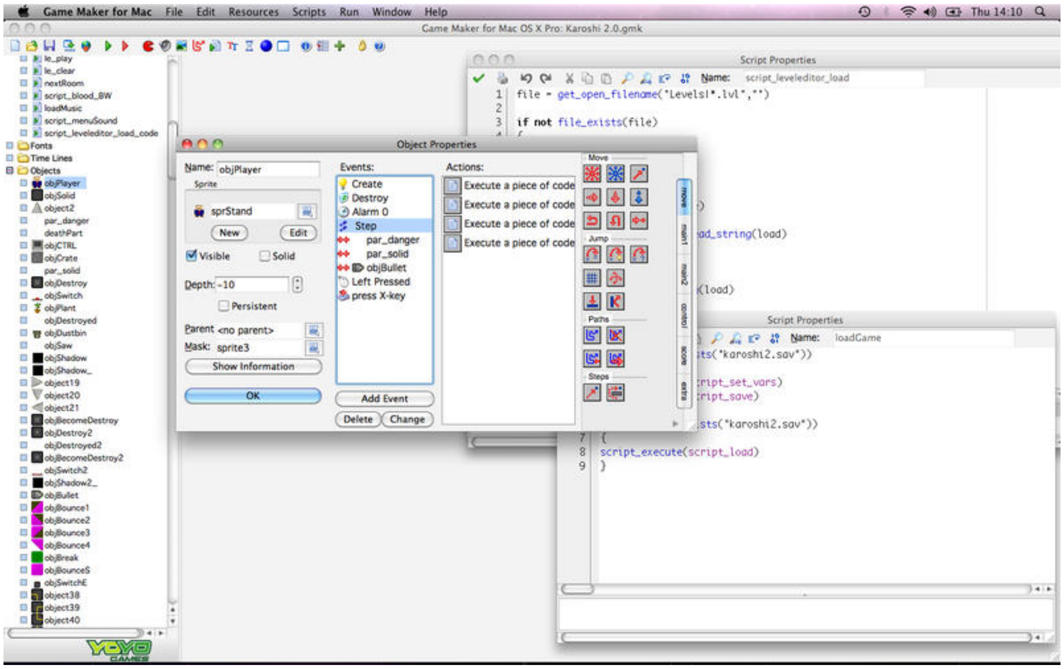

Figure 9.6: GameMaker [88].

Building your own terrain guarantees precision and flexibility. However, if you are impatient, (the undo feature is not easy to use), this can become an exercise in frustration. Building your own, it is possible to create a map with floating islands, to tunnel through mountains and to introduce lakes and rivers. Your world can quickly transition from ice and snow on the mountain top to a desert down in the valley.

In the end, we started with a created map and customized it. This involved removing extra trees (they got in the way of the camera) and deleting all the preset behaviours on creatures and the creatures themselves. Figure 9.7 shows a top down view of our map.

9.2.2 Koding

To create your own game in Project Spark, you use their version of coding which they call 'koding' since you write 'kode.' Koding is done by assembling kode tiles, where a kode tile may be a variable, an assignment operator or an action. This is done by opening up an object's *brain*, which contains all the kode controlling that specific object.

Each line of kode has two components. The first component is the *When* block and the second is the *Do* block, and both sides can be left empty. An example of kode is shown in Figure 9.8.

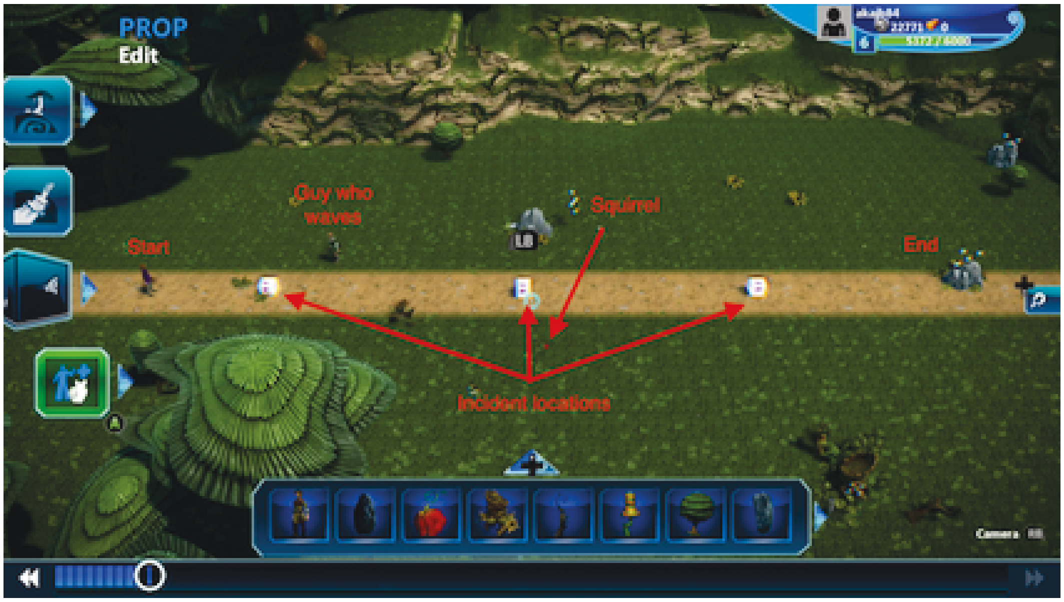

Figure 9.7: Map of our world in Project Spark.

The *When* component is used to put a condition on a line of kode. If it is left blank the line will always execute. The lines can be indented, so that a condition put on one line can be shared by all the indented lines below, similarly to how one would write an *if* statement in python with all contained lines indented below the condition line. Using the tile blocks accessible within the *When*, you can create conditions that are time based (countdown) or for conditions when a variable has a specific value, among others. You can also insert an "Ignore" tile to comment out a specific line.

The *Do* component is for the effect kode. Here you can change variable values, start animations, or tell a character to move.

To organize your kode, you can make multiple *pages* within a *brain*, and switch between them. The kode for switching is not at all intuitive, making this feature difficult to use.

Using the Tile Picker Wheel

Choosing the individual pieces of kode can be frustrating. The tiles are presented as a ring. Many of the options are categories that take you to an inner ring (which then potentially has "pages", if the options didn't all fit). Figure 9.9 shows an example of the tile picker.

Project Spark tries to be helpful by limiting your options based on the context of the tile's location (Are you in a *When* or a *Do*? Which tile is directly before this one?). However,

Figure 9.8: Project Spark 'Kode'.

depending on the ever changing context, the placement of the tiles is constantly shifting, making it difficult to rely on tiles being where you thought they were. Because the category names and contents are not particularly intuitive, you spend extra time looking through the options, searching for the one you need.

Each tile has a 'help' section associated with it. Some of these help sections are incredibly useful, with examples of how the tile could be used in a game. Unfortunately, many provide little to no information, leaving you to try and figure out the meaning by guess and check.

Debugging

Project Spark does not have explicit debugging tools. The best method available for debugging (and the recommended method) is to use the display commands to print out text and variable contents to the screen. While a useful method for debugging, the awkwardness of koding means adding and removing these extra lines is a lot of extra effort.

Figure 9.9: The tile picker in Project Spark.

9.2.3 Kinect Recording

One of the main reasons for selecting Project Spark over the other tools that we investigated, was the ability to use the Kinect to record our own animations. For previous studies, we hired an animator to make all the emotion specific animations. With the Kinect a designer is able to create animations using "acting motions" in addition to creating the emotional incidents using the koding.

Neon Orange Hats

Many of our animations required the character's head to move in a particular way. In some, it was turning to "look at" some object (you cannot specify a look target, so it must all be carefully simulated), and in others to look at the ground. However, it quickly became apparent that the Kinect has difficulty following head movement. If you turned around in a circle, the recording often became confused with the movement because you had done something impossible, i.e. rotated your body around while your head remained stationary. The Kinect did not guess/assume that the head rotated with the body.

We tried many solutions, including tying long hair up, adding/removing eye glasses, using shirts with different neck collars and a neon orange ball cap. The ball cap worked the best, although still provided only slightly better performance. Many of our animations took numerous recording trials and had to be slowed down (to give the Kinect a chance to keep up). Over-exaggerated motions were used with some success (hoping that the Kinect would notice that the head moved).

Hand and Arm Movements

We often used gestures expressed through hand and arm movements to convey emotional cues. Similarly to our previous study, we expected some participants to take special note of the hand movements "[I] focused mainly on the hands of the character when gauging emotions I found."

Regarding arm movements, the Kinect was generally able to create an accurate recording (although occasionally arms would slice into the body). However, the Kinect often assumed the hand was being held in a fist. Splaying fingers as far apart as possible often was not enough to overpower the Kinect's assumption that you were waving a fist around. This resulted in redoing animations many, many times. The farther the hands were from the body (out to the sides, not in front) the better odds at getting the correct hand position. In the end, some of our animations did end up with fists because, no matter how many takes we did, the Kinect would not budge.

Combining Animations

In real life, it is possible to walk down a street, occasionally drink from your coffee, and kick at that pesky rock that is in your way. With the Kinect, this is all *theoretically* possible, but not at all practical. Generally, if you stop walking to take a sip of coffee and kick a rock, you are able to do this all at the same time, by "blending" these actions together (your leg and foot move to kick, your core helps you balance, and the arm/hand with the coffee cup raises it to your mouth). Later, should you decide to stop and just kick a rock, you can easily do this while *not* drinking.

However, the Kinect does not allow blended animations. This is because every single recorded animation is a full-body animation. It is impossible to tell the Kinect to just record your arm or head. If you want to be able to drink and kick at the same time and also to just kick, then you need to record two separate animations. In the first, the character would perform the drink and kick animation simultaneously, while in the second, the character would perform only the kick animation.

In fact, with Project Spark you are limited to having only one animation play at a time. It is also *not* possible to stop an animation once started or to detect the end of an animation.

Facial Animations

The Kinect automatically includes facial animations as part of each recording. As we were not interested in having facial animations on our characters (we wanted them to have a neutral expression) we were worried this may cause a problem. However, it turned out that the Kinect had difficulty picking up on the small facial movements. In fact, for us, the Kinect felt that all characters should have one eyebrow raised and the other lowered. In the videos used in our study, we had the character situated far enough back and angled slightly away from the camera, so that even though our animations may include these odd facial expressions, it was not possible for the participants to observe them.

9.2.4 Walking

The most feedback we received was about the walk animation and it was not particularly positive.

- The way the person was walking was annoying.
- The abrupt stops in the females walking was kind of odd.
- Just a general comment but the movement seems a bit awkward. The arms look like they're only swinging in front of the body instead of naturally.
- the way she was walking was not normal
- The movements of the character seemed somewhat unnatural

Regarding the included walk animation, we had exactly one variable we could change – speed. However, it is supposed to be possible to create your own walk animation using the Kinect. Project Spark has an interesting feature for playing animations. If a character moves from their starting position during an animation, the character will stay at the final end position once the animation is over (instead of popping back to the start position). Therefore, theoretically you can have the Kinect record you walking, and then play that animation repeatedly to have a character use a custom walk animation.

In practice, this was too difficult because of the editing screen. Once you have finished recording an animation, you have one option for editing – cropping. For cropping, you have a slider tool that controls where to crop at the beginning. Once that is chosen, you get to choose the end point. It is incredibly difficult to control this slider accurately. If you want to make a walk animation, you will need to be able to perfectly walk two steps such that you can find a start and end point that match exactly. If you cannot, you end up with a character that takes two steps and then jerks as it jumps between the start and end positions of the animation, as you mimic the animation looping.

The lack of truly customizable walk animations left us mostly dependent on the emotional incidents for conveying emotional cues. We did use different speeds for the five scenes. The emotions, in order from slowest to fastest were: afraid, sad, neutral, happy, angry.

9.3 User Study

The purpose of our expedition into the land of Project Spark was to create animated scenes for our user study about emotional characters. For the study, the participants were presented with five video clips in a random order. They were asked to label each video according to the emotion they felt was shown. They were given five choices: Happy, Sad, Angry, Afraid, Neutral/None. We labelled the videos after shapes (line, circle, triangle, square, star) to prevent the users from making any connection between the label and the emotion. Users were able to watch the video clips as many times as they wanted and to change their answers. They were strongly encouraged to watch the videos from beginning to end.

In each video, a character would walk a set path and respond to three emotional incidents. First, a character waves at the walking character. Second, a squirrel runs across the path in front of the walking character. Third, the character stops on the path and responds to no external event (mimicking responding to an internal thought/emotion). The incidents and their specific emotion responses are detailed in Table 9.1.

9.3.1 Participants

We had 137 participants. They were undergraduate students taking a first year psychology class at the University of Alberta during the Fall 2014 semester. They were between the ages of 17 and 43 (mean of 19.7) with 91 females and 46 males. Thirty-five percent of the

Table 9.1: Emotional incident responses.

	Waving Character	Squirrel	Internal
Happy	waves	bends down to pet	jumps for joy
Sad	brief glance, sigh, then ignore	big sigh	big sigh, brief hand to forehand while looking down
Angry	raises fists and speeds up	kicks at squirrel	stomps foot
Afraid	small startle, then veers away	startle	quick startle, then side to side look

participants reported playing video games at least weekly ($\sim$ 19% of females and $\sim$ 70% of males).

9.3.2 Experiment Limitations

Again, our participant pool results in a very homogenous population in terms of age and education, distinct from the general population. This prevents the results from being extrapolated directly to the general population or the game playing community.

9.3.3 Statistical Techniques

In order to analyze the data, we used the same two statistical techniques we used in the previous study: confusion matrices and bootstrapping. For more information on the techniques, refer to the descriptions in Chapter 4.2.1 for confusion matrices and 4.3.4 for bootstrapping.

9.3.4 Results

	Happy	Sad	Angry	Afraid	None	R Sum	Recall	D&S Recall
Happy	**130**	0	1	1	5	137	0.956	0.922
Sad	0	**130**	3	1	3	137	0.956	0.878
Angry	6	0	**120**	6	5	137	0.883	0.867
Afraid	1	2	4	**116**	14	137	0.861	0.878
Neutral/None	1	6	24	1	**105**	137	0.774	0.700
P Sum	138	138	152	125	132			
Precision	0.942	0.942	0.789	0.928	0.795	**0.879**	**0.886**	**0.849**

Table 9.2: Overall confusion matrix for all participants. The final column provides a comparison with our earlier results.

In the table, we also present the combination results from our earlier study in Chapter 4. We choose to use the combination results, as those were the best overall results. We were able to achieve very similar recall rates, and slightly exceeded the values in four of the five cases. However, the differences are generally quite small and are not statistically significant. The greatest differences are between the sad and none results, followed closely by the neutral results. We were able to achieve recall rates above 70% for all four emotions as well as the neutral character.

However, it is not possible to do a direct comparison for four reasons. First, the neutral/none scene in the original study did not include any emotional incidents, while in the Project Spark scene, the incidents were still there but the character did not respond noticeably to any but the third, which was just a brief pause. Second, the combination results used emotional gaits. In this experiment, the best we could do was to change the speed of the gaits for each incident. Third, we included a third emotional incident, providing more opportunities for participants to pick up on emotional cues. Fourth, we presented all five videos at once, allowing participants to change their answers after watching later videos (in the original study participants were unable to change their earlier decisions).

9.4 Conclusion

Project Spark ended up being a much larger challenge than we initially thought when it was chosen for the task of recreating the experiment. We ended up changing some of our plans (such as removing emotional gaits and creating new response animations) in order to fit within the constraints of the software. However, we did receive positive feedback from the participants on the overall quality of the animations and graphics.

In the end, we were able to re-create the experiment in Project Spark. The results of the experiment support the use of emotional incidents in worlds outside of our original test world created in Unity and created without professional animation skills. This is a significant step in establishing the viability of using emotional gaits and incidents in a variety of virtual environments.

Whether or not Project Spark can be used successfully by non-programmers is another question. We believe that if a non-programmer relied on the prebuilt material in Project Spark (including the maps and character/object behaviour) it is possible to get a full game up and running relatively quickly. Once you get involved in the actual 'koding', the answer is not so clear. However, as Project Spark is still in its infancy, there is time for the documentation and quality of the system to make this possible.

Project Spark in its present form can be used successfully to create animations but is not suitable for coding emotional incidents. Therefore, it might be natural to assume you could use a different tool (such as ScriptEase or Alice or Scratch) to code incidents and connect them to the Kinect generated animations. However, there is no existing infrastructure for this. It is not immediately clear whether it would be better to create this infrastructure or to improve the coding experience in Project Spark.

Chapter 10

Conclusion

This dissertation describes how to use two techniques (emotional gaits and emotional incidents) to create background game characters that have non-verbal, non-facial emotion that can be easily and accurately identified. Chapter 6 explains the implementation we used, and the flexibility it allowed. We have detailed six user studies that we ran to evaluate these techniques.

The first two user studies (Chapter 4 and Chapter 5) examin whether or not participants could identify a virtual character's emotion. Chapter 4 compares the two techniques on their own, as well as combined, to figure out the best solution. The results suggest that while overall the combination is the best, some emotions could achieve equally good results by using only one technique. Using both techniques together answered the question: "Can our technique obtain at least 70% accuracy in player identification of character emotion?", with a positive response. Chapter 5 is designed to specifically examine whether or not participant gender and/or character gender influenced the results. The main difference we detected was that female participants were better at identifying afraid characters compared to male participants. However, future studies did not confirm these results. More research is required.

The next two studies detailed in Chapter 7 were our first attempts at placing these emotional characters into a "real" game. From these two user studies, we learned much about adding these characters to a game. And from the qualitative results, it was possible to tell that users *were* noticing the emotional behaviour and attributing meaning to character actions.

Our fifth study (Guess Who - Chapter 8) was our successful attempt at placing these characters into a "real" game. Participants who spent a larger percentage of their time interacting with our emotional characters compared to the neutral characters found the game characters more believable. This study answered the question: "Are characters who exhibit emotional cues using our technique more believable" with a positive answer. In the first two play-throughs of the game, participants were only asked to locate a single character by emotion, thus being directly exposed to, at most, two different emotions.

In the first game, participants were able to successfully identify the emotion with an accuracy of just over 50%, which is almost equivalent to chance (50%). By game two, they had improved their accuracy to just over 80% (chance was still 50%), showing that participants were learning. In the third game, participants now had to label all the characters by emotion. While there was a drop from game two, the participants were now being exposed to more emotions and were able to use what they had learned from the first two games to accurately label 70% of the characters. This is well above chance, which drops to 20%, as there were five possible labels.

For our final study, we moved our techniques out of our original park into Project Spark (Chapter 9). We made a series of videos similar to our first study, with a character walking a set path and reacting emotionally to events. This study showed that not only was it possible to implement emotional incidents in different game worlds and have them be accurately identified, but also, that even without professional animation experience, one could create believable emotional response animations with the help of the Kinect.

10.1 Future Work

It is important to note that, while this research made great steps towards creating emotional characters, the research is not complete. The two techniques presented here have been shown to successfully convey emotion to our research population. However, the studies were primarily limited to taking place within a park with a similar (and limited) set of characters, props, background and emotional incidents. It is important that further work be done, extending this research to a variety of emotions, emotional incidents, and scenarios (10.1.1).

We briefly examined the ability to detect change in character emotion in Chapter 6, but found a significant limitation to our implementation (the lack of emotional change signals). In order to use emotion in many commercial or serious games, it would be important for characters to be able to react to their current environment and change their emotional state correspondingly (10.1.2).

As well, the studies, and therefore the results, were done with a very homogenous population of university students that cannot be directly extended to either the general population or the game playing community. In particular, the results from the gender studies that we ran did not agree as to whether or not the participants' or characters' gender causes a significant effect on the ability for participants to identify character emotion (10.1.3).

Finally, while we did place these characters into an game (Chapter 8), the time spent playing the individual games was very short (less than five minutes in each game). Most commercial and/or serious games have a much longer duration. In order to fully understand the effects of adding these techniques (emotional gaits and emotional incidents), they will need to be studied within a larger game (10.1.4)

10.1.1 Greater Variety of Emotions

All of the studies detailed in this dissertation focused on at most four emotions (happiness, sadness, anger and fear). All characters of a given emotion shared the same combination of possible reactions. Two possible extensions would be to a) create emotional incidents and gaits to represent more emotions and b) to create a wider variety of emotional incidents for the current set of emotions.

Representation of more emotions would allow for a richer game world. Depending on the context of a game, the types of emotions that should be present would change. A game involving characters sneaking around back alleys late at night would likely involve more characters that show signs of fear or surprise. A game taking place at a family reunion could easily be speculated to represent many emotions from happiness to resentment to contentment to grief.

Creating a variety of emotional incidents would allow for characters to also start to express some personality through their actions. While there are similarities in how people react when angry, individual personalities also affect the reactions. In addition, as Magy Seif El-Nasr and Huaxin Wei [29] noticed, a character's role (boss vs subordinate) also affects the reaction.

10.1.2 Change in Emotion

The work presented in this dissertation does not include using an existing emotional model (such as OCC or EMA). This was because we focused on assigning each character an emotional state that they maintained during each game. During the research described in Chapter 7, we ran a study where we briefly looked to see if participants could identify emotional changes in our characters. This study pointed out that our design of emotional gaits and emotional incidents does not directly support the response change that occurs during a shift in emotion. A well known example of the emotional change that can occur is of a bird flying into a room startling an actor (as Stacy Marsella and Jonathan Gratch detail in [50]). The reaction has many steps but occurs over only a few seconds. In a game, the viewer needs some clues that something is going to happen, in order for them to anticipate it. As Joseph Bates explained "Anticipation denotes broadcasting in advance that a certain action is about to occur, to let people prepare to see it. This is necessary because realistic timing causes events to happen quickly, and people will miss them unless they are experienced watchers" [8]. Requiring players to wait to see a change in gait (for example from happy to sad) or in an emotional incident is not sufficient to signal a sharp change in emotional state.

10.1.3 Gender and Emotion

When we conducted a study examining the influence of gender on the perception of characters (Chapter 5), our results were not conclusive. A participant's gender appears to have some influence on labeling. However, our later studies were not able to reproduce these results. As games become played equally often by both genders, new techniques (such as adding emotion to characters) should be evaluated to ensure that they do not discriminate based on gender.

10.1.4 Longer Game Play

Our successful study of placing these emotional characters into a game world involved a very short game (Chapter 8). While the results were quite encouraging, it is important to examine the long-term effects of these characters in a game. While we believe that a game would be best suited to a combination of neutral and emotional characters, is there an ideal ratio? How frequently do emotional incidents need to occur such that participants notice them? Can this system be combined with a system such as Ben Sunshine-Hill and Norman Badler's alibi generation [74]? Do players start to tire of the emotional incidents over time?

10.2 Revisiting Contributions

In conclusion this dissertation presents eight contributions, each revisited briefly below.

Contribution 1. *Commercial game characters do not have the flexibility needed in animations to create relatable believable everyday characters needed for many serious games.*

Through the process of initially designing a serious game to help patients suffering from chronic depression, it became apparent that current game characters are not *normal* enough. While characters may appear believable within their own game environment, they cannot easily be moved into a serious game without issues. These issues ranged from the "too perfect" body shapes to aggressive character stances to the lack of variety in character movement and visual behaviour.

Contribution 2. *Players can recognize non-verbal, non-facial emotion in game scenes.*

For our research, we used a combination of emotional gaits and emotional incidents to convey non-verbal, non-facial emotion in background characters. We then ran a series of user studies to determine if participants would notice and be able to identify the emotions. In our first study (Chapter 4), the combination technique achieved an average recall of 0.886. In later studies, we placed the characters into "real" games. While the recall rate dropped in the real game, after playing two very short games, participants were still able to accurately identify emotions over 70% of the time (Chapter 8).

Contribution 3. *Gender does play a role, but what the role is needs more study.*

We ran a study focused specifically on examining whether the gender of the player or the game character influenced the emotion identification (Chapter 5). While some significant differences were found in the first study, later studies did not confirm these results. More research needs to be done to determine the exact role gender plays.

Contribution 4. *Participant filtering is required.*

Over the course of the six experiments we ran, we learned a lot about our participant pools. Specifically, many of them were showing up for their course credit, but their actual participation stopped there. We created specific strategies to detect these students in later experiments and remove their data from our results (Chapter 7, 8).

Contribution 5. *Emotional gaits and emotional incidents do not allow for participants to recognize a character is changing emotions because the cues happen over a period of time.*

While we obtained positive results for participants' ability to identify emotion as displayed through our emotional gaits and emotional incidents, participants struggled to recognize changes in emotion. We ran a brief experiment which confirmed that our system is missing key features to allow for the identification of emotional change (Chapter 7).

Contribution 6. *Emotional characters are more believable.*

We found, when we had a combination of emotional and neutral characters, that participants who spent a larger percentage of their time interacting with the emotional characters found the characters to be more believable compared to those who spent more time with the neutral characters (Chapter 8).

Contribution 7. *Players can learn to quickly recognize emotion in games.*

The participants in our Guess Who study (Chapter 8) were able to learn to identify our emotional characters and their specific emotion with an accuracy of 70% after playing two short games.

Contribution 8. *Adding emotional gaits and incidents is relatively inexpensive.*

In the Guess Who study (Chapter 8), only three emotional gaits and 12 reaction animations were needed to convey three different emotions (anger, sadness and fear). Our final study used Project Spark and did not require animation expertise in order to create emotion specific reactions with the Kinect.

10.3 Conclusions

Video games are a regular part of life for many people, with the majority of American households (59% according to the ESA [3]) having an average of two gamers. While video games may have started with a focus on entertainment, there are now many games designed with a different goal in mind, from using the power of crowd sourcing to searching for advances in science (*Foldit* [19]) to educating people on life as a refugee (*Darfur is Dying* [59]) to helping soldiers understand new cultures (*First Person Cultural Trainer* [92]).

Both games for entertainment and serious games are benefiting from advances in computer graphics and the ability to include more complicated story paths and new game play mechanics. However, advances in character behaviour have not kept pace, resulting in static and repetitive character behaviour. Creating realistic behaviour can be expensive – from the art to animation to character scripting. If the character is a filler character, existing more in the background to populate the world than for direct interaction with the player, this added expense may not seem worth it.

This dissertation answers three main research questions:

1. Does our technique(s) produce easily recognized emotion?
2. Are the emotional characters that use this technique more believable than neutral characters?
3. Can we implement our technique using only a small number of animations and scripts so that the amount of extra effort required is small compared to the effort required to create existing character behaviours?

We used two techniques (emotional gaits and emotional incidents) to add identifiable non-verbal, non-facial emotion to these background characters. We believe that the addition of recognizable emotion will enrich the game world by allowing players to provide their own reasoning and motivation for why characters are behaving as they are (something that is hard to do when the characters never change their behaviour or do anything interesting). Over the course of six user studies, we showed that these techniques are accurate at conveying emotion, exceeding our goal of 70% accuracy when using the combined technique (question one).

For question two, we placed our emotional characters into a game and compared them against neutral characters. The results of that study (Chapter 8) show that participants perceived our emotional characters as more believable (question two).

Finally, in our last two studies (Chapters 8 and 9) we show that the cost of adding our emotional techniques is low (question three). The implementation system we detail in

Chapter 6 can be shared among all characters, requiring the actual scripting code only be implemented once per emotional incident. By using newer technologies, such as the Kinect, we show that it is possible to create believable emotional reaction animations.

From the comments and feedback we received during our user studies, it was clear that participants were creating stories about our characters. These stories may have started from the information they could see, but quickly grew to include the participant's imagination, such as the use of the word 'glare' to explain the angry interaction in our first experiment (Chapter 4), even though the participants could not see the character's face. In closing, we present a couple of these comments.

> "She was more enjoyable to watch because while she was sitting on the picnic table, it showed that she was more relax[sic] which made me feel very relax[sic] and not stressed."

> "His eyes looked bloodshot, his face had stubble, and he[sic] his skin was kind of grey. He looked stressed out, but not sad because he did not seem relaxed. He was tense. I got the impression that he had not been sleeping and had gone days without changing his clothing."

Bibliography

[1] Kenji Amaya, Armin Bruderlin, and Tom Calvert. Emotion from motion. In *Graphics Interface*, pages 222–229. Citeseer, 1996.

[2] Entertainment Software Association et al. Essential facts about the computer and video game industry., 2005. Retrieved January 10, 2015.

[3] Entertainment Software Association et al. Essential facts about the computer and video game industry., 2014. Retrieved January 10, 2015.

[4] Anthony P Atkinson, Winand H Dittrich, Andrew J Gemmell, Andrew W Young, et al. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception-London*, 33(6):717–746, 2004.

[5] b33k3rz. Hitman: Mardi Gras Massacre (The Murder of Crows) , 2010. [YouTube Clip].

[6] Christine Bailey and Michael Katchabaw. An emergent framework for realistic psychosocial behaviour in non player characters. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pages 17–24. Association for Computing Machinery, 2008.

[7] C. Basak, W.R. Boot, M.W. Voss, and A.F. Kramer. Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and aging*, 23(4):765, 2008.

[8] Joseph Bates. The role of emotion in believable agents. *Communications of the Association for Computing Machinery*, 37(7):122–125, 1994.

[9] Berrin Beasley and Tracy Collins Standley. Shirts vs. skins: Clothing as an indicator of gender role stereotyping in video games. *Mass Communication & Society*, 5(3):279–293, 2002.

[10] Believe at University of Alberta. ScriptEase, 2015. [Game Development Tool].

[11] Bethesda Softworks LLC. Oblivion: The Elder Scrolls, 2006. [Video Game].

[12] Bethesda Softworks LLC. Skyrim, 2011. [Video Game].

[13] BioWare. Dragon Age Origins. Website, 2009. http://dragonage.bioware.com/dao.

[14] BioWare. Mass Effect 2, 2010. [Video Game].

[15] BioWare. Mass Effect 3, 2012. [Video Game].

[16] Carnegie Mellon University. Alice, 2015. [Game Development Tool].

[17] Cipscis. No NPC Greetings, 2011. [Video Game Mod].

[18] Cristina Conati and Xiaoming Zhou. Modeling students? emotions from cognitive appraisal in educational games. In *Intelligent tutoring systems*, pages 944–954. Springer, 2002.

[19] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, et al. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, 2010.

[20] Maria Cutumisu and Duane Szafron. An Architecture for Game Behavior AI: Behavior Multi-Queues. In *Artificial Intelligence and Interactive Digital Entertainment Conference*, 2009.

[21] Beatrice De Gelder. Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, 7(3):242–249, 2006.

[22] Neesha Desai, Duane Szafron, Liliane Sayegh, Russell Greiner, and Gustavo Turecki. Creating Video Games to Treat Chronic Depression. In *GRAND Annual Conference Work In Progress*, 2011.

[23] João Dias and Ana Paiva. Feeling and reasoning: A computational model for emotional characters. In *Progress in artificial intelligence*, pages 127–140. Springer, 2005.

[24] Tracy L Dietz. An examination of violence and gender role portrayals in video games: Implications for gender socialization and aggressive behavior. *Sex roles*, 38(5-6):425–442, 1998.

[25] Mike Dunn. BioWare using Facebook to select the look of Mass Effect 3 main character, 7 2011.

[26] Eidos Interactive. Thief, 1998. [Video Game].

[27] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[28] Paul Ekman et al. Are there basic emotions. *Psychological review*, 99(3):550–553, 1992.

[29] Magy Seif El-Nasr and Huaxin Wei. Exploring non-verbal behavior models for believable characters. In *Interactive Storytelling*, pages 71–82. Springer, 2008.

[30] Magy Seif El-Nasr, John Yen, and Thomas R Ioerger. FLAME - fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257, 2000.

[31] Electronic Arts Inc. The Sims, 2014. [Video Game].

[32] Jing Feng, Ian Spence, and Jay Pratt. Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, 18(10):850–855, 2007.

[33] Gamasutra. The Replay Interviews: Will Wright, Accessed January 2015. [Interview].

[34] Rockstar Games. L.A. Noire. Website, 2011. http://www.rockstargames.com/lanoire/agegate/ref/?redirect=.

[35] GamingEdus. Why I cannot wait to bring my Kano to school, 2015. http://www.gamingedus.org/2014/10/why-i-cannot-wait-to-bring-my-kano-to-school/ [Image].

[36] Yue Gao and Regan Mandryk. The Acute Cognitive Benefits of Casual Exergame Play. In *Computer?Human Interaction '12: Proceedings of the 30th international conference on Human factors in computing systems.*, Austin, Texas, USA. 1863-1872, 2012.

[37] Garry Bardin. Conflict, 1983. [Video] [Accessed on January 2015].

[38] J.A. Hall. Gender effects in decoding nonverbal cues. *Psychological bulletin*, 85(4):845, 1978.

[39] P.J. Hardin and J.M. Shumway. Statistical significance and normalized confusion matrices. *Photogrammetric engineering and remote sensing*, 63(6):735–739, 1997.

[40] Tilo Hartmann and Christoph Klimmt. Gender and computer games: Exploring females dislikes. *Journal of Computer-Mediated Communication*, 11(4):910–931, 2006.

[41] Hasbro. Guess Who, 2015. http://www.hasbrogames.com/en-us/product/guess-who-game:8EFEF4A8-6D40-1014-8BF0-9EFBF894F9D4 [Board Game].

[42] Ion Storms. Deus Ex, 2000. [Video Game].

[43] James D Ivory. Still a man's game: Gender representation in online reviews of video games. *Mass Communication & Society*, 9(1):103–114, 2006.

[44] Jason Hamilton. Learn programming with 3D video producer 'Alice 3', 2015. http://www.freewaregenius.com/learn-programming-with-3d-video-producer-alice-3/ [Image].

[45] Joystiq. MIGS06: Ken Perlin on "the illusion of life", Accessed January 2015. [Interview].

[46] Kano Computing LTD. Kano, 2015. [Computer Kit and Software].

[47] Nick Kraayenbrink, Jassin Kessing, Tim Tutenel, Gerwin de Haan, Fernando Marson, Soraia R Musse, and Rafael Bidarra. Semantic crowds: reusable population for virtual worlds. *Procedia Computer Science*, 15:122–139, 2012.

[48] Sandy Louchart, Ruth Aylett, Michael Kriegel, João Dias, Rui Figueiredo, and Ana Paiva. Authoring emergent narrative-based games. *Journal of Game Development*, 3(1):19–37, 2008.

[49] Raymond A Mar, William M Kelley, Todd F Heatherton, and C Neil Macrae. Detecting agency from the biological motion of veridical vs animated agents. *Social cognitive and affective neuroscience*, 2(3):199–205, 2007.

[50] Stacy Marsella and Jonathan Gratch. EMA: A computational model of appraisal dynamics. In *European Meeting on Cybernetics and Systems Research*, 2006.

[51] Stacy Marsella, Jonathan Gratch, and Paolo Petta. Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, pages 21–46, 2010.

[52] Martha Alderson. Connecting with Audiences Through Character Emotions, 2015.

[53] Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona Newell, and Carol O'Sullivan. Evaluating the emotional content of human motions on real and virtual characters. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 67–74. ACM, 2008.

[54] Microsoft. Project Spark, 2014. [Video Game].

[55] Microsoft. Kodu Game Lab, 2015. [Video Game].

[56] Microsoft Research. Programming as a Creative Medium, 2015. http://research.microsoft.com/en-us/projects/kodu/ [Image].

[57] MIT Media Lab. Scratch, 2015. [Game Development Tool].

[58] Joann M Montepare, Sabra B Goldstein, and Annmarie Clausen. The identification of emotions from gait information. *Journal of Nonverbal Behavior*, 11(1):33–42, 1987.

[59] mtvU. Darfur is dying, 2009. [Video Game].

[60] National Academy of Sciences. Women in Science and Engineering Statistics, 7 2015.

[61] A Ortony, GL Clore, and A Collins. The cognitive structure of emotions, 1988.

[62] Andrew Ortony. On making believable emotional agents believable. *Trappl et al.(Eds.)(2002)*, pages 189–211, 2002.

[63] Andrew Ortony, Terence J Turner, et al. What's basic about basic emotions. *Psychological review*, 97(3):315–331, 1990.

[64] Marco Pasch and Ronald Poppe. Person or puppet? the role of stimulus realism in attributing emotion to static body postures. In *Affective Computing and Intelligent Interaction*, pages 83–94. Springer, 2007.

[65] Ken Perlin. Polly's World. Website, 2011. http://mrl.nyu.edu/ perlin/experiments/polly/.

[66] Pixar. Luxor Jr. Website, 1986. http://www.pixar.com/shorts/ljr/.

[67] Primary ICT Development. Teaching Scratch-Beginners guide, 2015. http://primaryictdevelopment.blogspot.ca/2013/04/download-lesson-1-of-teaching-scratch.html [Image].

[68] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, et al. Scratch: programming for all. *Communications of the Association for Computing Machinery*, 52(11):60–67, 2009.

[69] Mark O Riedl and R Michael Young. An objective character believability evaluation procedure for multi-agent story generation systems. In *Intelligent Virtual Agents*, pages 278–291. Springer, 2005.

[70] Paola Rizzo, Manuela Veloso, Maria Miceli, and Amedeo Cesta. Personality-driven social behaviors in believable agents. In *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents*, pages 109–114, 1997.

[71] Claire L Roether, Lars Omlor, Andrea Christensen, and Martin A Giese. Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6), 2009.

[72] Randy Smith and Smith Harvey. Practical Techniques for Implementing Emergent Gameplay. Presented at Game Developers Conference, San Jose, California, 2004.

[73] Square Enix. Hitman: Blood Money, 2006. [Video Game].

[74] Ben Sunshine-Hill and Norman I Badler. Perceptually Realistic Behavior through Alibi Generation. In *Artificial Intelligence and Interactive Digital Entertainment Conference*, 2010.

[75] Fabien Tencé, Cédric Buche, Pierre De Loor, and Olivier Marc. The challenge of believability in video games: Definitions, agents models and imitation learning. *Preprint arXiv:1009.0451*, 2010.

[76] The Unofficial Elder Scrolls Pages. Skyrim:People, 2014. [Wiki of Skyrim Characters].

[77] Frank Thomas, Ollie Johnston, and Frank. Thomas. *The illusion of life: Disney animation*. Hyperion New York, 1995.

[78] Angela Tinwell, Mark Grimshaw, and Debbie Abdel-Nabi. Effect of emotion and articulation of speech on the uncanny valley in virtual characters. In *Affective Computing and Intelligent Interaction*, pages 557–566. Springer, 2011.

[79] C. J. van Rijsbergen. Information Retrieval, 1979. [Video Game].

[80] Hal Varian. Boostrap Tutorial. *The Mathematica Journal*, 9(4):768–775, 2005.

[81] Ivar E Vermeulen, Christian Roth, Peter Vorderer, and Christoph Klimmt. Measuring user responses to interactive stories: Towards a standardized assessment tool. In *Interactive Storytelling*, pages 38–43. Springer, 2010.

[82] Harald G Wallbott. Bodily expression of emotion. *European journal of social psychology*, 28(6):879–896, 1998.

[83] Harald G Wallbott and Klaus R Scherer. Cues and channels in emotion recognition. *Journal of personality and social psychology*, 51(4):690–699, 1986.

[84] Infinity Ward. Call of Duty Modern Warfare 2. Website, 2009. http://www.modernwarfare2.com/.

[85] YouTube xCheezbrgr. Call of Duty Modern Warfare 2 - No Russian - Airport Mission Uncut - Veteran - High Quality. Website, 2009. http://www.youtube.com/watch?v=gXBDkevx5lM.

[86] Geogios N Yannakakis. Game AI revisited. In *Proceedings of the 9th conference on Computing Frontiers*, pages 285–292. Association for Computing Machinery, 2012.

[87] YoYo Games. GameMaker: Studio, 2015. [Game Development Tool].

[88] YoYo Games Ltd. Game Maker, 2015. https://itunes.apple.com/us/app/game-maker/id408626090?mt=12 [Image].

[89] Richard Zhao and Duane Szafron. Generating Believable Virtual Characters Using Behavior Capture and Hidden Markov Models. In *Advances in Computer Games Conference (ACG)*, pages 342–353, 2011.

[90] Richard Zhao and Duane Szafron. Using Cyclic Scheduling to Generate Believable Behavior in Games. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.

[91] Marjorie A Zielke, Monica J Evans, Frank Dufour, Timothy V Christopher, Jumanne K Donahue, Phillip Johnson, Erin B Jennings, Brent S Friedman, Phonesury L Ounekeo, and Ricardo Flores. Serious games for immersive cultural training: Creating a living world. *Computer Graphics and Applications, IEEE*, 29:49–60, 2009.

[92] Marjorie A Zielke and Thomas E Linehan. The first person cultural trainer. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, volume 2009. NTSA, 2009.

Appendix A

Challenges and problems of prototyping a serious game

A short paper from the GRAND-nce (Graphics Animation and New Media) annual conference in 2011, held in Vancouver, British Columbia, Canada, motivated the research described in this dissertation.

Challenges and problems of prototyping a serious game

Neesha Desai
Department of Computing Science
University of Alberta
neesha@ualberta.ca

Duane Szafron
Department of Computing Science
University of Alberta
duane@ualberta.ca

Liliane Sayegh
Douglas Mental Health University Institute
Department of Psychiatry
McGill University
lilian.sayegh@douglas.mcgill.ca

Russell Greiner
Department of Computing Science
University of Alberta
rgreiner@ualberta.ca

Gustavo Turecki
Douglas Mental Health University Institute
Department of Psychiatry
McGill University
gustavo.turecki@douglas.mcgill.ca

ABSTRACT
Video games are a thriving industry that is quickly catching up to the movie industry. Serious games, a sub-genre, are a more recent trend. A serious game is a video game that is designed to teach the players about some concept or idea. We have been working as a team of computing scientists and psychologists to prototype a game to help treat people suffering from chronic depression. The game is based on a common treatment technique called a situational analysis and is designed to teach the concept of a situational analysis to the patient. Over the course of developing a prototype of the game, we have discovered a number of challenges, some expected, but many were unforseen. This paper discusses these challenges and some potential solutions.

Keywords
Video Games, Serious Games, Psychology, Depression

1. INTRODUCTION
Video games are a multi-billion dollar industry in North America. Consumers spent $18.85 billion on games and consoles in 2007 [1]. Video games are a fast-growing industry whose profits are quickly catching up to the movie industry [1]. This rapid growth can be explained by a changing consumer market that includes more older gamers and more female gamers and the increase in the length and interactivity of the games which has improved game replay and sequel value [1].

"Serious games" are an emerging genre of video games. Serious games are designed to teach the player through game play. Many of the games that fall in this category are health related. For example, *Kidney Heroes* [7] teaches pre-teen patients with chronic kidney disease how to manage their condition, and *Darfur is Dying* [6] teaches players about living in a refugee camp. A benefit of serious games is that they can teach important concepts without placing the player in a potentially dangerous situation. The player can try various scenarios without having to worry about any possible consequences.

We are investigating the use of serious games for people suffering from chronic depression. In a common treatment technique, patients describe to their therapist a real-life scenario that caused them to feel depressed. The patient and therapist then work through a *situational analysis*, in which the therapist asks the patient a series of questions about the scenario. The goal of a situational analysis is for patients to develop coping strategies and techniques for dealing with these situations in their everyday lives. We are attempting to create a game that simulates a situational analysis to help patients learn the process.

Creating serious games presents an interesting challenge in that these games requires two types of specialized knowledge. First, they require people who are able to build and produce a game. And second, they require domain experts - people who understand *what* the game is trying to teach.

This paper presents the challenges or stumbling blocks we have come across in our journey. Section 2 gives a deeper overview of situational analysis. Section 3 discusses how we are attempting to turn it into a video game. Section 4 presents the problems we've discovered, along with some potential solutions. Finally, section 5 is a brief conclusion.

2. WHAT IS A SITUATIONAL ANALYSIS?
A situational analysis is a common treatment technique used by psychologists to treat patients suffering from chronic depression. It begins with a patient telling the psychologist about a recent event where the patient was unable to obtain his or her desired outcome. The patient and the therapist then discuss and analyze the scenario by following a predetermined set of questions. The following is an example of a

GRAND 2011, May 12-14, 2011, Vancouver, BC, Canada

story as presented by a patient to a psychologist from the Patient's Manual for CBASP [5]:

> My son was in his high chair, and I was feeding him lunch. The doorbell rang. I stopped feeding him, left my son in the chair, and went to the front door. It was my next-door neighbour, who is pushy and aggressive. She said that she needed a cup of sugar for a cake she was baking. I told her that this was not a good time and asked her to come back. I opened the door and let her in. She mumbled something about this not taking long and then asked me where the sugar was. I pointed her to the sugar bin on the counter and went back to feeding my son. She got the sugar and left. I was frustrated, mad, and then I got depressed and thought to myself: "I've been screwed again."

After presenting the scenario, the therapist then proceeds through a set of questions with the patient, where they discuss the scenario and the patients interpretations of the events. The goal is to help the patient learn coping strategies and techniques for future situations.

3. THE GAME

The goal of our game is to turn a situational analysis into a video game that can be used by patients either alone or in psychologists-led group therapy to learn the process. We decided that this could be easily done by having two steps. First, by presenting a scenario to the patient in the form of a cutscene (video clip) and second, a discussion of the scenario with a virtual therapist.

However (and not surprising), it turns out that trying to directly map a situational analysis into a game creates a boring game. And if the game is boring there will be no motivation for the patients to play it, much less finish. It is also difficult to argue that having patients watch a scenario and then proceed through a series of dialogue questions even qualifies as a game. The patient does not control a character or get to make any decisions that have obvious consequences.

Therefore, we proposed a third section we could add to the game that we call replay. What's replay? Replay is the idea of role-playing the scenario, but with the ability to go through multiple versions of it. It allows patients to enter the original scenario in the role of the main character and make their own choices as to what actions to take and what dialogue to say. What happens if Mary ignores the door-bell?Or if Mary slammed the door in the neighbor's face? What actions can Mary take that will have a positive impact in letting her get her desired outcome? This gives the patient a chance to test out different theories in a non-threatening situation with no real-life consequences.

4. PROBLEMS

Over the course of developing the prototype, we discovered a variety of problems. Some were issues we had identified before starting, but others were complete surprises. The problems included game art, animations, dialogue, and realism. We discuss these problems and some potential solutions.

4.1 Game Art

The game art is often what attracts players to a game and keeps them engaged. If done well, a player can believe s/he is immersed in the game world. If not, it can be a distraction and can actually cause a game to fail. Our prototype is being built using BioWare's *Neverwinter Nights* (NWN) game tools [2]. This presented a set of immediate challenges as the game is set in a fantasy medieval world. The game was released in 2002 (which makes it "ancient") and therefore there is little facial detail, no facial animations, and limited character animations.

We knew from the start that we wouldn't be able to only use the default art from the game. While the game is filled with common objects like tables and chairs, and normal animals like penguins and bears, it also has magic spells, orcs, trolls and dragons. A kitchen in NWN has a fire pit with a pot hanging over it - not like a kitchen one would find today. The game does not have ovens or sinks, or characters dressed in jeans and t-shirts. However, by using some community-made art assets we are able to create a more "modern" world, as shown in Figure 1. However, if you look closely on the left side of Figure 1, you'll notice that it is still not as modern as you would expect. There is a medieval banner hanging down in front of the counters with an old fashioned light attached.

Figure 1: A kitchen using the modern toolkit.

Unfortunately, there are few games that take place in the present day world that don't have a science fiction or fantasy twist. It is difficult to find pre-built modern art assets that can be used between games. However, the psychologists pointed out another problem that exists even with more contemporary game art. The characters are too perfect, specifically their body shape, as shown in Figure 2. The psychologists feel, that since the patients are often overweight due to medications they take and are not dressed as well as the character art portrays, the patients would find it difficult to relate to the characters.

A digital artist on our research team is creating some game models that more closely match our expectations. Using newer art techniques, we are able to create faces that look much more photo-realistic and are able to express emotions, see Figure 3. By creating our own art, we can re-design character body types to reflect a more average body. Figure 4 provides a couple of sketches of Mary dressed in sweats with a more suitable body shape.

Figure 2: A screenshot of a modern female in NWN.

An important part of the game is for the patient to understand what emotions characters are supposed to be expressing. Are they upset, happy, depressed, angry or resigned? Identifying emotions is something we do every day, and a lot of our clues come from facial expressions. Is the person you are talking with smiling or frowning? Are their brows furrowed? In NWN, the characters have very few facial features and it is not possible to make them smile or frown. In fact, the lack of facial detail in NWN (as visible in Figure 2) actually makes the characters appear a bit creepy when looked at too closely.

4.2 Animations

Most video games consist of a limited number of animations. These animations are usually quite basic and include a walk, run, stand, and sit. It was these basic animations that were identified as a problem by the psychologists. They found that the characters walk too fast. When characters are idle they sort of bob about with their arms hanging down by their sides. Most people would not think twice about their stance. However, the psychologists immediately identified the body stance as being very aggressive. To them, the characters looked ready to start fighting at any moment and the psychologists felt that would counteract the tone of the voiced dialogue.

Figure 3: Artist version of a woman's face.

Figure 4: An example of more realistic clothing and body shape.

In NWN, and in some other games, characters are unable to both sit and hold a conversation at the same time. Likewise, characters are often unable to actually carry an arbitrary object (non-weapon)in their hands (instead it gets placed in the character's inventory). We are working on dealing with these problems by creating our own character animations. For example, we want our characters to be able to choose between different speeds and types of walking, and to be able to sit and converse at the same time.

4.3 Dialogue

Dialogue is arguably the most important part of the game. However, patients suffering from chronic depression often have difficulty with reading and comprehension. A large wall of text will seem overwhelming to them. Therefore the text needs to be as short as possible. However, one goal of the game is to teach the patients about the situational analysis process and it is difficult (impossible) to do so without using dialogue to explain what is going on and why.

The analysis section of the game contains a large amount of dialogue. If presented without breaks, it is easy to become overwhelmed by the volume of content and to have no clear sense of progress. Our solution to this was looking at how we could divide the dialogue into pieces. We did this by turning each step into its own complete section of dialogue. Now, instead of proceeding through dialogue without breaks, the patients only needs to complete a single step at a time. When starting the analysis section, only the first step is available. Each time the patient completes a step, a counter is increased and the next step is made available. Each step ends with an option to exit the dialogue or continue directly to the next step. This gives the patient the ability to control the pace of the analysis. If the patient is feeling overwhelmed they can stop and start again later. Each time the patient restarts the dialogue, all the steps the patient has completed, plus the next step are available. The previous steps remain available so a patient can re-play a previous step. In addition, we allow the patient to replay the initial cut scene as many times as the patient wants, between steps.

The other change we made was to voice all the dialogue lines. As mentioned earlier, the characters in NWN do not have facial expressions. We voiced the dialogue to help convey the tone of what is being said. Voicing the lines also means the patients do not have to read all the dialogue that is presented.

4.4 Realism

Another discovery we did not expect was the problem with realism. Even though everyone involved in the project knows we are creating a game and that it is not real, the lack of realism became problematic. Particularly, the psychologists do not believe the game world looks real enough. The believe that it will be difficult for patients to identify with what they are watching. This problem is not limited to our game as Eleni Stroulia and her team [3] working on EMT training in *Second Life* [4] came across the same problem when presenting their research to the EMTs. The EMTs must stretch out the leg of the patient and in the game the leg appears to stretch out at an odd angle. The EMTs found this took them out of the game. In our game, the psychologists identified not just the game world, but issues like characters stance, lack of facial features, the speed the characters walk and the too perfect character bodies as all contributing to the problem.

Some of these issues can be fixed, such as using different character models, but others are problems of the NWN game environment, such as the walking speed and (lack of) facial features. There are two possible solutions. First, try to make everything closer to reality. This solution is difficult (if not impossible) to achieve, as there are limits on how realistic we can possibly make characters and character actions. The second solution is to move away from being realistic and making the characters more cartoonish. This removes the expectation that the characters will move and act identically to humans.

5. CONCLUSION

In this paper we have discussed a number of challenges that we discovered while creating a prototype for our serious game. The challenges range from problems with the available game art to the dialogue structure and dealing with (lack of) realism.

These problems bring us to the question: Where do we go from here? We know that we need to reconsider the game engine we are using. However, one of our initial goals was for the psychologists to be able to create the games on their own. This meant that the system we use needs to be easy enough for them to build scripts, fill in the dialogue, and design the game world. We believe that the combination of ScriptEase and the Aurora toolset meets these needs and that other game engines are more complicated.

However, it is also apparent that NWN is not the proper tool for developing this game. Some of the problems mentioned are NWN specific. However, most of the problems are inherent to the way most commercial video games are developed. Commercial game tools are becoming too advanced for novices to be able to learn and produce games as they requires too much game-specific knowledge. The Aurora toolset that comes with NWN spoiled novices. It came with a pile of art from which to populate a world, a dialogue editor for structuring your dialogue, and the ability to add scripts to game objects using a C like language.

The next step is to investigate the other game engines and tools available. We may want to consider moving from a 3D to a 2D environment. We will need to determine how important it is to have realistic looking characters. We need to figure out which of these problems are most important for patients to be able to identify with the game and be therefore willing to play it.

6. REFERENCES

[1] E. Bangeman. Growth of gaming in 2007 far outpaces movies, music. Website, 2009. http://arstechnica.com/gaming/news/2008/01/growth-of-gaming-in-2007-far-outpaces-movies-music.ars.

[2] BioWare. Neverwinter nights. Website, 2009. http://nwn.bioware.com/.

[3] D. Chodos, P. Naeimi, and E. Stroulia. A simulation-based training framework for health-science education on video and in a virtual world. *Journal of Virtual Worlds Research*, 2009.

[4] I. Linden Research. Second life. Website, 2011. http://www.secondlife.com/.

[5] J. P. McCullough Jr. *Patient's Manual for CBASP (Cognitive Behavioral Analysis System of Psychotherapy)*. The Guilford Press, New York, 2003.

[6] mtvU. Darfur is dying. Website, 2010. http://www.darfurisdying.com/.

[7] D. Pozefsky and M. Ferris. Kidney heroes: Improving health management through video games. Presentation, 2009.

[8] ScriptEase. Scriptease. Website, 2010. http://webdocs.cs.ualberta.ca/ script/.

Appendix B

ANOVA for the Gender Study

An ANOVA analysis comparing emotion, character gender, participant gender, gaming ability and consistency for the gender study presented in Chapter 5

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Emotion	3	5.21	1.738	16.356	**1.99e-10**
Character	1	0.03	0.028	0.261	0.60921
Participant	1	0.52	0.522	4.910	**0.02689**
Gamer	1	0.45	0.450	4.238	**0.03973**
Consistent	1	10.14	10.141	95.460	**<2e-16**
Emotion:Character	3	0.37	0.124	1.172	0.31925
Emotion:Participant	3	1.51	0.503	4.735	**0.00273**
Character:Participant	1	0.15	0.151	1.424	0.23305
Emotion:Gamer	3	0.15	0.050	0.470	0.70301
Character:Gamer	1	0.09	0.086	0.808	0.36878
Participant:Gamer	1	0.19	0.192	1.811	0.17863
Emotion:Consistent	3	1.18	0.395	3.717	**0.01118**
Character:Consistent	1	0.02	0.017	0.158	0.69143
Participant:Consistent	1	0.00	0.000	0.001	0.97366
Gamer:Consistent	1	0.00	0.000	0.003	0.95913
Emotion:Character:Participant	3	0.13	0.042	0.397	0.75515
Emotion:Character:Gamer	3	0.23	0.078	0.730	0.53414
Emotion:Participant:Gamer	3	0.17	0.056	0.526	0.66462
Character:Participant:Gamer	1	0.00	0.000	0.001	0.97850
Emotion:Character:Consistent	3	0.37	0.122	1.153	0.32655
Emotion:Participant:Consistent	3	0.07	0.025	0.234	0.87276
Character:Participant:Consistent	1	0.18	0.175	1.650	0.19914
Emotion:Gamer:Consistent	3	0.49	0.163	1.537	0.20311
Character:Gamer:Consistent	1	0.11	0.105	0.993	0.31930
Participant:Gamer:Consistent	1	0.18	0.177	1.670	0.19649
Emotion:Character: Participant:Gamer	3	0.11	0.038	0.358	0.78365
Emotion:Character: Participant:Consistent	3	0.13	0.043	0.403	0.75092
Emotion:Character: Gamer:Consistent	3	0.25	0.083	0.782	0.50387
Emotion:Participant: Gamer:Consistent	3	0.03	0.011	0.102	0.95914
Character:Participant: Gamer:Consistent	1	0.01	0.013	0.119	0.73027
Emotion:Character:Participant: Gamer:Consistent	3	0.20	0.065	0.615	0.60549
Residuals	1232	130.88	0.106		

Table B.1: ANOVA summary for emotion, participant gender (Participant), character gender (Character), gaming ability (Gamer) and consistency (Consistent). Bold indicates statistically significant at 95%.

Appendix C

Comments from Return Tablet Game

C.1 Emotional Characters

Comments that imply the character has an emotion.

- He was walking around very slowly looking very sad.
- He was walking faster than other characters, a little paniced, he was the one who lost the ipad.
- Walking sulkly through park
- run around agrilly
- she was really mean and she was mopping around
- Engaged in friendly conversation when I talked to him. Appeared relaxed. Could not see his eyes, but from his conversation and his body language he still seemed happy.
- He just was a happy guy
- She was the happiest of the characters. She was nice and conversed in a very friendly way even though we had never met. She gave more than a yes or no answer the first time I talked to her, and the second time she furthered our conversation. Her clothes were also clean, nice, and colourful. Her face and eyes looked happy and friendly. She looked healthy.
- She was sad at first because she lost her ipad, but when I found her and asked her if

she had lost an ipad she was happy because it was found.

- she said get out of my way, was not in the mood for socializing. obviously something was bothering her.
- He maybe did not want to be disturbed and therefore was angry every time I tried to initiate a conversation.
- Jake was happy to be enjoying the outdoors on a nice and sunny day. He was just talking a stroll in the park.
- His shoulders looked slumped and his expression looked sad or depressed
- His shoulder was bended that it looks like he has something bad happened to him.
- He was walking slowly and he was hunched. He was looking down and his eyes looked sad and droopy. He looked unkempt: He had stubble on his face and was dressed sloppily in a dirty looking shirt, like he had given up.

Comments about the characters behaviour (but not emotion specific).

- It appeared as if he was searchin for his lost iPad
- He was looking on the ground while walking, so I thought it might be his but he also looked like he was acting drunk...
- walking around with hands in his pockets
- Ran away screaming
- looked like she was looking around for something
- He seemed like he was acting as though he was too cool for you, the sunglasses and shirt gave me the impression as well as the way he walked.
- He had lost his ipad and was searching frantically for it.

C.2 Neutral Characters

Comments that imply the character has an emotion.

- She kept telling me to go away, seemed frustrated, angry.
- He was walking around with a fishing pole. In the game when you tried asking him if he had lost his ipad he was angry and said Leave me alone!"
- walk around grumpily

- He seemed really anxious. He stuttered when answering my question if the ipad was his or not.
- He seemed a little nervous when answering my question.
- He said something along the lines of good luck finding the owner". Sounded cheery and smiled.
- I can't remember whether she was happy or afraid. She talked to me plesantly the first time, and gave me more than just a yes or no answer, which would indicate that she was happy. But she had a frightened look in her eyes and appeared tense, and I think when I talked to her the second time she seemed more distracted and afraid in what she said.
- maybe sad because he sat lonely and maybe was reminising on things
- He didn't seem to want to talk with me so I thought he was angry about something.
- I do not know why she was acting happy and helpful, but it was much nicer than interacting with people who werent nice to you.
- She was probably happy because she saw her friend and she is very friendly.
- No conversation box popped up, I guess she just looked at me holding an iPad and shouted at me that it's not hers. She also told me to leave her alone and go away, I find her rather rude. She appeared... irritated or angry judging by the orange words on top of her head, she was probably shouting those words if I can see those words from a distance away.
- I do not know why she was exhibiting such behaviour, but she was not happy. She said to get lost and get out of her way
- She cwas clearly very upset and angered by something. Why she was so rude and unhelpful to my character, I have no idea.

Comments about the characters behaviour (but not emotion specific).

- Everytime I asked if the iPad was hers, she would yell at me to 'get away' or 'No, it's not mine. Get lost'
- stalked around, usually with her back turned
- walking quickly, was not in the mood for converstation
- She was waving at a friend.
- He was walking with a purpose.

- He was walking around mischeviously.
- he sat on a bench by the lack. he was lonely.
- He was standing by the trees, almost deciding which way to go.
- Walking around the pool like a badass
- He was pacing back and forth in the corner. I think he may have been waiting for someone else who might deliver some good/bad news.
- Fred was not a nice guy. Maybe he couldn't catch any fish that day. Funny that he'd be fishing in a pond in the middle of a park the size of a quarter. What did he think he would catch.
- She was rather helpful, telling me to look around the picnic table for the owner of the iPad, I suppose the owner of the iPad was asking around for her iPad.
- He looked desperate as he ran around and wouldnt speak with me.
- He was walking quickly and purposfully and was standing stright. He did not stop to talk to me, he just yelled mean things. He didn't even look at me (but he was wearing sunglasses so it was hard to tell). The sunglasses also added to his indifference towards me.

Appendix D

ANOVA Results for the Guess Who Study

The resulting ANOVAs on participants ratings of believable characters (Table D.1), presence (Table D.2), agency (Table D.3), and enjoyment (Table D.4) based on gender, frequency of game playing, if they won the first game and the number of conversations they had.

D.1 Believable Characters

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gender	1	7.6	7.60	0.901	0.3478
Game	1	0.6	0.64	0.076	0.7839
Won	1	9.4	9.43	1.118	0.2963
Convos	1	38.6	38.63	4.577	**0.038**
Gender:Game	1	0.1	0.07	0.009	0.9266
Gender:Won	1	11.2	11.23	1.331	0.2550
Game:Won	1	7.0	6.98	0.828	0.3680
Gender:Convos	1	10.2	10.21	1.210	0.2775
Game:Convos	1	13.2	13.24	1.568	0.2172
Won:Convos	1	19.6	19.56	2.317	0.1353
Gender:Game:Won	1	0.2	0.20	0.023	0.8793
Gender:Game:Convos	1	48.2	48.20	5.711	**0.0213**
Gender:Won:Convos	1	22.5	22.55	2.672	0.1094
Game:Won:Convos	1	2.8	2.81	0.334	0.5666
Residuals	43	362.9	8.44		

Table D.1: **Believable Characters** - ANOVA summary for gender, frequency of game playing (Game), winning the first game (Won), and number of conversations (Convos). Bold indicates statistically significant at 95%.

D.2 Presence

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gender	1	2.1	2.09	0.078	0.781
Game	1	0.1	0.06	0.002	0.964
Won	1	7.7	7.73	0.290	0.593
Convos	1	7.1	7.13	0.267	0.608
Gender:Game	1	2.8	2.84	0.106	0.746
Gender:Won	1	0.0	0.01	0.000	0.985
Game:Won	1	48.3	48.32	1.810	0.186
Gender:Convos	1	30.5	30.46	1.141	0.291
Game:Convos	1	6.1	6.14	0.230	0.634
Won:Convos	1	24.6	24.62	0.922	0.342
Gender:Game:Won	1	72.2	72.22	2.705	0.107
Gender:Game:Convos	1	123.7	123.71	4.634	**0.037**
Gender:Won:Convos	1	9.1	9.11	0.341	0.562
Game:Won:Convos	1	18.5	18.48	0.692	0.410
Residuals	43	1148.0	26.70		

Table D.2: **Presence -** ANOVA summary for gender, frequency of game playing (Game), winning the first game (Won), and number of conversations (Convos). Bold indicates statistically significant at 95%.

D.3 Agency

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gender	1	29.0	28.98	1.767	0.191
Game	1	8.1	8.07	0.492	0.487
Won	1	2.2	2.16	0.131	0.719
Convos	1	14.1	14.05	0.857	0.360
Gender:Game	1	0.0	0.02	0.001	0.974
Gender:Won	1	0.2	0.25	0.015	0.903
Game:Won	1	2.3	2.25	0.137	0.713
Gender:Convos	1	26.1	26.14	1.593	0.214
Game:Convos	1	1.0	1.04	0.063	0.802
Won:Convos	1	5.7	5.67	0.346	0.560
Gender:Game:Won	1	28.5	28.48	1.736	0.195
Gender:Game:Convos	1	36.0	36.03	2.196	0.146
Gender:Won:Convos	1	9.4	9.36	0.571	0.454
Game:Won:Convos	1	0.1	0.13	0.008	0.928
Residuals	43	705.4	16.40		

Table D.3: **Agency** - ANOVA summary for gender, frequency of game playing (Game), winning the first game (Won), and number of conversations (Convos). Bold indicates statistically significant at 95%.

D.4 Enjoyment

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gender	1	3	2.91	0.033	0.8569
Game	1	212	211.98	2.393	0.1292
Won	1	292	291.52	3.291	0.0766 .
Convos	1	24	24.02	0.271	0.6052
Gender:Game	1	73	72.93	0.823	0.3693
Gender:Won	1	0	0.38	0.004	0.9478
Game:Won	1	168	168.28	1.900	0.1752
Gender:Convos	1	3	2.71	0.031	0.8620
Game:Convos	1	12	11.78	0.133	0.7171
Won:Convos	1	115	114.70	1.295	0.2614
Gender:Game:Won	1	17	17.16	0.194	0.6620
Gender:Game:Convos	1	290	290.19	3.276	0.0773
Gender:Won:Convos	1	3	2.87	0.032	0.8580
Game:Won:Convos	1	74	74.47	0.841	0.3643
Residuals	43	3809	88.58		

Table D.4: **Enjoyment** - ANOVA summary for gender, frequency of game playing (Game), winning the first game (Won), and number of conversations (Convos). Bold indicates statistically significant at 95%.