Understanding Intentions in Animacy Displays Derived from Human Motion

Philip E. McAleer
B.Sc.

Submitted in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

To the

Faculty of Information and Mathematical Science

And

Department of Psychology,

University of Glasgow

Submitted 13th October 2006

Spectemur Agendo

[*Let us be judged by our actions – McAleer Family Motto*]

Abstract

As humans we live in a world where we are constantly interacting with those around us. To achieve this we must be able to successfully anticipate the intentions of others by correctly interpreting their movements. In studying how humans interpret intention from motion, we make use of simplified scenarios known as animacy displays where it has been shown that observers will attribute human-like qualities to the motion of geometric shapes (Heider and Simmel, 1944). This thesis advances the research into the attribution of social intentions by re-addressing the methods for the creation of animacy displays, leading to previously unexplored avenues of research. Where animacy displays are normally made via clever animations or mathematical algorithms, we introduce a method for creating these displays directly from video recordings of human motion, there by producing the first examples of animacy displays that are truly representative of human motion.

Initially, explorative steps were taken to establish this technique as successful in creating displays that will be perceived as animate, using video recordings of simple and complex human interactions as a basis. Using a combination of tasks, including free response tasks and 10 point Likert scales, the use of this technique for stimulus production was validated. Furthermore, results showed that the viewpoint from which animacy displays are to be perceived from, comparing a side view and an overhead view, has effects on the ability to judge intentions in the displays, with a clear preference to the elevated viewpoint.

Following this, the intentions of Chasing, Fighting, Flirting, Following, Guarding and Playing, thought to be generic to animacy displays, were used to

create displays via this new method of stimulus production. Using a six Alternative Forced Choice (AFC) task it was shown that participants are successful at recognising these intentions, however, that the addition of ordinal depth cues, as well as cues to identity and boundaries, has little impact on increasing the ability to perceive intentions in animacy displays. Next, an experiment on the ability to judge intentions in animacy displays of brief durations was performed. Using the same 6 intentions as before, displays were created lasting 1, 5, and 10 seconds. Results of a 6 AFC task showed that observers are accurate at all durations, and furthermore, results indicate that participants are as accurate at recognising the intention in a display after 5 seconds, as after viewing longer durations of approximately 30 seconds.

We then perform a comprehensive analysis of the animacy displays used, looking at the motion patterns and the kinematic properties such as speed, acceleration and distance of the agents. This analysis shows clear differences in the displays across viewpoints, and across intentions, that are indicative of the cues that participants may use to differentiate between intentions. We also perform a stepwise regression analysis to find the motion and positional predictors that best explain the variance in the behavioural data of previous experiments in this thesis. It is found that speed and acceleration cues are important for the classification of intentions in animacy displays.

Finally, a study is presented that attempts to advance research into the perception of social intentions by people with Autistic Spectrum Disorders (ASDs), using video recordings of human motions and the resultant animacy displays. The intentions of Chasing, Fighting, Flirting, Following, Guarding and Playing, were again used in conjunction with a 6 AFC task. Comparing people

with ASDs to an age-matched control population, results indicate that people with ASDs are poorer at judging intentions in animacy displays. In addition, results reveal an unknown deficit, not seen in the control population, in judging intentions from an elevated position in video displays.

This work may be considered of interest to various groups of people with a wide range of research interests, including the perception and cognition of human motion, the attribution of social intent and "Theory of Mind", and the surveillance of people via video techniques.

Declaration

I declare that this thesis, submitted to the University of Glasgow for the degree of Doctor of Philosophy, is the result of my own research, except where otherwise acknowledged, and that this thesis has not been submitted for a higher degree to any other university or institution.

Signed: …….
(Philip E. McAleer)

Date: ……..

Acknowledgements

I have sought the help and guidance of many people over the course of this thesis, in both an academic and personal sense, and I wish to take a moment to thank them.

Firstly I wish to thank all the participants that have endured hundreds of displays of white circles moving around grey screens, for without you I would never have obtained my data. I would also like to thank the numerous actors that I have used to create my displays including Cat Mulvenna, Andy McCourt, Betty Friedrich, Helena Paterson, Ulla Martins, Andy Paterson, Dave Sheridan, Alan McGhee, and the Italian dancers. Further thanks goes to Nigel Hutchins, Media Support, University of Glasgow, for teaching me everything I ever needed to know about video cameras, and for always being accommodating when I lost them. Thanks to John Shaw for creating the stage that allowed the filming to take place. Thanks must also go to Antonio Camurri, Barbara Mazzarino and Gaultiero Volpe, for the creation of Eyesweb and for helping in programming. Thanks also to the members of the PACO lab who have helped me including Vaia Lestou and Helena Paterson. In addition, thanks goes to Dave and Kathy Forster for providing me with the quiet solitude I needed to write this thesis.

I would like to give great thanks to Marc Becirspahic for his unwavering patience and help with Matlab, and constant supply of coffee. Great thanks also to Lawrie McKay for always answering my statistical problems, and his help along with David Simmons in obtaining data from the ASD population.

Much love and thanks go to my office mates over the years that have endured my rants and vocal attempts, particularly Cat Mulvenna, Lawrie McKay

and Lisa O'Kane, and especially to David McCormack: we will always have Florida man.

Love and thanks also to my parents, Matt and Amelia, my brothers, Richard, Matthew and Gordon, and to the best friends a guy could have: Darren O'Reilly, Peter Fleming, Chris Hannah, Jamie McVey, Vlad Valiente, Dave Sheridan and Alan McGhee: all of you have been responsible for the completion of this opus as much as anyone.

My continual gratitude and love goes to Frank E. Pollick, my supervisor and friend. You encouraged me to start this thesis and you have more than aided me in seeing it through to fruition. Also, thanks for letting me ruin your milking shed.

Finally, I would never have got to the end of this journey without the constant love and support of my partner Lisa. To you I have nothing but unending love and thanks.

The Engineering and Physical Sciences Research Council funded this research.

Contents

Chapter 1 – Introduction

The Perception of Animacy

For many of us, upon opening our eyes to a crowded scene, we would appear to be able to understand the actions and intentions of those around us. Though this ability may seem effortless for most people, beneath the surface there is a complex processing of visual information that is necessary to perform this task. This thesis will examine how people attribute intention to the movement of others. We will examine if the ability to judge intention from human motion is efficient for a range of actions and intentions from differing viewpoints, and investigate, via behavioural experiments, what cues observers use to make these judgements about agency and intention. In order to study a person's ability to judge intention from motion we make use of simple displays, with tractable amounts of information, and a phenomenon known as animacy.

The perception of animacy is the attribution of human-like qualities such as beliefs, desires, intentions and emotions by observers to displays of moving geometric shapes. Depending on how these shapes move, they may be described in ways that would suggest that their movement is intentional or purposeful, as opposed to the perception of the motion being random or of one shape causing the movement of another, such as in Michotte's perception of causality literature (1946). In life we are constantly interacting with the people and objects around us. Whether we are walking through a busy urban street or through a thick forest, it is essential that we are able to make quick and accurate judgements about those that cross our path. We must be able to decide if they are living or not, and if they are friend or foe: allowing us to take appropriate action. The

visual system and the cognitive brain would appear to be able to handle this task reasonably well, as people do not continually walk in to each other, or into stationary objects such as poles. Animacy displays can be used to show how people make such judgements regarding if an object is living or not, and its' intention. Though these displays are sparse in their appearance, they have consistently been shown to elicit from observers, tales regarding social interactions, with shapes having needs and desires as though they were people. As part of this research, a main focus will be on the way that the animacy displays are produced, and their relationship to actual human movement. As will be shown, the common methods for creating these displays involve synthetic approaches, such as computer animation and graphic design. Our attempt to create animacy displays in this research, will involve the recording of humans moving and interacting, and creating the displays from these movements. It is proposed, and will be discussed later, that in doing so, we fill the gap that currently exists between relating results from animacy experiments to real world human actions.

Animacy was first reported in the seminal work of Fritz Heider and Marianne Simmel (1944) who examined the importance of spatial configurations of stimuli in determining how emotions and intentions are attributed to moving geometric shapes. They created a 2½-minute film, see Figure 1.1, that showed 3 shapes, a large triangle, a small triangle and a

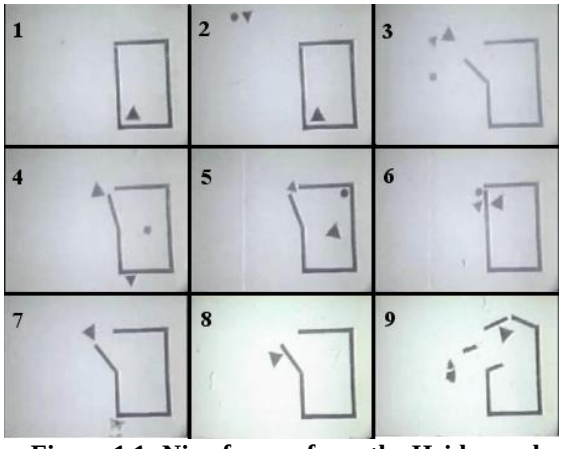

Figure 1.1: Nine frames from the Heider and Simmel movie (1944), depicting geometric shapes moving with human-like qualities.

small disc/circle, moving around a hollow box that had a moving part on one of its sides. The movie was created using a trick-film method, akin to stop-gap or key frame animation, where a frame is filmed, characters are moved and then the next frame is filmed, and so on. The shapes moved in such a manner that on viewing, all but one participant retold the movie as a story involving people interacting, as opposed to geometric shapes aimlessly moving around. The most common type of story was of a love-triangle relationship resulting in a fight between the large triangle and the small triangle, both seen as male, over the helpless feminine circle. Similar stories were found when the response task was either a free response method, asking observers to simply write down what happened in the movie, or a questionnaire task asking questions about each character and their roles in the movie. It was also found that showing the movie in reverse still resulted in animate descriptions but with increased variation as to the actual story, perhaps due to the lack of coherence in the movie now being shown backwards. The experimenters discussed the results in terms of movements that lead to certain percepts, suggesting that relationships between the shapes are interpreted based on the temporal succession and spatial proximity of the shapes, coupled with coherent lines of movement. This suggests a bottom-up processing of the motion. However, the authors also suggested that movements are attributed meanings based on the personalities of the actors, indicating a top-down involvement in the attribution of intentions to motion.

Subsequent work on the topic of animacy has focused on the movement of the shapes and prior knowledge of the characters being portrayed as key to perceiving the shapes as human-like (Hashimoto, 1966; Shor, 1957; Thayer and Schiff, 1969), with Tagiuri (1960) showing that it was possible for observers to

attribute emotions and characteristics to unseen agents, simply from seeing diagrams of their motion paths.

Two papers that show the importance of the spatial and temporal contingencies in the attribution of intention are Bassili (1976), and Berry, Misovich, Kean & Baron (1992). Using differing techniques, these authors demonstrated that spatial and temporal cues create the perception of animacy rather than the aesthetic qualities of the displays. Bassili (1976), using a series of movies involving two circles, one solid and one outlined, performed a systematic analysis of the role of temporal contingencies and spatial configurations. Observers were asked to describe the displays in their own words and then to give ratings of animacy and levels of interaction. The original display showed one circle chasing the other with subsequent movies created by systematically altering both the temporal relationship of the two circles and the spatial configuration. The temporal relationship was changed by altering the delay, short to long, that the second circle moves after the movement of the first circle. The spatial configuration was changed by altering shapes moving on a similar motion path to being randomly displaced. Results indicated that good temporal contingency between the shapes resulted in higher ratings of interaction and that the spatial configurations of the shapes determined the meaning of this interaction.

Berry et al. (1992) showed the importance of global motion cues over aesthetics by using variations of the Heider and Simmel movie. Using a subtractive quantization technique where each frame is converted to a configuration of blocks, of dimension 16 by 16 pixels, they created 3 variations of the Heider and Simmel movie: 1) disruption of shapes and structures but

preserved dynamic cues; 2) preserved structures but disrupted dynamic cues by removing frames; 3) a static version where structural and dynamic aspects are disrupted. They showed the three variations and the original movie to participants in a between design experiment and analysed the free response data in terms of the use of descriptions of human-like actions. Results showed that free responses to the movie where dynamic information was maintained were similar to the original Heider and Simmel study, with participants using themes such as aggression, escape and cooperation. Participants also included more terms relating to human motion in their descriptions of the dynamic displays compared to the displays of variations two and three, however the number of words used was reduced compared to the original Heider and Simmel movie. The authors concluded that the results indicated that the tendency of participants to report a display in terms describing human motion relies heavily on the dynamic properties of the events, rather than structural characteristics of the display.

Dittrich and Lea (1994) investigated the importance of the motion sequences using a movie display that depicted moving letters across the screen. Participants' task was to spot the letter that looked like it was following another. They were either informed to view the display as a lamb seeking its mother, or as a wolf stalking its prey. This difference in prior knowledge would also allow for investigation into the effect of top-down processing via instructions. Dittrich and Lea altered various aspects of the motion of the letters and the appearance, including altering the visibility of the goal, the number of items, the speed and directness of the target and distracters, and the relentlessness of the target movement, i.e., how much effort did the primary agent use to find its goal. The

authors argued that there was a two-stage process in the interpretation of such displays: firstly a bottom-up process where specific motion features are selected, and a later stage where they are visually encoded and conceptually integrated in such a way that top-down processes activate intentional percepts. Their experiments indicated that the perception of intentionality was directly dependent on the variation of movement parameters, such as directedness and speed of target, and the degree of goal orientation. Dittrich and Lea generally found no influence of instructions, which suggested that the perception of intentionality could be a relatively immediate, bottom-up process, probably occurring in the early visual processing stages. They concluded that intentionality is derived from the trajectory of the target's absolute spatiotemporal kinematics, i.e. the relationship between an agent and its background, and that the perception of an interaction is a function of the relative spatiotemporal kinematics, i.e. the relationship between two agents.

More recently, Bloom and Veres (1999) showed that the perception of intentionality could be attributed to groups and not just singular shapes, investigating the manner in which we sometimes give intentions to countries, teams, clouds, etc. in statements such as Britain attacked, and the New York Giants beat, etc. They devised a variation on the Heider and Simmel movie to see whether participants would attribute emotions and intentions to groups of circles and triangles. The authors created a display that followed a similar pattern as the original Heider and Simmel movie with 3 geometric shapes, and also created a variation that had 3 groups of shapes of different colours, see Figure 1.2, (page 10). Each group consisted of 5 homogenous shapes. Bloom and Veres showed the movies to participants both in forward play and reverse

play, along with two control movies, one where the objects are static and another where they are just moving up and down, vertically on the screen. Using a Likert scale response, results showed that there was no significant difference in intentionality perceptions between the group videos and the single shapes video. It was also found, similar to Heider and Simmel (1944) that videos played forward were rated higher on intentionality than those played backwards, again most likely due to an understandable streaming of the display. The control stimuli were either given low or no ratings of intentional attribution. As the group displays were always described in the plural then it was concluded that the participants did in fact see them as different entities but grouped them together based on intentions.

Figure 1.2: Two frames taken from Bloom and Veres (1999) investigating the perception of animacy with groups of objects. Colours based on report.

At the opposite end of the scale, Tremoulet and Feldman (2000) showed that it was possible to achieve strong percepts of animacy from the movement of a single shape in a plain background. The sparseness of the displays in this research would reaffirm the importance of the movement of the shapes as being the major contributor to the percept of animacy. As shown in Figure 1.3, (page 11), they presented a single white shape moving across a circular grey background.

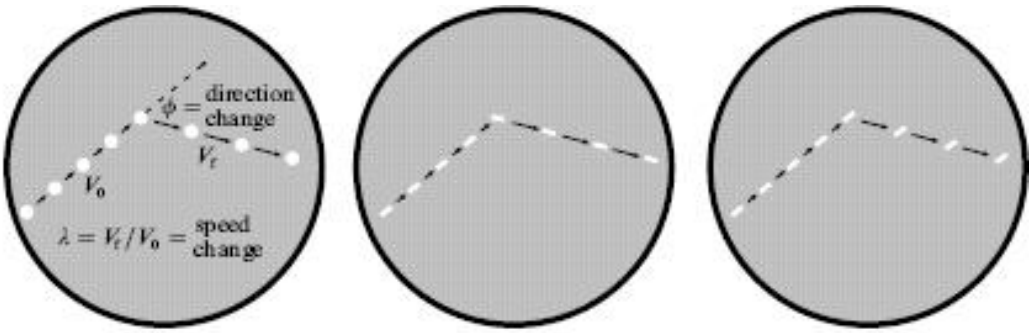

Figure 1.3: From Tremoulet and Feldman (2000) showing the position of the change of direction and speed of a dot, an aligned rectangle, and a misaligned triangle.

Participants were positioned so that they were looking straight down onto the screen and told to believe that they were scientists using a super-sized microscope to determine whether particles were alive. Participants used a seven point Likert scale to rate the perception of animacy. The shape (target) would start at one side of the screen and move across at a constant speed and direction, then at the midpoint the target would change its speed and direction and maintain them until it had left the screen. The directional changes ranged from 0 to 80 degrees from its original path, and the speed changes ranged from 0.5 to 4 times its initial speed. The researchers' main aim was to investigate the hypothesis that observers would classify an object as being animate when the trajectory of that object was unlikely to have occurred in that environment. In other words, they suggested that people would perceive the object as animate if there was no clear explanation from the surrounding environment that would explain the change in the target's speed and direction. Therefore, the target must be under its own control or have an internal energy source, as it is sometimes stated. Furthermore, it was their claim that using long complex trajectories and multiple interactions, as had been used in previous experiments, made it almost impossible to isolate the motion factors responsible for the perception of animacy, and thus they chose to use a single shape with a solitary motion pattern. Using the one target also

allowed for investigation into the importance of the shape of the target and therefore the alignment or misalignment of the target with its trajectory. To investigate alignment they used two shapes: a disc, which has no alignment or an ambiguous alignment, and a rectangle, which may or may not be aligned. Tremoulet and Feldman hypothesised that alignment would increase the rating of animacy. Results showed that changes in direction, speed and alignment all had effects on the ratings of animacy. Animacy ratings were highest in displays that showed a rectangle that was aligned with its trajectory, followed by the disc displays and lowest in the misaligned rectangle displays. It was also shown that targets that had the largest changes in speed and the largest changes in direction were given the highest animacy ratings. Interestingly, shapes that decelerated were rated lower in animacy than shapes that did not change in speed. The authors concluded that unexplainable changes in velocity and direction resulted in changes in ratings of animacy. They argued that participants attributed the targets with the intention to interact with an unseen goal, thus giving explanation to the changes in speed and direction. The basis of the argument is that people are perhaps aware, at a basic level, of Newtonian laws of motion (Newton, 1687): that objects do not just stop and start or change direction without something controlling them, or without the objects themselves being in control of their own movement. The proposal that animacy perceptions are based on a violation of Newtonian laws suggests that whenever a target changes speed or direction for an unexplained reason, and no external force can be attributed, then the target must have an internal force and therefore will be perceived as animate.

Stewart (1982), in her unpublished thesis, is generally attributed with the violation of Newtonian laws hypothesis (Gelman, Durgin & Kaufman, 1995).

Participants viewed displays that showed computer-generated circles moving across a screen. Displays were created where the circles were either consistent or inconsistent with Newtonian laws, resulting in three possible perceptions: animate, inanimate or neutral. For the neutral condition, it was proposed that a circle moving along a straight path at a constant speed would offer no information about what initiated the motion, and therefore no information as to whether it is animate or inanimate. Twenty-one participants were asked to fill in the blanks in sentences about the displays using appropriate nouns. Results indicated that collision events were seen as inanimate by the majority of participants, and were never attributed animate nouns. Displays in which the circles had multiple stops and starts, in the absence of any external cause, were the most likely to elicit animate nouns. Furthermore, displays in which there was a mid-point acceleration, or paired or co-ordinated motion paths, were all good at eliciting animate nouns. However, not all displays produced results that would be consistent with the Newtonian-violation hypothesis. Displays in which a circle would deliberately avoid a barrier, thus avoiding a collision, were not particularly seen as animate, and overall, many of the displays that did violate Newtonian laws were seen as ambiguous. In a follow up experiment, participants were either informed to view the circles in an animate or inanimate manner. Results from this experiment showed that participants were now more likely to attribute animate or inanimate nouns accordingly. This last experiment shows the influence that top-down processing can have on the perception of such displays. Stewart concluded that the perception of animacy involved both perceptual and conceptual processes.

Gelman, Durgin and Kaufman (1995) attempted to enhance the findings of Stewart (1982) by investigating the interaction between trajectories of targets and their surrounding environment, and the resultant impact on the perception of animacy. Using displays similar to Stewart (1982) they presented computer-generated displays of one or two circles (targets) moving across a screen, and collected descriptions and ratings of animacy. The descriptions were extracted using open-ended responses, and the animacy ratings were given using a bar on the computer monitor which participants would click on to give a rating from 0 to 400 of aliveness. The authors introduced various features to the displays: they altered whether the path of the circles was straight or curved; whether it had inflections or not; whether it was one circle or two; whether the circle stopped in view or out of view, i.e. off-screen; and whether there was an object in the local environment and the interaction of the circle in relation to the object. There were four possible situations of environment: "none" – no object; "odd" – the object was located in a place that would require no change in trajectory of the target; "okay" – the object was positioned so that it was acceptable for the target to alter its course; "good" – the object was positioned so that the target had to alter its course. Due to the large number of possibilities of displays, a quasi-Latin-square design was run so each participant saw only 8 different displays. Results of the rating data indicated that animacy ratings were enhanced when there was a purposeful change in direction of the target based on the position of the object, i.e. in the "good" environment condition. However there was no effect of a stopping-starting motion, or of having more than one object. In the free response data, participants gave more animate interpretations to "okay" and "good" environments and gave more inanimate interpretations to "odd" or "none"

environments. Also there was a tendency to describe the movement in animate terms when there were two circles as opposed to one, especially in curved paths. The authors concluded that the characteristics of the trajectories influence the interpretation of the display but they do not determine it, and that perceptions of animacy can be influenced by a good interaction with the surroundings.

Tremoulet and Feldman (2006), in a follow on to their previous study (Tremoulet and Feldman, 2000), recently researched the interaction of environment and target trajectory on the perception of animacy. Using displays similar to their original study, where a target moves along the screen then changes speed and direction at the midpoint, they investigated the influence of the environment in 3 incremental experiments. The basis of the experiment was to examine two underlying hypotheses that are often used to explain the perception of animacy. According to the authors these hypotheses are the Intentionality hypothesis, and the previously mentioned Newtonian-violation hypothesis. The Intentionality hypothesis (Bassili, 1976; Dittrich and Lea, 1994; Gelman, Durgin and Kaufman, 1995; Stewart, 1982) states that animacy is perceived only when intentionality is perceived, though this intention does not have to be a visible one. The Newtonian-violation hypothesis suggests that animacy is perceived when an object's trajectory indicates that the object must have a hidden energy source that can create movement (Stewart, 1982). Though these are similar in ways, the underlying difference is that the Newtonian-violation hypothesis does not depend on how the target interacts with its environment, whereas the Intentionality hypothesis does. In the first experiment to examine these theories, Tremoulet and Feldman (2006) introduced a static object into the environment whose position, relative to the target's trajectory,

could be manipulated. Five conditions were examined: "Prey" - after 375ms the target changes its speed and direction and can now be perceived to be moving towards the static object; "Predator" – after 375ms the target moves away from the object; "Obstacle" – target moves towards object in first half of trajectory but after 375ms looks like it is moving away from the object; "Irrelevant" – object was positioned so that it looked like it had no influence on the target; "None" – no object shown in display. Fifteen participants viewed all displays and used a 7-point Likert scale to give ratings of aliveness of the target. Similar effects of changes in speed and direction were seen as in the previous experiment (Tremoulet and Feldman 2000) but also, an effect of environment was found. The Prey scenario was given the highest animacy ratings, followed by the Predator and Obstacle displays, with the Irrelevant and None displays getting the poorest ratings. The results showed that a minimal contextual cue could influence the perception of animacy, similar to that shown by Gelman et al (1995), and it was concluded that environments that facilitate an intentional interpretation would enhance the perception of animacy. However, it should be noted that the effect of environment was smaller than the effects of speed and direction, indicating that it may have played a lesser part in the interpretation. Experiment 2 was created to examine the effect of the target and the static object coming into contact with each other. Four environmental variations were created: "Goal" – the target moves toward the object after changing speed and direction at the midpoint; "Bounce" – target touches middle of object when it changes direction and moves in a manner that would suggest an inanimate rebound; "Skew" – similar to Bounce condition but target moves off at an angle of 30 degrees more than an inanimate object would move at; "Irrelevant" – target

never touches, moves towards or away from the object. Twelve new participants used the same Likert scale procedure. Results showed that there was no difference in ratings of animacy for the Goal and Irrelevant conditions but both were rated significantly higher than the Bounce and Skew conditions. This suggests that environments that can lend themselves to the perception that a target passively rebounds of an object will reduce the ratings of animacy. It was again found that as final speed increased so did the perception of animacy, and as this happens in a sparse environment then it would appear that acceleration is a very compelling cue to animacy. Experiment 3 aimed to test whether making the acceleration cue at the midpoint appear mechanical could reduce the effect of this cue in the perception of animacy. To achieve this, the authors created four variations that would show the object moving in the first half of the target's trajectory in a way that might create a perception of mechanical motion, see Figure 1.4: Goal – the object moves in a parallel trajectory to what the target would move after the direction change; Bounce and Skew – same as Experiment 2 but object moves into position and therefore looks like it transfers some kinetic energy to the target; Irrelevant – a control condition. A further 12 participants were run and results again showed similar effects of speed, direction and environment changes. The effect of the acceleration cue was not hampered by the cue to mechanical motion,

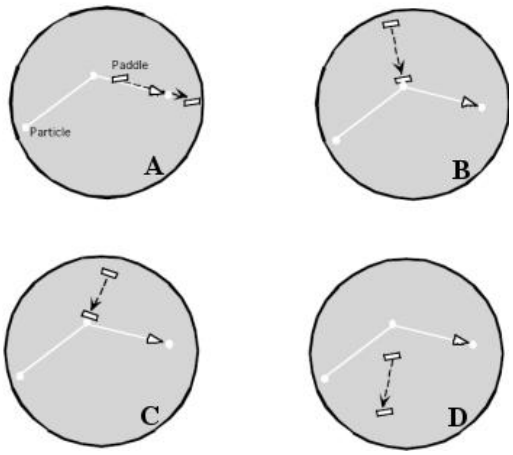

Figure1.4: Four experimental displays from Expt. 3 of Tremoulet and Feldman (2006). White line shows motion of dot (particle) and black hash line shows motion of paddle. The four displays were called: (A) Goal, (B) Bounce, (C) Skew and (D) Irrelevant.

i.e. the collision. The authors concluded that their results showed support for the Intentionality hypothesis, that animacy is perceived if and only if an intentional mental state is inferred to be the cause of the motion of an object. Although the effect of environmental context was small compared to other cues such as acceleration, Tremoulet and Feldman argued that the perception of animacy should be influenced by environment, because the mental state of an agent is normally influenced by its environment and how it will interact with its environment, be it objects or other agents.

To this point the previously mentioned research on the perception of animacy and the attribution of social intentions has focused primarily on the influences of spatial and temporal dynamics and the relationship of the agents to the surrounding environment. The general themes of the displays have ranged from long movies showing a wide variety of intentions, such as in the Heider and Simmel movie, to a single moving shape in the work of Tremoulet and Feldman. An aspect of the perception of animacy and intention that has received little attention is that of which specific actions of behaviour facilitate the ascription of social intentions. This issue was addressed by Blythe, Todd and Miller (1999).

Blythe, Todd and Miller's (1999) goal was to investigate what cues we use to distinguish categories of behaviour and to examine cognitive algorithms for making this process fast and automated. Their research would focus only on what they deemed the most typical forms of intention-driven interactions between two agents. The authors narrowed down animate motion to a few basic survival and reproductive based movements, and assumed these to be general enough that people would be able to perceive and infer the correct intentions from them. The overall goal was to identify useful objective motion cues and the

heuristics used to process them to infer animate motion. They established the intentions of Chasing, Evading, Courting, Being Courted, Fighting and Playing, as being obvious basic dyadic interactions to study. The authors created displays of each of these intentions using a computer game scenario where ten pairs of participants were asked to control two "ants" on a computer screen and make the ants act out these intentions. Ants were used instead of the standard geometric shapes as it was thought they would be more interesting to viewers and furthermore, ants have a clear front and back. The trajectories of the ants were recorded at high spatial and temporal resolutions to allow for subsequent analysis. Analysis of the trajectories revealed that certain intentions have clear characteristics, for example pursuit, evasion and fighting are characterised by high speeds combined with large degrees of turning and looping. Play on the other hand involves a combination of pursuit, evasion, fighting and courting, with loops, rapid dashes, and long still pauses.

Ten participants were shown portions of displays of each intention and asked to categorize them using a six alternative forced choice task. Overall, the success rate was around 50%, chance being 16.67%, suggesting that naïve participants can categorize behaviours based on whole body motions alone. High false alarm rates were seen in play, which the authors suggested is due to an underlying belief that play is more common, and that play has an important role in helping us to learn the other intentions. An interesting follow up experiment showed that if you remove one ant and only show the movement of the remaining ant, then performance drops to 30%, suggesting the importance of the relative motion cues between the two agents. The authors examined what cues are required to distinguish intentions and established seven cues useful for

categorization: (1) relative distance – distance between the two agents; (2) relative angle – angle between an agent's heading and the other agent's position thus showing who is in front; (3) relative heading – difference in heading from one agent to another thus showing if they are facing the same direction; (4) absolute velocity – forward velocity of an agent with respect to the background; (5) relative velocity – difference between two agents' velocities; (6) absolute vorticity – change in heading of an agent with respect to the background; (7) relative vorticity – difference in vorticity between two agents. In order to ensure that these seven cues contained sufficient information to make accurate judgements of intention, the authors trained a neural network on all the experimental motion trajectories using the data of the seven cues as input and the predetermined categories as output. They found the network to have an overall classification performance of 82%. This shows that although the cues do not capture all the information, they do contain enough information to outperform human participants. The authors combined these cues with a simple, fast and frugal algorithm called Classification by Elimination (CBE). A main characteristic of CBE is that it uses the minimal number of cues needed to make accurate judgements of intention, using each cue one at a time in a predetermined order. It is therefore important to enter the cues in the order that will allow for best judgement. Analysis showed that absolute velocity was the best classifier by itself, followed by, in decreasing order, relative angle, relative velocity, relative heading, relative vorticity, absolute vorticity and relative distance. The authors concluded that CBE performed as well as more traditional algorithms, using only half the number of cues that traditional algorithms use, and that CBE

is a much faster cognitive algorithm. They therefore showed these seven cues to have strong diagnostic qualities in the classification of basic dyadic interactions.

A subsequent paper by Barrett, Todd, Miller and Blythe (2005) explored whether the ability to judge these intentions from only the global motion was indeed a universal human ability. They ran a cross-cultural study comparing ability to judge intentions in a German adult population and a Shuar adult population from Amazonian Ecuador. The Shuar people are a hunter-horticulturalist society who had limited experience with technology and many were reported to be surprised to see shapes acting on a screen. As in the previous study, the authors established intentions that would account for a great deal of animate motion, especially intentions that are beneficial for survival reasons. They established six intentional categories: four from the previous experiment, Fighting, Courting, Chasing and Playing, and two new categories of Following and Guarding. Barrett et al (2005) created new displays based on these six intentions using the same computer game technique as before but using arrowheads instead of ants. Using a similar experimental procedure as Blythe et al (1999), results indicated no significant difference in the ability to categorize intentions between the Shuar adults and the German adults. The authors concluded that perhaps the motion schemas for these intentions are not culturally specific and may develop for evolutionary purposes for inferring intentions from motion. The authors again examined the seven cues established in Blythe et al (1999) and found them to have strong diagnostic qualities across these two very distinct social cultures. It was concluded that the ability to infer intentions from global motion is a fundamental cognitive ability and that, by the process of

natural selection, this ability has become fast and accurate, and has ultimately remained universal.

Zacks (2004) examined various features of the global motion of animated geometric shapes in order to establish how people perceive intentions in a series of event segmentation experiments. Previous work, using video displays, has shown that people are better at segmenting events that they have a clear understanding off, or that are common to them, like making a bed (Zacks, Tversky and Iyer, 2001), and that often different people would segment these displays at similar points. It is thought that having an event schema is important for comprehending an activity and for understanding when one activity changes into another (Zacks and Tversky, 2001). This would mean that people, when viewing displays such as the Heider and Simmel movie, know which parts are chases and which parts are fights because they already have an understanding of the movements involved in these intentions. Zacks (2004) introduced a model for understanding how people perceive events and goals. The model states that sensory characteristics are continually processed in a bottom-up manner, providing feature detectors with input, allowing for event segmentation, however, that these sensory cues are controlled in a top-down manner by such things as instructions and prior knowledge, allowing the same cues to lead to different understandings or different event segmentations. According to Zacks (2004), the model revolves around four postulates: (1) that movement features contribute to fine/narrow event segments; (2) that grouping fine segments into larger units can be based on other information rather than just motion features and observers rely less on motion features as the segments get larger; (3) that prior inferences about intentions of agents can affect which motion features

affect segmentation and to what extent; and finally, (4) that inferences about an agent's intention can be influenced by both motion features and top-down information. Zacks tested these postulates through a series of 3 experiments on the role of movement features and intentions in event segmentation using movies that depicted two geometric shapes, Figure 1.5, whose speed, acceleration and positions were controlled by mathematical equations. Throughout the experiments participants were instructed to segment the movies in either a coarse or fine manner, and were informed that the movies depicted either random or goal-directed movement, though the movement was always generated in the same manner. In the 3rd experiment Zacks used a similar method to create movies to that of Barrett, Todd, Miller and Blythe (2005) in order to compare actual goal-directed motion against random motion. The results of the experiments confirmed the postulations of the model. It was found that movement features do help observers determine when an event boundary has occurred but are used more for short events. The features that explained most variance were ones that were related to the distance between agents and the acceleration. Finally, it was shown that observers' attributions of the intentions of agents do appear to rely on an interaction between intrinsic motion features and top-down information provided by prior knowledge or instructions. Similar to previous authors, it was concluded that people could

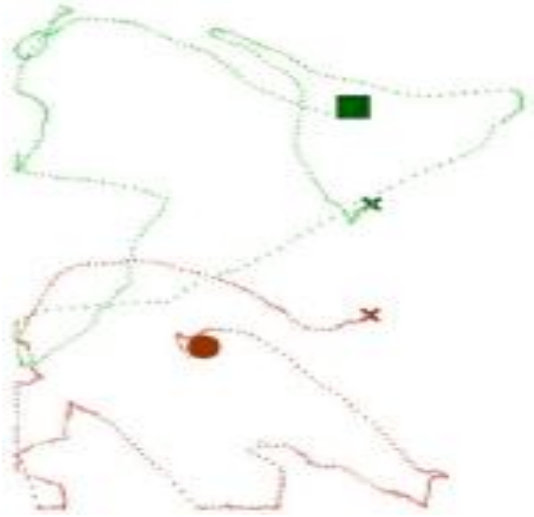

Figure 1.5: Zacks (2004) used the motion trajectories (shown in dash) of a circle and square to examine event segmentation. The motion trajectories were created via mathematical equations.

make use of the physical structure of events and combine this with top-down knowledge to make sense of intentional movement.

This review has so far shown an overview of the general theories and studies that have been explored in the understanding of attribution of social intentions from simple visual displays. It has shown that on viewing displays of simple geometric shapes it is possible for people to describe these displays using animate terms, personify the shapes, giving them intentions, emotions and thoughts. Common thought is that the perception of animacy is due to both a combination of bottom-up processing of motion features and top-down influences of prior knowledge and instructions. Further research, discussed below, will show that this perception of animacy would appear to be a skill that is either innate or is learnt from a young age.

Perception of Animacy in Early Life

It is a common sight to see young children using their toys in a manner that would suggest understanding of intentional motion; a young boy moving two toy cars around in the form of a chase, or a young girl moving dolls around pretending that a party is in full swing. Though this is a step above displays involving moving geometric shapes, anecdotes of childhood playing and imagination would appear to suggest that at a young age, children can understand and imitate intentional motion. Child developmental research, discussed below, suggests that very young children can attribute intention to animacy displays, suggesting that the ability to perceive the intent of others, if not innate, is established early in our development.

Barrett, Todd, Miller and Blythe (2005), using the intentions of Chasing, Fighting, Playing and Following, tested the ability of children aged 3, 4 and 5 years old. The experiment employed a two alternative forced choice, where one answer was correct and the other was randomly selected from the other intentions. Overall hit rate was approximately 60%, with the 4-year-old and 5-year-old children scoring above 60% but the 3 year olds scoring below chance, 50%. This shows that by the age of four we can determine the intentions of moving shapes but not earlier, though this may have been due to either difficulty with coping with the experiment or the intentions being too complex. The 3-year-old children may have recognised the movements as being intentional but not known how to classify them. Work by other authors indicates that very young children, approximately 9 – 12 month, can perceive movement as being intentional and goal-directed (Csibra, Gergely, Biro, Koos and Brockbank, 1999; Dasser, Ulbaek and Premack, 1989; Gergely, Nadasdy, Csibra and Biro, 1995; Kuhlmeier, Wynn and Bloom, 2003; Schlottmann, Allen, Linderoth and Hesketh, 2002).

Gergely et al (1995), using a habituation study, showed that children of 12 months old could attribute intention to the movement of a goal-directed agent, giving it desires and beliefs. The authors created two movies depicting an interaction between a large ball, a small ball and a rectangle, shown in Figure 1.6 (page 26). The first movie showed the small ball on the right of the screen and the large ball on the left of the screen, separated by the large rectangle. At the start, both balls expanded and contracted slightly. The small ball then moved to the rectangle, stopped and moved back. The small ball again moved to the rectangle and then moves over the rectangle to the side of the large ball, moved

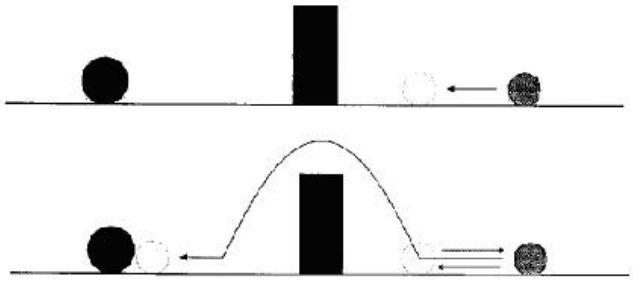

Figure 1.6: Two frames from Gergely et al's (1995) habituation movie. Children were tested on a similar movie, minus the large rectangle. Results showed that young children can understand intentional motion.

next to the large ball, stopped, and the movie ends with both balls expanding and contracting again. This movie was termed the rational approach, as it makes sense for the small ball to go over the rectangle to get to the large ball, its goal. When this display was shown to adults the majority described it as a mother (large ball) calling to its child (small ball) and the child trying to find its way home. The second movie, called the irrational approach, showed the exact same motion trajectories however this time the rectangle was positioned behind the small ball at the start. This motion is irrational because there is no barrier for the small ball to get over yet it still makes the movement of going over the barrier. Two groups of 12 month old children were habituated to these displays in a between groups design and then shown the test movies. There was two test movies: both depicted the small and large balls but no rectangle. The first test movie showed the small ball moving in a direct path to the large ball. The second movie showed the small ball moving in the looping manner seen in the habituation movies to the large ball, as though there was a barrier to overcome. It was hypothesised that if infants saw the small ball as an agent in the rational approach then they would prefer the test movie that showed the small ball moving in a direct path, as this would be a logical path. If the infants did not see the small ball as a rational agent then they would prefer the old movement, as this is what they were familiar with. No hypothesis was made about children

trained on the irrational approach as it was thought that they would not see the small ball as an agent. Results showed that children who had been trained on the rational approach did indeed prefer the logical, direct path of the small ball. The authors pointed to features of the display that may have facilitated the perception of animacy including self-propelled motion and reciprocal activity at a distance, i.e. the expansion and contraction of the balls at the start and end of the movies. It was concluded this experiment was evidence that by the end of the first year, children can attribute intention to the motion of goal-directed, rational agents.

Csibra et al (1999) ran a series of follow-up experiments to test younger children on a similar experimental procedure and to investigate the roles of cues such as self-propulsion. The initial experiment was a carbon copy of the Gergely et al (1995) experiment, however the participants were 6 month olds and 9 month olds. The results showed that the 9-month-old children behaved the same as the 12-month-old children in the previous paper and attributed intentions and desires to the small ball. The 6 month olds, on the other hand, did not attribute any thoughts or desires to the small ball. The second experiment examined whether cues such as self-initiated movement and contraction/expansion, i.e. non-rigid surface movements, affected the attribution of intention. To test this, the contraction/expansion of the balls at the start of the movie was removed, and the small ball started from off screen. The movie started with the large ball next to the rectangle in the centre and the large ball moving away from it to the left edge of the screen. The small ball then appeared on the screen already on its trajectory over the rectangle. When shown to adults it was reported to look as though it was a tennis ball that had been thrown from off the screen. The test phase was again without the rectangle: the large ball moved slightly to its left and the small

ball either came in on a looping trajectory, as though going over a boundary, or in a direct path across the horizon. The authors tested 9 month old and 12 month old children: results showed that both groups did prefer the rational test phase, suggesting that self-propulsion of the small ball is neither necessary nor sufficient for attributing intention to behaviour. The final experiment tested reciprocal activity and used the same habituation movie as Experiment 2 but without the initial movement of the large ball. Again results showed that both groups preferred the rational test phase. It was concluded that rational intentional movement can be perceived in young children even when there is an absence of cues to agency, such as self-propulsion, non-rigid surface motion and reciprocal activity at a distance.

Later work by Kuhlmeier et al (2003) showed that it was possible for 12-month-old infants to make generalisations about the actions of geometric shapes based on what they had seen the shapes do in different situations. Displays were created that depicted a circle trying to reach the top of a hill, with a triangle and a square floating in the sky above the hill, Figure 1.7.

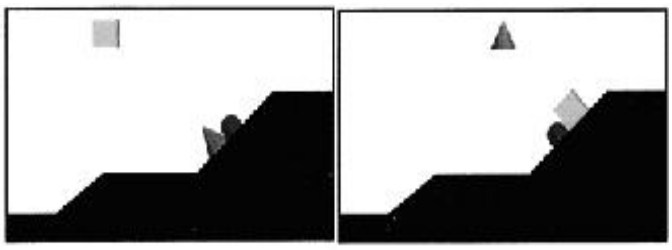

Figure 1.7: Two frames taken from Kuhlmeier et al (2003) showing the circle being helped by the triangle (left) and hindered by the square (right).

Half way up the hill the circle begins to struggle. In one variation of the movie, the triangle comes down and helps the circle to the top of the hill. In the second variation the square comes down and pushes the circle to the bottom of

the hill. In the test phase the hill was removed and only the three shapes were shown. The circle would then either move to beside the triangle or the square. Pilot data with adults confirmed that the circle should like the triangle and dislike the square. The experiment was run as a habituation study on 5 month old and 12 month old children. If the children understood that the triangle helped the circle and therefore it made sense for the circle to go towards the triangle in the test phase, and the reverse for the square, then this would show that they could attribute intention to shapes and generalise it to a different situation. Results, based on looking times, showed that 12 month olds did make this generalisation but 5 month olds did not. The fact that the 5 month olds did not make this generalisation is consistent with Csibra et al (1999) who showed 6 month olds could not attribute rational movement to shapes. An extension of the Kuhlmeier et al study, where the test phase was altered so that either the triangle or the square moved to the circle, suggests that the results are due to previous learned attributes about the behaviours of the shapes and not due to positive associations between the triangle and circle.

These experiments by no means cover the full scope of the literature of the attribution of social intention and agency in child development. Discussing them emphasises the role of animacy displays understanding the onset of the attribution of social intentions. The results clearly show that from a young age, approximately 9 months old, children can infer intentions to simple shapes and motion patterns and that they can use this information to make generalisations about what the shapes will do in different situations. However, though this ability is clear in young children, the research of Barrett et al (2005) suggests that the actual ability to categorize intentions, not just recognise intentional motion,

does not happen until around the age of 3 or 4 years old. This would be consistent with the theory by Frith and Frith (2001) who suggest that all the required tools for processing and understanding intention are present at a young age, and though not efficient in early infancy, rapidly become developed.

A Neurological basis for the Perception of Animacy

Animacy displays are not only useful in determining at what age people are capable of making inferences about intentions, they have also been used in determining areas of the brain that are specifically developed for the understanding of the intensions of others. Though no brain imaging studies are carried out in this thesis, it is important to discuss research showing areas of the brain that have become specialised for attributing intentions. Castelli, Happé, Frith and Frith (2000) ran a Positron Emission Topography (PET) study on healthy adults viewing silent animations designed to elicit mental state attributions to the geometric shapes in the movie. 3 types of animations were explored, see Figure 1.8: (1) Theory of Mind (ToM) – an interaction between two shapes designed to suggest complex mental states, such as deception, where the action of one agent appears to be determined by what the other agent thought; (2) Goal-Directed (GD) – an interaction on a simpler level, where the action of one agent is determined by what the other agent did, such as chasing; (3) Random Action (RA) – the two agents do not interact and the behaviour was not contingent. The ToM displays

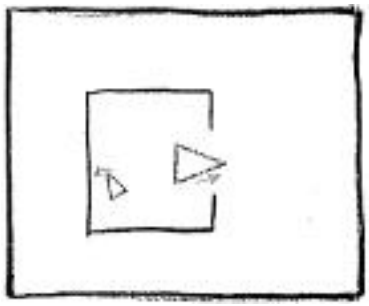

Figure 1.8: A frame from a ToM animation used in Castelli et al (2000). The large triangle was viewed as "mother" and the small triangle as "child". The same displays were also used by Abell et al (2000) and Castelli et al (2002).

showed interactions such as the shapes bluffing each other or mocking each other. The GD displays showed the shapes chasing, imitating, and leading each other. Finally, the RA displays showed the balls bouncing of walls. Six healthy male adults were each shown four different displays of the 3 variations of the animations, divided into 2 blocks: in one block the displays were cued in order to test the effect of prior knowledge. After each display participants were asked, "What was happening in the animation?" The answers that the participants gave were coded on intention, appropriateness of answer, certainty of explanation (i.e. if they hesitated), and length of answer. Results showed that, as expected, more intentionality was given to the ToM displays than the GD displays and finally the RA displays. Longer descriptions were also used for the ToM displays than the other two variations. The PET scan results showed increased activation in four main regions when viewing ToM animations: the medial prefrontal cortex, the temporoparietal junction at the end of the superior temporal sulcus, basal temporal regions including the fusiform gyrus and temporal poles immediately adjacent to the amygdala, and the extrastriate cortex (occipital gyrus). This increased activation was seen in both hemispheres but more so in the right hemisphere except for the medial prefrontal cortex. The authors concluded these areas to form a socialising network in the brain, for the mentalizing of peoples' actions and the attribution of intentions. In addition, similar areas including the fusiform gyrus, the superior temporal, the medial parietal and the dorsomedial prefrontal cortices show changes in activation levels when participants are instructed to passively view video displays of actors performing scenes of social interactions, compared to viewing scenes of no interaction (Frith and Frith, 2003; Iacoboni, Lieberman, Knowlton, Molnar-Szackas, Moritz, Throop and Fiske,

2004). These findings would suggest that similar brain networks are involved in the understanding of social interactions in real-life situations and animated sequences.

Blakemore, Boyer, Pachot-Clouard, Meltzoff, Segebath & Decety (2003) found similar areas of activation in participants viewing goal-directed motion. They studied changes in brain activation using functional Magnetic Resonance Imaging (fMRI) techniques whilst participants viewed movies displaying animate or inanimate motion involving two unusual shapes. The movies always displayed the same two shapes, referred to as the prime mover and the reactive mover, but varied in the contingency of the motion, i.e. whether the behaviour of the reactive mover was resultant of the movement of the primer agent, and also varied in the animacy of the motion, i.e. whether the reactive agent moved as a result of the primer or by itself. Furthermore, the authors varied the focus of the displays for the participants by either drawing participants' attention to the contingency of the motion or to physical aspects of the display, via the experimental instructions. Results showed activation bilaterally in the superior parietal cortex, extending to the intraparietal sulcus in the left hemisphere when attending to the contingency in the animate movements. Furthermore, the right middle frontal gyrus and the left superior temporal sulcus showed increased activation when participants specifically attenuated to the contingent nature of the stimuli. Blakemore et al (2003) reiterated the findings of Castelli et al (2000), suggesting that areas of the brain have become specialised for social understanding.

Schultz, Imamizu, Kawato and Frith (2004) employed a display showing a chase scenario between two agents to explore activation levels in the superior

temporal gyrus during observation of goal attribution. The displays were altered so that the agent chasing had two ways of interacting with the agent that was being chased: the chaser either followed the path of the chased, or would predict where the chased agent would go. It was hypothesised that there would be greater activation in the superior temporal gyrus and sulcus when the chaser appeared to predict and anticipate the motion of the object it was chasing, rather than just following its path. This means that more activation is expected when it is assumed that the chaser can infer about the behaviour of the object it is chasing. Results were consistent with the hypothesis confirming that, bilaterally, these brain regions are involved in the process of inferring intentions in goal-directed motion.

Finally, a study by Heberlein, Adolphs, Pennebaker and Tranel (2003) has shown damage to certain parts of the brain can reduce or remove people's ability to infer intent and agency to movement. Heberlein et al (2003) showed that people with damage to the right somatosensory cortices (RSS), an area known for its role in emotion recognition, used fewer words of emotional and social content when describing the Heider and Simmel movie, compared to normals, even though the number of words used overall were equivalent.

Animacy and Autistic Spectrum Disorders

One of the major areas that animacy displays have been of use is in the understanding of Autistic Spectrum Disorders (ASD). Insight has been gained into the understanding of intentions by people with ASD via behavioural experiments incorporating displays similar to the Heider and Simmel movie, and combining these displays with neuroimaging studies. Again, this following

section is by no means a complete discussion on the field of ASD but serves as a brief overview depicting the importance of animacy displays in the research of Autistic Spectrum Disorders. The theories below, put forward as possible underlying causes of ASD are discussed in greater detail in Chapter 6 of this thesis.

Autism, first diagnosed by Kranner (1943) and Asperger (1944), is regarded as a developmental disorder that is diagnosed when an individual is shown to have abnormalities in a triad of behavioural domains: social development, social communication, and repetitive behaviours and obsessive interests (APA, 1994; WHO, 1994; Wing and Gould, 1979). Although a vast quantity of research has been carried out with the purpose of understanding ASD, the basis of these conditions is still unknown. Main theories include the Empathizing-Systemizing theory (Baron-Cohen, 2002), Weak Central Coherence theory (Frith, 1989) and the Social Brain theory (Brothers, 1990). The Empathizing-Systemizing theory suggests that people with ASD have a disadvantage in empathizing which may explain social and communication problems, and a superiority in systemizing, which may explain traits such as repetitive behaviours, obsessive interests and islets of abilities (Baron-Cohen, Wheelwright, Lawson, Griffin, Ashwin, Billington and Chakrabarti, 2005). Weak Central Coherence theory suggests that individuals with ASD have an inability to bind local details of a scene or situation into a global percept, indicated by their attention to finer details and disregard for the whole. The Social Brain theory, established via single cell studies in animals and neurological studies in humans, finds that areas associated with social understanding, such as the amygdala, medial frontal cortices and the superior

temporal sulcus, show abnormalities, such as reduced activation and differences in mass, in people with ASD compared to typically developing people. Another theory, put forward by Baron-Cohen and Belmonte (2005), is that of neuronal hyperactivity. This suggests that due to a combination of hyperactivity in areas of the brain, reduced connectivity between distal networks of the brain, and inappropriate attentional foci, irrelevant information or noise is not successfully disregarded and therefore the capacity for vital information, be it for dealing with a situation or understanding someone's intent, becomes reduced.

Though the underlying basis of autism is as yet unknown, via Theory of Mind tasks and animacy displays, we can begin to understand and make inferences about the difficulties faced by people with ASD. Theory of Mind is the ability to attribute independent mental states to oneself and others so that we may explain and predict others' behaviour. Premack and Woodruff (1978) coined the term when they showed that chimpanzees would either help or deceive a lab assistant if they thought, or had learned, that the assistant would feed them or take food away. A major characteristic of autism is impairment in Theory of Mind (Baron-Cohen, Leslie and Frith, 1985; Baron-Cohen, Leslie and Frith, 1986) with people with autism ranging on this ability from severe to mild detriment, irrespective of their intelligence. Baron-Cohen et al (1985) showed that children with autism lacked the ability to make inferences about others' beliefs using a false belief task. They compared children with autism to normal children and children with Downs' Syndrome, a condition characterised by reduced mental ability. Children were shown a scenario depicting two dolls, Sally and Anne, with two baskets. Sally placed a marble in her basket and left. Anne took the marble from Sally's basket and moved it to her own. Children

were asked where they thought that Sally would look for the marble. Approximately 80% of the normal children and the children with Downs' Syndrome passed this test as they said that Sally would look in her basket as she had not been present when the marble had been moved. However, around 80% of the children with autism failed the task by saying that Sally would look in Anne's basket, thus not being able to correctly judge the beliefs' of Sally, and showing reduced Theory of Mind. The authors concluded that this demonstrated a cognitive deficit largely independent of general intelligence and may explain the social impairment that people with ASD show, as well as the lack of pretend play that has been noted as another characteristic (Wing and Gould, 1979).

Baron-Cohen et al (1986) ran a follow-up experiment on the same groups of children to investigate the understanding of mechanical, behavioural and intentional movement using a picture-sequencing task. The mechanical displays showed either two objects reacting in a causal manner, or a person and an object reacting in a causal manner, such as a man being knocked over by a rock rolling down a hill. The behavioural displays showed one person moving or two people interacting in ways that required no understanding or knowledge of beliefs or desires, such as a boy getting dressed. Finally, the intentional displays showed two people interacting in manners that required understanding of their beliefs and intentions, such as a girl wondering about where her teddy has gone after a boy has stolen it. The children were given the first picture in the sequence and asked to sort the other pictures into an order that would make sense. They were then asked to tell the story of the pictures. Results showed that in the mechanical and behavioural sequences, children with ASD performed as well as the controls and actually better than controls in the mechanical condition. However they showed

significantly reduced ability on the intentional movement sequences. These results are consistent with theories that show people with ASD to be good at understanding fine, local details but having reduced ability to judge the desires and beliefs of others. In the descriptions of the stories, children with ASD were more inclined to use causal terms or simple descriptive terms of behaviour, as opposed to the children with Downs' Syndrome and the typically developing children who used more terms suggesting mental states and desires.

More recently, animacy displays reminiscent of the original Heider and Simmel movie have been used to give better understanding of the attribution of mental states in people with autism. Abell, Happé and Frith (2000), using the same displays as Castelli et al (2000), explored the use of descriptive words when children viewed these displays. The displays involved the same variations of movement as before, shown in Figure 1.8 (page 30): Theory of Mind sequences, Goal-Directed sequences and Random Action sequences. The authors compared four groups: children with autism, children with general intellectual impairment, typically developing children, and adults. Participants were told to view the triangles differently depending on the type of display: in the Random displays they were told to view the triangles simply as triangles; to view the triangles in the Goal-Directed displays as animals such as ducks; and the triangles in the Theory of Mind displays as people. The task was a free response task to the question, "What happened in the cartoon?" Overall, the children with autism used more interaction descriptions when viewing the random action displays and they gave significantly more inappropriate and fewer appropriate answers, compared to the other groups. Prior to this experiment, some of the children with autism had been shown to be able to pass a false belief

task similar to Baron-Cohen et al (1985). Abell et al (2000) concluded that although some children may be able to pass false belief tasks, this gives no indication that they will be able to attribute the correct mental states to agents in an on-line situation involving moving shapes or agents.

Using the same set of displays, Castelli, Frith, Happé and Frith (2002) performed a Positron Emission Tomography (PET) study on people with autism and controls to look for differences in brain activity during the attribution of metal states. The authors compared 10 adults with autism or Aspergers' Syndrome with 10 typically developed adults. The procedure was the same as Castelli et al (2000). In particular the authors wanted to examine 3 regions that previous work (Castelli et al, 2000) had suggested were involved in the process of attributing mental states: an anterior region of the medial prefrontal cortex; an area in the anterior temporal lobes close to the amygdala; and the superior temporal sulcus (STS) at the temporoparietal junction. Free response data showed that people with ASD used fewer and less appropriate words when describing the Theory of Mind displays, with a tendency to attribute incorrect mental states when they did so. Examining the brain activation levels via cerebral blood flow, people with autism showed reduced activity in the superior temporal sulcus and in the medial prefrontal area, areas both associated with attribution of mental states. The authors found no difference in levels of activation in regions of the extrastriate cortex, compared to the typically developed adults, when viewing Theory of Mind displays. However, using Statistical Parametric Mapping (SPM99, Wellcome Department of Cognitive Neurology, London, UK) it was found that the extrastriate cortex area did show reduced functional connectivity with the STS. Given the role of the STS in

understanding biological motion and mentalizing, the authors suggested that this reduced connectivity might result in a lack of transfer of information resulting in a poor mentalizing of the agents in the displays. Castelli et al (2002) concluded that these results do indeed suggest a physiological aspect of problems with mentalizing in people with autism. The authors go onto suggest that reduced functional connectivity and brain activity may result in a poor combination between the bottom-up processors and the top-down modulators which allow typically developed people to attribute appropriate mental states, thus stopping people with autism from being able to perceive the intention in motion.

Studies by Bertone, Mottron, Jelenic and Faubert (2003; 2005) and Bertone and Faubert (2006) have shown support for a theory that suggests people with ASD have reduced ability in neural processing of motion. Though these studies looked at the processing of first-order and second-order motion patterns, and not the understanding of social movement, the results are important if it is accepted that animacy displays can essential be broken down to motion patterns, and therefore understanding the ability of people with autism to perceive simple motion patterns may be useful. Bertone et al (2003, 2005) created first-order (luminance defined) and second-order (texture defined) motion stimuli of translating, radiating and rotating patterns. First-order patterns are processed by standard motion selective mechanisms in the primary visual cortex (V1), whereas additional neural processing is required for second-order patterns to be perceived, as standard motion mechanisms are unable to process it. For this reason, second order motion patterns can be thought of as being more complex. Bertone et al (2003; 2005) asked their participants to determine the motion direction of the patterns and set a threshold at 75% correct. The results showed

that there was no difference between controls and people with ASD when viewing first-order patterns, however the people with ASD were significantly poorer when processing second-order motion patterns. This effect was also shown for static second-order gratings using a two alternative forced choice task for orientation discrimination (Bertone and Faubert, 2006). Though these results do not specifically enhance our understanding of the ability of people with ASD to mentalize beliefs and desires, they do confirm that people with ASD appear to have problems processing complex motion.

This brief look at studies into Autistic Spectrum Disorders has shown that people with ASD are poor at attributing mental states to agents in social situations and overall appear to be poor in processing complex motion patterns. On the other hand, people with ASD have also been shown to be capable of processing mechanical interactions and simple behavioural interactions that require no knowledge of mental states. Furthermore, results suggest that they are just as capable of processing simple first-order motion patterns as typically developed people. This review has also introduced various theories as to the cause of Autistic Spectrum Disorders, with one of the main current theories suggesting that people with ASD have reduced activity in areas of the brain previously associated with mental attribution and understanding of intention, such as the amygdala and the STS, and that this reduction may be due to a lack of filtering out of irrelevant noise. Animacy displays have been useful in the study of Autistic Spectrum Disorders, and it is one of the main aspects of this thesis to develop a new method of creating such displays that will help reveal unknown aspects of these disorders.

Outlining the Aims of this Thesis

The aforementioned research stands as the basis of the experiments carried out in this thesis. The aim of this thesis is to further the understanding of the perception of animacy and to further enhance the role of such displays in the research of autistic spectrum disorders. One of the main ways that we intend to study this perception is by looking at how the displays are created and in turn, introduce a new method for achieving this. Heider and Simmel (1944) created their display using a means of stopgap animation with a camera, some cut out shapes and a light box. Since then the method of creating these displays has ranged from variations of this display such as in Berry et al (1992) to complex algorithms that control the movement, speed, acceleration, etc. of these displays, such as those used by Zacks (2004). In general, the displays are created by clever forms of animation, or mathematical algorithms, and even by complex computer game systems as seen in the work of Blythe, Todd and Miller (1999). Researching the literature on animacy, there would appear to be a gap in the research for displays that are directly related to human movement. Although all results are described in terms of why people perceive these shapes to be moving as humans, no work has been reported that has used actual human movement as a starting point for the creation of the displays. It is one of the main goals of this thesis to explore a way in which we can track the positional co-ordinates of people and have this motion represented by geometric shapes, in order to examine whether observers will still see these movements as being animate and if they will be able to determine intentions from the movement. To study this, we will make use of a programme called Eyesweb (Camurri, Trocca and Volpe, 2002) that was originally designed for the study of emotions and intentions from

gestures in artistic and dance performances. The role of Eyesweb is introduced in the next chapter and elaborated upon in the experimental chapters. We intend to use a combination of Eyesweb and other programmes to show that it is possible to create animacy displays where the basis of the movement in the displays is derived from video recordings of actual human movement.

Another major aspect of this work will look at the effect of viewpoint on the perception of animacy. The vast majority of previously used displays have depicted the movement from an overhead viewpoint, as though we are looking down onto the shapes (for example Bassili, 1976; Blakemore et al, 2003; Bloom & Veres, 1999; Blythe et al, 1999; Castelli et al, 2002; Gelman et al, 1995; Heider & Simmel, 1944; Tremoulet & Feldman, 2000). Indeed, Tremoulet and Feldman (2006; 2000) specifically instructed their participants to take this viewpoint. Only research involving children seems to incorporate displays that are viewed from the side viewpoint, our natural viewpoint (Csibra et al, 1999; Gergely et al, 1995; Kuhlmeier et al, 2003). In the course of this thesis, the role of viewpoint in the perception of animacy will be examined from two viewpoints: an overhead and a side view. Experiments will compare participants' perceptions of animate movement and intentional movement from these two viewpoints to explore if one viewpoint has an advantage over the other.

All experimental displays in this thesis, apart from the use of an adaptation of the Heider and Simmel movie by Nevarez and Scholl (2000), will be created by tracking human actors to obtain positional co-ordinates and using these co-ordinates to create animacy displays showing shapes performing the movements. The early experimental work in this thesis will use complex dance movements to generate the displays to see if participants will perceive this

complex motion as being animate. From here, work will focus on more simple motions such as walking and running, and simple interactions such as chasing. These movements are generally the motions that people describe when they are asked for a free response to the Heider and Simmel display, and so we will examine if participants perceive these as animate and if they attribute meaning to these movements. This will allow for a comparison between the perception of simple movements to the perception of the complex dance movements, as bases of animacy displays. After experimental examination of these movements, we will begin to look at movements with more precise intentions, and in particular the intentions stipulated by Blythe et al (1999) and Barrett el al (2005) of Chasing, Courting, Fighting, Following, Guarding and Playing. Displays involving these intentions will be explored in behavioral experiments, and in a quantitative analysis, looking at the kinematics of the displays. An attempt will be made to link the movements of the actors to the behavioral data, showing which actions are used to differentiate each intention.

The final part of the thesis will explore the use of animacy displays derived from human motion in studies involving people with Autistic Spectrum Disorders. We will compare the ability of people with Autistic Spectrum Disorders at judging intentions in both animacy displays and the original video displays of the actors, to see if their ability to judge intention changes as the available visual information is reduced or increased. We will also compare the effect of viewpoint on people with Autistic Spectrum Disorders, comparing them with a control population to see if the two populations show similar trends in ability to judge intention across viewpoint. It is believed that this work will

reveal unknown aspects of the ability of people with ASD to understand intentional movement in others.

In summary, the main aim of this thesis is to advance the understanding of the way people attribute intentions to human movement by exploring a new way for creating animacy displays, where the resultant displays are derived from original recordings of human actions. We will look at simple and complex human movements, as well as a series of interactions. We will investigate the perception of intentions from an overhead and side-view perspective and compare the perceptions from these two viewpoints. Finally, we will use these displays to look at the attribution of intentions by people with autistic spectrum disorders, comparing the results to that of a control population.

Chapter 2 – The Eyesweb open platform for multimedia application and motion analysis

The Eyesweb open platform

Fundamental to the creation of animate stimuli in this thesis is the extraction of positional co-ordinates of actors from video displays across time. To achieve this, we make use of the Eyesweb open platform for multimedia application and motion analysis (www.eyesweb.org), developed by the Infomus lab at the University of Genoa, Italy (Camurri, Trocca and Volpe, 2002). This brief chapter will show the underlying basis of the Eyesweb application, its implementation in research, and its role in the research carried out in this thesis.

Eyesweb was designed with the intention of creating a tool that could perform real-time analysis of full-body gestures and movements of one or more persons at a time. Of particular interest in the original design of the programme was the extraction of high-level parameters of expressive intentions in a performance; for example, the developers wanted to create a system that was capable of distinguishing between two performances of the same movement that differed on the emotion expressed in the movement. Hitherto, the dominant use of Eyesweb has been to look at the success of musical performers or dancers in modern dance at expressing emotional content to audiences. The Eyesweb system has thus far been successfully implemented in numerous theatre and museum exhibits (Camurri, Mazzarino and Volpe, 2003), and research is now underway as to assess the validity of this system in the assistance of the treatment of people with disorders of the motor system, such as Parkinson's disease (Camurri, Mazzarino, Volpe, Morasso, Priano and Re, 2003).

Understanding Eyesweb

Eyesweb is a visual programming language that consists of a development area with an accessible set of libraries containing software modules which can be used repeatedly, and can be interconnected with each other to create a processing patch, or series of modules, for motion analysis. A screen shot of an Eyesweb patch can be seen in Figure 2.1, (page 52), at the end of this chapter.

Analysis of movement in Eyesweb is performed using a four-layered bottom-up approach (Camurri, De Poli, Leman and Volpe, 2001) that assumes that the percept of the audience can be explained by the measurable features of the motion and sound in a performance. Layer 1 involves the capturing of the movements and sound with video cameras, etc., and the early processing of this signal such as background subtraction and creation of a Silhouette Motion Image (SMI). Background subtraction relies on a contrast between the static background and the moving actors with the easiest method being colour; for example in the experiments presented in this thesis actors were dressed in white suits performing on a black stage, or vice versa. The output of the background subtraction process is the SMI, an image that carries information about variations in the silhouette shape of each actor and about their relevant positions over the last few frames, Figure 2.2. Layer 2 involves the extraction of low-level features of the movement to create a collection of motion cues that describes the movements being

Figure 2.2: A Silhouette Motion Image (SMI) calculated over four frames. The grey area is the SMI and the dark contour is the current silhouette (Camurri, Trocca and Volpe 2003).

performed. Various motion cues are looked at, including the Contraction Index and the Quantity of Motion. The Quantity of Motion (QoM) is a first, rough approximation of the physical momentum of an actor, found by calculating the area of the SMI over the previous few frames. Therefore, the Quantity of Motion is calculated by multiplying the mass of the silhouette image by the velocity. The Quantity of Motion does depend on the distance of the actors from the camera but the problem is solved by scaling the area of the current SMI of each actor by the area of the most recent succession of silhouette frames, resulting in the measures being independent of the distance of the camera, and expressed as fractions of body area moved. The Contraction Index (CI) is the area surrounding the body and gives an indication of the movement of the limbs: a high CI indicates that the arms are tight into the body. Layers 3 and 4 are more psychologically based aspects of the model and look at the starting and stopping in the movements and how these and the previously determined cues relate to the percept of emotion and expression by the audiences.

The role of Eyesweb in Human Motion Research

Timmers, Camurri and Volpe (2003) and Camurri, Lagerlöf and Volpe (2003) employed Eyesweb to examine the emotional percepts of audiences of piano and dance performances, respectively. This was achieved by correlating measurements of movements by performers, calculated by Eyesweb, such as the QoM, with ratings of emotional engagement from observers in behavioural experiments. Timmers et al (2003) investigated which aspects of a pianist's performance were most strongly related to the emotional engagement of the audience, looking at cues present in the sound and in the pianist's movements.

They created two experimental groups: the first were presented with just the audio sequence and the second group were presented with both the audio and video sequences. Subjects viewed or heard the piece twice. On the first trial the subjects indicated the phrase boundaries of the piece by clicking a button, and on the second trial they indicated their own emotional engagement with the piece by moving a slider up and down. Results indicated a large overlap between aspects of the performance with the emotional engagement of the listeners: high levels of energy in the audio and large movements of the pianist were correlated with high emotional engagement in the listeners.

Camurri, Lagerlöf and Volpe (2003) explored the cues that are important for the recognition of emotions in dance and compared the results of human spectators to automatic techniques for recognizing emotions. Five dancers performed the same dance movement four times, each time employing a different intended emotion: fear, anger, grief and joy. Two sets of subjects viewed the dance performances: the first group responded using a four alternative forced choice, and the second group responded using a multiple choice schemata, rating the emotion of the movement on the four scales of fear, anger, grief and joy. There was a high degree of consistency in intended emotions between dancers when the Quantity of Motion and Contraction Indexes were compared. This suggests that although each dancer performed independently, the manner in which they performed each emotion was similar to the other dancers. Results of the two groups of spectators were similar. Both groups had poor recognition of fear. The highest level of recognition was grief, followed by anger then joy. Analysis showed that there are longer duration times and higher contraction indexes in grief, meaning that the movements are longer over time and the arms

and legs are held close to the torso. The automated models were poor at recognizing joy, and often classified fear as anger, but were better at classifying grief than spectators.

Camurri, Krumhansl, Mazzarino and Volpe (2004) performed a study looking at anticipating human movement in dance. They examined if stopping a dance display at the midpoint of a segment would affect participants' ability to accurately judge where the dancer would finish the movement. Another aspect of the study was to look at the saliency of the barycenter of the dancer as a cue of movement. The barycenter is described as the first-order moment of the 2D silhouette of an actor and is an approximation of the center of mass. It is a means of obtaining a single measure of the combined locations of the torso and limbs. Nineteen clips were extracted from a 3-minute recording of a moving dancer; each clip picked as a coherent unit of movement after discussion between the authors, choreographers and dance experts. The clips were stopped in the middle of the movement at a point of maximum dynamical motion, with the average full duration of the clips being 845msec (range = 350 – 1725msec). The barycenter position over the previous four frames was indicated by a graphical trail. Subjects saw each clip five times and had to indicate the position of the barycenter at the end of the gesture. Results showed that subjects' ability to judge the end position of the barycenter did improve. Although there is some aspect that these results could be partly due to a memory effect, the authors concluded that this experiment does validate the use of the barycenter as a point of information in movement analysis, suggesting that observers can make use of the barycenter to judge the motion of a person.

The role of Eyesweb in this Thesis

In this thesis Eyesweb is used to extract the positional co-ordinates of actors across time from video recordings of their movements. The extracted co-ordinates represent the barycenter of the actors, thus when the actors are viewed from the overhead position this is approximately the center of the heads of the actors, and when they are viewed from the side view this is approximately the stomachs of the actors. We then use these positional co-ordinates to create animacy displays where the actors are represented by geometric shapes.

Though the implementation of Eyesweb has been fundamental to the creation of animacy displays in this thesis, two main problems with the programme were noted during the process of this research that resulted in the final experimental displays being created using a combination of Eyesweb, Matlab (the Mathworks, Natick) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The problems that arose were due to the tracking system employed in the Eyesweb programme. The first problem was that if the actors moved too close together then Eyesweb would average the co-ordinates of the two actors together and the resultant display would appear as though the actors jumped together rather than moved smoothly towards each other. This problem was solved via a low pass Butterworth filter and rendering the displays as QuickTime Movies using a combination of the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) and Matlab (the Mathworks, Natick). Though this technique did smooth the motion of the actors slightly, the final displays appeared more similar to the original displays than the non-filtered displays.

The second problem arose because the tracking system appeared to separate the screen in half, down the vertical midline, with separate trackers tracking either side of this midline. This resulted in confusion of co-ordinates when the actors crossed this midline. Therefore, some of the co-ordinate files would repeatedly become switched between actors whilst extracting the co-ordinates from the displays. In order to solve this problem it was necessary to manually locate the switched regions and reverse them. This was a fairly straightforward procedure, as generally these regions were characterised by large differences in sequential co-ordinates. The details of how the role of Eyesweb altered in the course of the thesis and the relevant aspects of its role will be explained in each experiment.

Eyesweb has shown itself to be a useful tool for the extraction of co-ordinates of human actors from video recordings of movements, particularly in this thesis. Eyesweb is continually developing and being developed not only by its creators but also by its expanding user group. Though this thesis only deals with single actors and dyadic interactions, it is now possible to use Eyesweb to track multiple actors at one time. It is clear that future research will benefit from such an adaptable programme and, in turn, would reveal further potential of this programme in the field of understanding emotions and intentions from human movement.

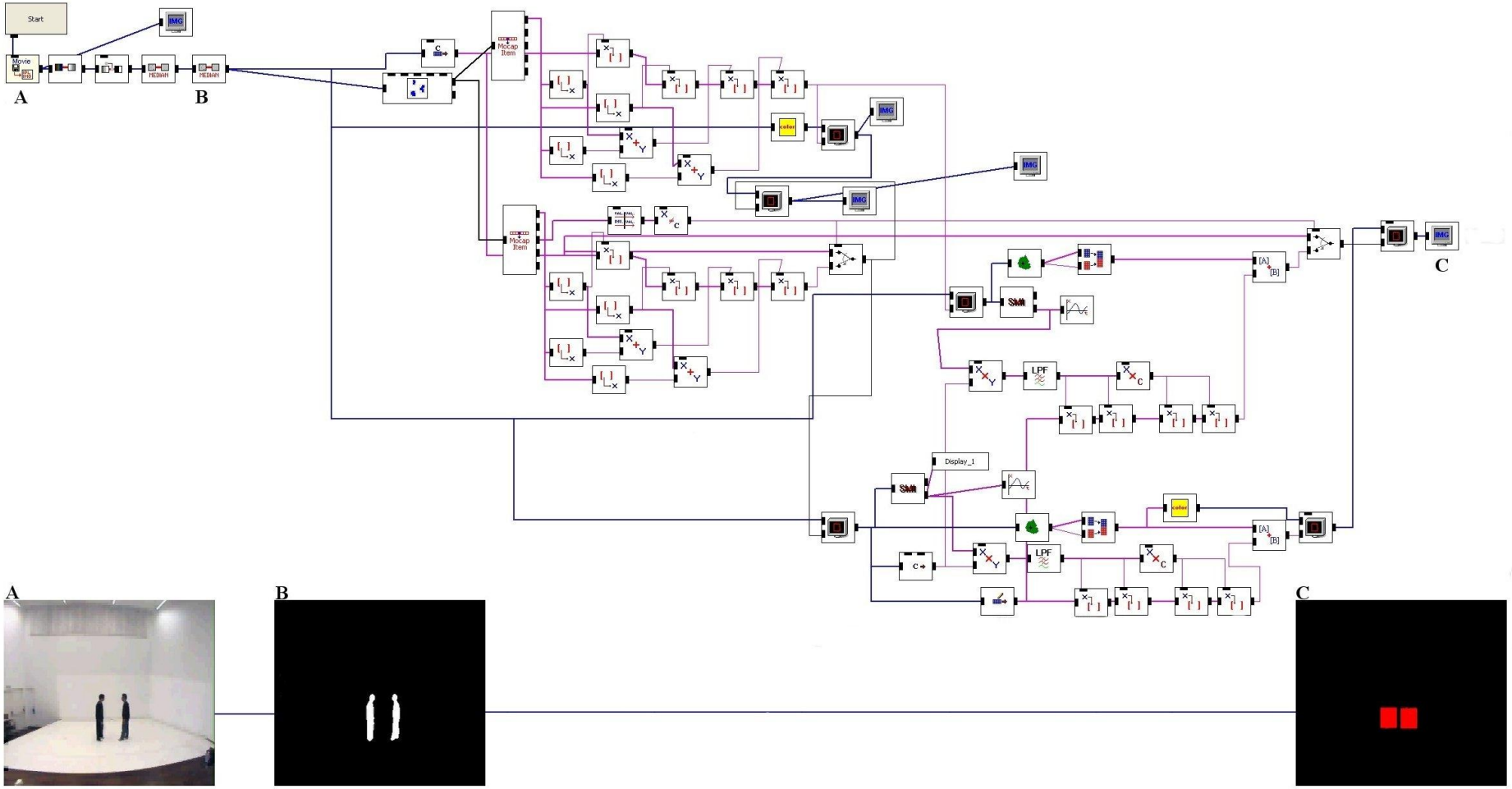

Figure 2.1: Eyesweb screen shot showing the interlinked software modules that complete the patch. The original video enters the process at position A, where it is quickly converted to the silhouette image via background subtraction (position B). From there the number of silhouettes are counted, tracked at the barycenter and the resultant image is created where the actors are represented by squares (position C).

Chapter 3 – Creating Animacy Displays from Human Actions

Chapter Introduction and Overview

Research has shown that in certain situations people will attribute human intentions and emotions to moving geometric shapes, based on the movement of the shapes. The classic work of Heider and Simmel (1944) investigated how observers would describe a video clip that depicted a large triangle, a small triangle and a small disc, moving around a hollow rectangle. All but one of the participants described the clip in terms of an interaction between people and not as three shapes moving aimlessly. Various papers have subsequently found similar results (Bassili, 1976; Berry, Misovich, Kean and Baron, 1992; Bloom & Veres, 1999; Dittrich and Lea, 1994; Gelman, Durgin and Kaufman, 1995; Stewart, 1982; Tremoulet & Feldman, 2006; 2000) and have gone on to show that the perception of animacy is heavily reliant on a combination of prior knowledge and of the spatial and temporal dynamics of the shapes rather than the aesthetic characteristics of the display. Generally animacy displays involve two or more shapes interacting, with subsequent research showing that animacy can be perceived with solitary moving objects (Tremoulet and Feldman, 2006; 2000) or numerous groups of shapes (Bloom and Veres, 1999). Possible cues to the perception of animacy have been researched including self-propulsion, reciprocal movements, non-rigid surface movements, speed and directional changes, acceleration changes, the presence of more than one shape, and prior knowledge (Bassili, 1976; Csibra, Gergely, Biro, Koos and Brockbank, 1999; Gelman, Durgin and Kaufman, 1995; Heider and Simmel, 1944; Stewart, 1982; Tremoulet and Feldman, 2006; 2000). The purpose of this chapter is to examine the

perception of animacy displays containing similar cues, by exploring and developing a new technique for the production of these displays that will firmly base the displays in human movement.

Previous methods for producing displays of animate objects have relied on various techniques including handcrafted shapes and clever camera work (Heider and Simmel, 1944), mathematical equations (Zacks, 2004) or complex computer game scenarios (Blythe, Todd and Miller, 1999). Though these techniques have been successful in producing displays that people will describe as animate, there appears to be no research involving displays that are directly representative of human movement. The main aim of this chapter is to explore a new approach for the production of animacy displays derived from actual human movements. This will be achieved by extracting the positional co-ordinates of movements from recordings of actors, and using these co-ordinates to create displays where the movement of the actors is represented by geometric shapes. This will allow us to examine if it is possible to create animacy displays directly from recordings of human motion, and furthermore, to examine if this technique will work with all human movements, thus exploring what motion properties need to be preserved to obtain animacy.

Following from this, the second aim of the chapter will be to draw comparisons between using complex human movements and simple human movements to create animacy displays. In the first two experiments of this chapter we will use complex modern dance movements to create the animacy displays, and the latter two experiments of the chapter will use simpler movements, such as walking, running and chasing. This comparison will allow

for postulations about what movements are more readily perceptible in animacy displays.

Finally, we will end the chapter by exploring the importance of the intended viewpoint for perceiving animacy displays. The vast majority of displays used in the literature show the shapes moving as though we are looking at them from above, i.e. a bird's eye view (Abell, Happé and Frith, 2000; Bassili, 1976; Blakemore, Boyer, Pachot-Clouard, Meltzoff, Segebath & Decety, 2003; Bloom & Veres, 1999; Blythe et al, 1999; Castelli, Frith, Happé and Frith, 2002; Gelman et al, 1995; Heider & Simmel, 1944; Tremoulet & Feldman, 2000). There are some examples of animacy displays where a side-on viewpoint is the intended percept, our natural viewpoint; however these would appear to be in the minority (Csibra, Gergely, Biro, Koos and Brockbank, 1999; Gergely, Nadasdy, Csibra and Biro, 1995; Kuhlmeier, Wynn and Bloom, 2003). Given that the technique we will employ to create animacy displays starts with the recording of human movement using video cameras, the viewpoint of the resultant animacy display will depend on the positioning of the video camera during the original recording. In turn, multiple video cameras give the possibility of recording the same action/interaction from different perspectives. We will make use of this advantage of the new technique to draw conclusions about the perception of actions and interactions from two viewpoints: an overhead and a side-view viewpoint.

Experiment 3.1 – Creating Animacy Displays from Dyadic Dance Displays

In this experiment we introduce a new technique for the production of animacy displays. We will record dancers performing a modern dance routine

then use the Eyesweb open platform for multimedia application and motion analysis to track the positional co-ordinates of the dancers and create animacy displays where the exact movements of the dancers are represented by shapes. We will then examine, using a free response task, whether participants will describe the motion of these shapes in animate terms, and compare their descriptions to those of the original video recordings. Furthermore, as the process of creating the animacy displays involves a stage of background subtraction where the dancers appear as silhouettes, this will allow a third display condition and we will use these silhouette images to explore the use of animate terms for describing silhouette displays. Finally, a fourth display condition will be used where the size of the shapes representing the actors will increase or decrease in size depending on the Quantity of Motion (QoM) of the dancers. The Quantity of Motion is a first, rough approximation of the physical momentum of an actor calculated by using the mass and velocity of the silhouettes of the dancers over the previous few frames, see below. All of the displays in this experiment will be recorded from a side-view perspective of the dancers.

This experiment will also make use of an on-line slider task to compare changes in the level of arousal experienced by the participants whilst viewing the displays.

It is expected that as the movement in all display conditions is either performed by humans, as in the full video and silhouette conditions, or is directly representative of humans, as in the animacy display conditions, there should be no difference in the use of animate terms between the displays. No expectations are hypothesised for the level of arousal across the displays.

METHODS

3.1.1 PARTICIPANTS

36 participants from the University of Glasgow undergraduate participant pool took part in the experiment. All were naïve to the purpose of the experiment, had normal or corrected-to-normal vision and received course credit for their participation.

3.1.2 STIMULUS PRODUCTION

A modern dance involving two male dancers of similar body structure was recorded using a Panasonic portable video camera mounted on a tripod, positioned approximately 10 feet from the front of the stage. The dance was choreographed by Italian choreographer Giovanni Di Cicco and displayed a man struggling with his inner-self, with one actor playing the man and the second playing his inner-self. Both actors were dressed in black clothing and performed on a white stage. From this original recording we employed the Eyesweb open platform for multimedia application and motion analysis to create four experimental display conditions: (1) original video recording; (2) body silhouettes; (3) Pulsing Blocks – the dancers are represented by red shapes changing in size in accordance with the respective motion energy of the dancer (Camurri, Trocca and Volpe, 2002); (4) Blocks – dancers are represented by red shapes that do not change in size.

Condition 2, body silhouettes, was created by removing colour information and applying a background subtraction technique to the original input video frames. In Condition 3, Pulsing Blocks, the size of the shapes representing the actors are related to the Quantity of Motion (QoM) as measured

by algorithms included in the Eyesweb Expressive Gesture Processing Library (Camurri, Mazzarino and Volpe, 2004). QoM is computed as the change in area of the dancer, in the silhouette format, and is found by first calculating the area of the Silhouette Motion Image (SMI) over the previous few frames (four in this experiment), and then by multiplying the mass of the silhouette images by the velocity. This is done separately for each dancer. Scaling the area of the current SMI of each actor by the area of the most recent succession of silhouette frames, eradicates any discrepancies in the QoM due to the distance from the actor to the camera. QoM can be assumed as a measure of the global amount of detected motion and it can be thought as a first rough approximation of the physical momentum (Camurri, Lagerlöf and Volpe, 2003). Condition 4, Blocks, uses techniques for the tracking of the barycenter of the respective dancer that are programmed in Eyesweb (Camurri, Krumhansl, Mazzarino and Volpe, 2004). In Conditions 3 and 4, the dancers were represented by red shapes on a black background. A representation of the four display conditions can be seen in Figure 3.1.1:

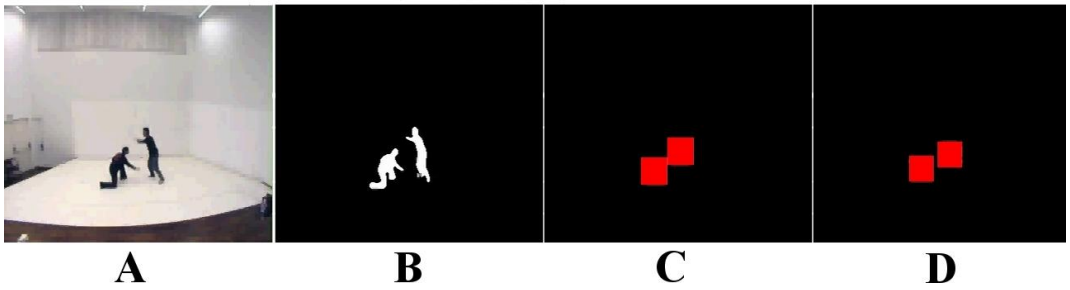

Figure 3.1.1: A frame from each of the experimental display conditions of the dyadic dance display: (A) Original Footage; (B) Body Silhouette; (C) Pulsing Block and (D) Block.

3.1.3 PROCEDURE

Incorporating a between participants design, each participant viewed only one experimental display. Participants were randomly allocated to an experimental display condition, nine per condition. The duration of each display was 1 minute 46 seconds. Whilst watching the display, participants were administered with an on-line slider task. Participants were instructed to move the computer mouse towards and away from themselves to indicate their level of arousal/excitement: half moved the cursor towards for low engagement, away for high engagement, and vice versa. The vertical position of the cursor on the screen, as controlled by the mouse, was tracked and used to give an approximate measurement of arousal. The cursor was centred on the screen at the start of the trial and Eyesweb recorded the position of the cursor at a sample rate of approximately 25 samples/second. A similar method incorporating a foot pedal can be found in Krumhansl and Schenk (1997) to measure emotional engagement in dance and music. After viewing the display participants were given a free response task. They were asked to write a description of the display that they had just viewed, giving as much detail as possible. Participants were never instructed on how to perceive the shapes and thus were in no way biased to describe the displays in animate terms.

The experiment was run using Eyesweb on a Dell Precision 450 computer, and presented at full screen on a 20'' monitor. Screen resolution was set at 1600x1200. Participants were sat at a distance of approximately 1m, giving a starting visual arc of each dancer in conditions 1 and 2 of 6.05', and of 4.6' in conditions 3 and 4. Luminance in the real video displays was measured

as averaging 74.8 cd/m^2. In the silhouette condition, luminance of the black background was 0.33 cd/m^2 and the luminance of the actors was 145 cd/m^2, giving a contrast of 1[6] (Michelson, 1927). In the animacy conditions 3 and 4, the luminance of the red squares was 36.8 cd/m^2, with the black background luminance recorded as 150 cd/m^2, giving a contrast value of approximately 0.7. All displays were shown at 30fps.

3.1.4 RESULTS

This experiment explored changes in levels of arousal/excitement that participants experienced as visual information in displays became more reduced, and examined if a movie depicting two men performing a modern dance could be used to create animacy displays that would elicit animate terms from participants in a free response task. We hypothesised that as all the displays either depicted actual human movement or directly represented human movement, then they would all be described in animate terms. No expectations were hypothesised about changes in the level of arousal/excitement experienced by participants.

Data from the on-line slider task were normalised and averaged (Krumhansl and Schenk, 1997) and the changes in average arousal for each stimulus are shown in Figure 3.1.2. The values plotted are the averages adjusted to an arbitrary range of -1 to 1 to compensate for variation in the use of the slider by participants.

[6] Calculated using Michelson contrast: Lum(max)-Lum(min)/Lum(max)+Lum(min)

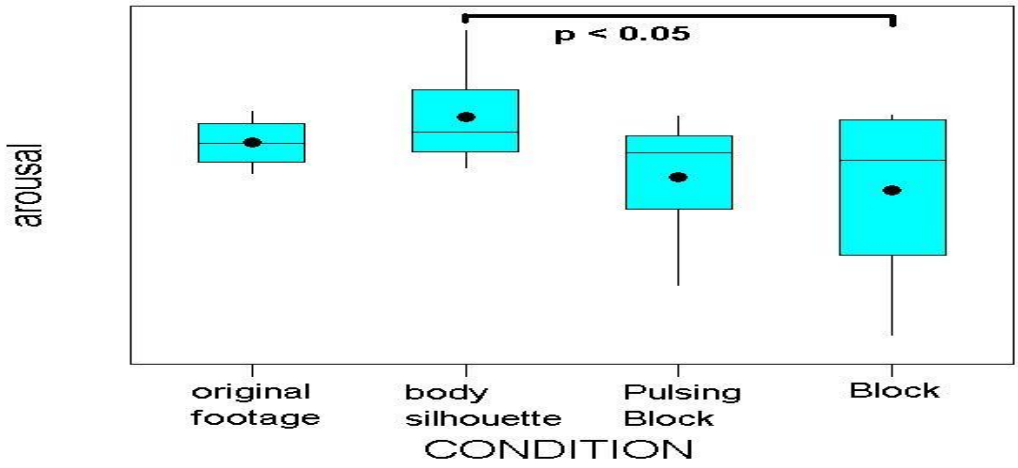

Figure 3.1.2: Average Arousal for participants across each condition with means represented as black dots and medians as lines across the boxes

Average arousal between display conditions was similar across the four experimental display conditions with the body silhouette condition receiving the highest overall rating. A one-way ANOVA looking at levels of arousal across Display Condition (original footage X body silhouette X pulsing block X block) revealed there was a significant effect of Display Condition, $F(3,32) = 3.46$, $\boldsymbol{p < 0.05}$. Post-hoc analysis, using Tukey's HSD, showed that there was a significant difference between the body silhouettes condition and the block condition, where the size of the shapes was fixed. Differing from the expectation of the real video footage being the most arousing, the body silhouette condition had the highest mean arousal.

For the free responses data, similar to Heider and Simmel (1944), we looked for terms and statements that indicated that participants had attributed human movements and characteristics to the shapes. These were terms such as touched, chased, followed, and emotions such as happy or angry. Other guides to animacy were the shapes generally being described in active roles, as opposed

to being controlled in a passive role. Data from the free response task can be seen in Figure 3.1.3:

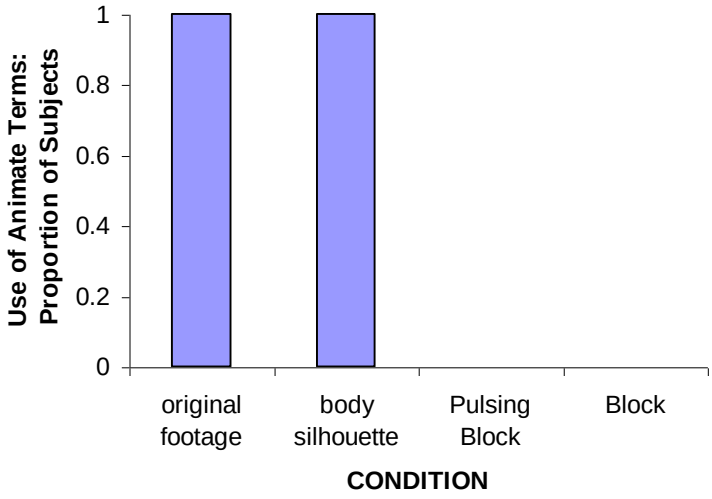

Figure 3.1.3: Occurrence of animate terms used to describe each display condition. The original footage and body silhouettes appeared as two men moving, where as the Pulsing Block and Block conditions appeared as moving red shapes on a black background.

By looking for such terms we concluded that conditions 1 and 2 were described in animate terms but conditions 3 and 4 were not (100% for the original footage and body silhouette displays Vs. 0% for the Pulsing Block and Block displays). Condition 1 was generally seen and described for what it was, a dance:

"I saw two people (male) dancing in a room with a white floor and what appeared to be padded walls (?). Possibly was indicating some type of relationship (the dance seemed to be characterised by the two people coming together and then pushing each other away)?" [Participant 9, Original Footage]

Condition 2, body silhouettes, resulted in confusion as to the purpose of the movement: though always seen as human figures, a proportion of participants reported it as some form of fight between the two men or some form of martial arts:

"Two people were dancing but it sometimes looked as though they were fighting or drunk. Occasionally one of the figures was blocked out. The figures were white and the background was black. It looked like they both died at the end, or they fell over in a drunken mess." [Participant 1, Body Silhouette]

Conditions 3 and 4 were not reported using animate words, but were instead reported merely as shapes moving passively around in varying degrees of speed:

"A red square bouncing/moving about, randomly. Occasionally splitting into two boxes then returning to one. Sometimes it stopped altogether, other times it moved erratically about. When there were two boxes they overlapped, not bounce of each other. Never went to screen extremities." [Participant 8, Pulsing Block]

"A red square was moving on the screen. The movement was very limited at first. Then another red square emerged. The squares kept moving more and then went apart." [Participant 7, Block]

3.1.5 DISCUSSION

This experiment was designed to measure changing arousal levels between the varying visual presentations of a human interaction and to explore the use of recordings of human actions as a means of creating animacy displays

that people would attribute human-like qualities to when describing the motion. It was shown that arousal did vary between conditions and that the body silhouette condition was the most arousing. In relation to the creation of animacy displays from scenes of human actions, as measured by the use of animate terms, only the full video and body silhouette conditions were described as animate. Though no expectations about arousal levels across display conditions was made, there is some suggestion that arousal should have decreased in accordance with decreasing visual information. Thus the full body image would have been most arousing, then silhouette, etc. That the body silhouettes were viewed as most arousing may be explained by the free response task. The silhouette display being viewed as a fight would coincide with the higher arousal than the full video display, which was described as an abstract dance that did not conform to any commonplace social schema, and higher arousal than the two conditions showing only moving shapes. On the other hand, the novelty of seeing the silhouette displays, as opposed to seeing the normal viewing condition, may also have contributed to higher arousal levels, rather than any specific perception of the movement. Furthermore, as contrast and luminance levels were not held constant from one display condition to another, it is possible that fluctuations in these factors may also have contributed to the increased levels of arousal when viewing the silhouette displays, and careful attempts should be made to keep these measures constant in future research, to enable stronger conclusions to be drawn.

The nature of the dance may also have played an important part in the lack of animate terms in describing conditions 3 and 4. The modern dance genre is not one that a lot of people might be familiar with, especially in the abstract

representations and it is therefore possible that the motion of the shapes was not recognisable by people as being typical of human movement. This would suggest that if participants do not recognise aspects of the motion as human then the motion would be viewed as random motion and described as such. If this is true then perhaps a simpler dance sequence may facilitate the use of animate terms.

Experiment 3.2 – Creating Animacy Displays from Monadic Dance Displays

In a follow-on from the previous experiment, we explore if using footage of a solo modern dancer, rather than footage of complex interaction sequence, would facilitate the use of animate terms when describing the display. We also examine the emotional engagement between participants and the display, across the four experimental conditions introduced in the previous experiment. Tremoulet and Feldman (2006; 2000) showed that a single shape, whose direction and speed changes without a visible external force, would be perceived as alive. As the solo dancer display used in this experiment will satisfy these criteria we expect that animate terms will be used to describe all experimental display conditions. We also expect that emotional engagement will change in a similar manner to arousal in the previous experiment with overall levels being slightly higher in display conditions 1 and 2 (original footage and body silhouette), than in 3 and 4 (pulsing block and block).

METHODS

3.2.1 PARTICIPANTS

32 participants from the University of Glasgow undergraduate participant pool took part in the experiment. All were naïve to the purpose of the experiment, had normal or corrected-to-normal vision and received no incentive for their participation.

3.2.2 STIMULUS PRODUCTION

Using an excerpt of a solo dancer performing a modern dance of duration 37 seconds, the four experimental conditions from the previous experiment were created using the same technique: (1) full video recording; (2) body silhouette; (3) motion of the barycentre of the dancer represented by a single small block changing in size in accordance with the motion energy of the dancer; (4) motion of the barycentre of the dancer represented by a small block that did not change in size. In Conditions 3 and 4 the dancer was represented as a white circle on a black background. This was done to eradicate any possible connotations that the colour red may have in social situations. A representation of the four display conditions can be seen in Figure 3.2.2:

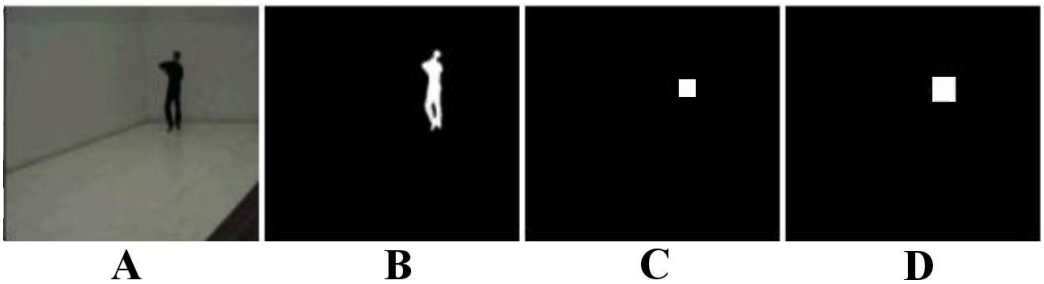

Figure 3.2.2: A frame from each of the experimental display conditions of the monadic dance display: (A) Original Footage; (B) Body Silhouette; (C) Pulsing Block and (D) Block.

3.2.3 PROCEDURE

The same experimental design from Experiment 3.1 was repeated. Each participant was randomly allocated to a display condition group, eight per group, and saw only one display. Similar to the measurement of arousal in the previous experiment, observers were instructed to move the cursor on the screen, via the mouse, towards and away from themselves as their emotional engagement with the clip changed: half moved the cursor towards for low engagement, away for high engagement, and vice versa. The vertical position of the cursor was again used as the approximate measure of emotional engagement. Following the clip, the participants were again instructed to write a free response of what they perceived.

Participants were sat at a distance of approximately 1m, giving a starting visual arc of the dancer in conditions 1 and 2 of 6.9', and of 1.8' in conditions 3 and 4. Luminance in the real video displays was measured as averaging 27.4 cd/m^2. In the silhouette condition, luminance of the black background was 0.21 cd/m^2 and the luminance of the actor was 140 cd/m^2, giving a contrast of 0.98 (Michelson, 1927). In the animacy conditions 3 and 4, the luminance of the white square was 0.28 cd/m^2, with the black background luminance recorded as 145 cd/m^2, giving a contrast value of approximately 1. Displays were shown at 30fps.

3.2.4 RESULTS

We designed this experiment to examine the change in emotional engagement between the participant and the clip as the visual information

available was reduced, and to examine if animate results would be achieved when the animacy displays were created from a monadic dance routine and thus contain only a single dynamic block. Similar to arousal we hypothesised that emotional engagement should reduce with decreasing visual information, and that all conditions of display would result in animacy. Data were again normalised and averaged, and the changes in average emotional engagement for each stimulus are shown below in Figure 3.2.2.

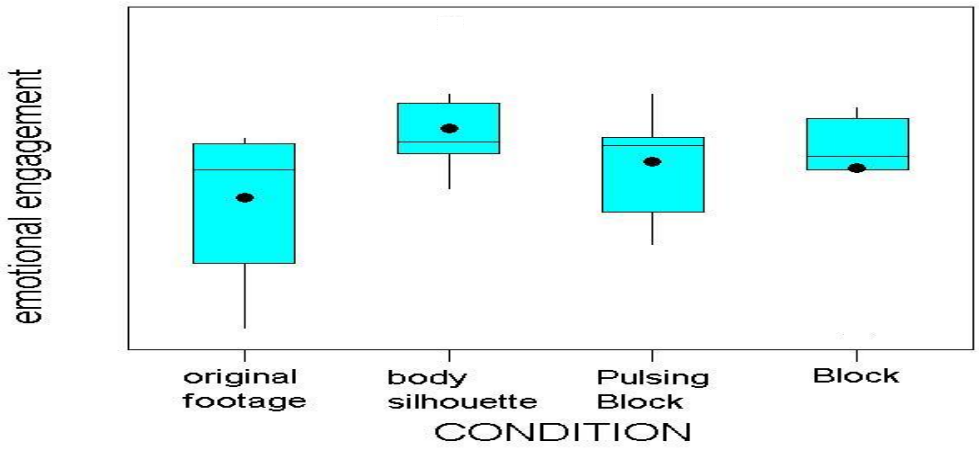

Figure 3.2.2: Average emotional engagement of participants across each condition with means represented as black dots and medians as lines across the boxes. Arbitrary scale of –1 to 1 to compensate for variations in use of slider.

Average emotional engagement for participants viewing the clip was similar across all viewing conditions. Though subjectively from the graph, body silhouette condition would appear to have the largest mean emotional engagement, a one-way ANOVA comparing emotional engagement levels across Display Conditions (original footage X body silhouette X pulsing block X block) showed there was no significant difference between the conditions.

The free response data, Figure 3.2.3, indicated that the original footage and body silhouette conditions were always described in animate terms, however,

the pulsing block and block conditions were not (100% for the original footage and body silhouette displays Vs. 0% for the Pulsing Block and Block displays).

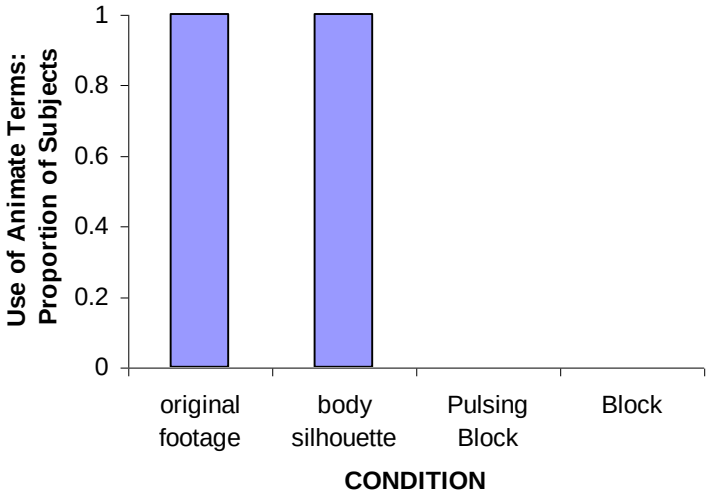

Figure 3.2.3: Occurrence of animate terms used to describe each display condition. The original footage and body silhouettes appeared as one man moving, where as the Pulsing Block and Block conditions appeared as a moving white shape on a black background.

Conditions 1 and 2 were always seen and described as a man dancing by himself:

" A man dancing in an empty room" [Participant 6, original footage]

" I saw a man begin by dancing with lots of movements which were very expressive. Then it moved to a slower pace with less expressive movements and finished with a walk" [Participant 8, body silhouette]

Conditions 3 and 4, where the figure had been reduced to a white rectangle, were reported as a block moving around the screen in altering speeds and directions:

"A white square, increasing and decreasing in size, moving around the black background of the screen" [Participant 2, Pulsing Block]

" A white square moving around slowly" [Participant 5, Block]

The movement of the block in either condition 3 or 4 was never given any purpose or intention, and therefore could not be counted as perceived as animate or alive.

3.2.5 DISCUSSION

The first purpose of this experiment was to examine changes in emotional engagement across the varying visual presentations. Similar to arousal levels in the previous experiment, highest emotional engagement was found when the clip was presented in the body silhouette format. In Experiment 3.1, it was suggested that increased arousal levels for the body silhouette displays could be due to the perception that it was a fight, or possible due to the novelty of the situation. In this experiment, the body silhouettes were described in manners similar to the original footage, i.e. a man dancing. This would suggest that the increases in arousal levels in Experiment 3.1 were indeed due to the perception of a fight, rather than any novelty value, for novelty value should be equivalent across experiments as two different sets of participants were used.

The second purpose of the experiment was to examine if changing the display from a complicated dyadic interaction to a solo dancer would facilitate the use of animate terms being used to describe the animacy display conditions. As was the case in Experiment 3.1, only the original footage and the body silhouette conditions were described in animate terms, with the two abstract conditions, the pulsing and non-pulsing blocks, being described in geometric terms. Changing the display to the movement of one dancer did not result in the use of animate terms for describing the abstract conditions 3 and 4.

Tremoulet and Feldman (2006; 2000) showed that movement of a single dot or rectangle would be described more alive if during the course of a trajectory the shape underwent a change in direction or a change in kinetic energy. The animacy displays produced for this experiment and for Experiment 3.1 both contained moments of these two descriptors yet failed to result in animacy terms being used to describe the displays by participants. These results do not argue against the validity of previous research, as quite clearly an object that suddenly and unexplainably changes direction and speed is likely to be determined controlled by an unseen internal source. There would appear to be qualities lacking in the human motions used to create the displays of the current research that are present in the non-human or pseudo-human motions used to create displays in previous research, as social intentions are more readily perceived in these pseudo-human animacy displays, such as in the work of Heider and Simmel (1944). It is perhaps plausible that in the case of the displays used in the current research, some human actions are much more subtle than the motion of animated objects such as research mentioned in the introduction using displays showing chasing and fighting (Barrett, Todd, Miller and Blythe, 2005; Dittrich and Lea, 1994; Heider and Simmel, 1944). It rather seems the case that when an object lacks human-like qualities, the attribution of animacy relies on an exaggerated display of social behaviour or a limited large deviation in behaviour that cannot be explained by an external source. Subtle and continuous behaviours such as those depicted in our displays do not result in the same kinds of attributions. This is not an unreasonable assumption, since an object blowing in the wind will change direction and speed many times, though when viewed through a window, with no sound or visual cues to the wind, it will still be

appropriate to categorise such an object as something blowing in the wind rather than a self-propelled being. However, it would not be impossible to view such an object as animate, for inhibiting our prior knowledge about such scenarios would indeed allow us to perceive an inanimate object being propelled, as a self-propelled animate agent. Our results therefore suggest that while motion might represent an important cue to animacy, the quality and context of this motion is also important. These features will form the focus of our future research.

Finally, though animacy was not seen in all conditions, we still believe that using this new technique of starting with a full body video and transforming it into basic geometric shapes may result in important information being realised about the attribution of animacy and social meaning. This work serves as a springboard for further study into creating animacy stimuli using this automatic approach.

Experiment 3.3 – Creating Animacy Displays from Common Scenarios

We propose that our attempt to create animacy displays derived from human motion will be validated by using movements that are more common in general life as the starting blocks. This experiment will test if common movements, such as chasing and following, are more likely to facilitate the use of animate terms than the complex dance scenarios used in the previous experiments. The results of Experiments 3.1 and 3.2 showed that the animacy displays created using footage of modern dances did not elicit animate terms. It is suggested that a cause of this could be the unfamiliarity of participants with the movement of modern dance, and that some form of understanding of the

movement or knowledge of modern dance may have increased the likelihood of animate terms being used. That prior knowledge of an event is required to understand the event is a common topic of discussion in the understanding of social movements with Zacks and Tversky (2001) suggesting that a concept of the goal of an event is important to comprehend the meaning of the event.

In this experiment it was decided to use more familiar scenarios that are easily comprehensible to people. The Heider and Simmel display and the free responses to it were analysed for the dominant scenarios that were reported with regularity. Common scenarios included chases, running, following, fighting and descriptions of romantic attraction between the agents. It was decided to attempt to create new animacy stimuli based on these scenarios.

Furthermore, it was suggested that the on-line task used in the previous experiments might have reduced the perception of animacy by distributing attention between the task and the display, and that perhaps undivided attention may facilitate the perception of animacy. As a result of this it was decided to change the on-line slider task to a post-rating of self-propulsion as used by Tremoulet and Feldman (2000).

We hypothesised that as we are now using simple movements that are more common in life and which should be more familiar to people, this would facilitate the use of animate terms used to describe the displays in a free response task. Also, we hypothesised that the post-rating test of self-propulsion would show levels indicating that people perceived these shapes to be self-propelled.

METHOD

3.3.1 PARTICIPANTS

32 participants took part in the experiment. All were naïve to the purpose of the experiment, had normal or corrected to normal vision, and were given course credit or a monetary incentive for taking part.

3.3.2 STIMULUS PRODUCTION

Two actors (1 male) of similar stature were instructed to perform six simple scenarios[7]. The actors performed on an approximately 8 foot long by 2 foot wide, white polythene sheet in front of a white wall, wearing black clothing that exposed only there hands and feet. The contrast between the actors and the surroundings is essential for subsequent video processing and background subtraction. The actors were filmed from their side using a SONY DCR-TRV950E with 12X, f/1.6 optical zoom lens (3.6 - 43.2mm), positioned on a tripod approximately 5 feet high and 8 feet from the centre point of the polythene sheet.

A total of six new scenarios were filmed, four of which involved a dyadic interaction between the actors and the remaining two involved movement by just one of the actors. The scenarios are described as follows, along with the working titles (in bold) that were used to distinguish between the displays: **(1) Social**

[7] At this recording session, three other actors, 2 female and the author, were present and filmed performing similar interactions and situations. However, due to problems with processing of the video footage, only the six situations described were ever used in experiments. The problems with the processing of the video footage have subsequently been eradicated and this footage still exists for later experimental use.

Interaction: actor one is crouched on far right of camera viewing area and actor two is standing up-right on the far left. Actor two approaches actor one who remains crouched. When actor two is within a step of actor one, actor one jumps up from crouched position; actor two acts scared and is chased off screen to the left by actor one; **(2) Attraction:** actors enter from opposite sides of viewing area and walk towards the center of the acting area. When they are positioned roughly a foot from each other, they stop, then proceed to circle three times, as though they are both walking around the same invisible pole, whilst looking at each other, pretending to be romantically interested in each other. Actor two proceeds to walk off screen to the left with actor one a step behind her; **(3) Clap in Middle:** actors enter from opposite sides of viewing area, meet and stop in middle approximately a foot apart. Proceed to act out 3 "High 5's" and then turn and each exit from the side they entered from; **(4) Quick Chase:** both actors start off screen on the left. Actor two chases actor one from starting point to off screen right as fast as possible; **(5) Walk Single:** actor walks from off screen left to off screen right at normal walking pace; **(6) Walk Jog:** actor walks from off screen left to the far right edge of the viewing area, turns to face opposite direction and runs to off screen left as quick as possible.

The four experimental conditions as described in Experiment 3.1 were created for all six afore-mentioned scenarios: original footage, body silhouettes, pulsing blocks and blocks. The actors in conditions 3 and 4 were represented as white shapes on a black background, similar to Experiment 3.2.

The above six scenarios were combined with the dyadic dance display from Experiment 3.1 (**dance x2)**, the monadic dance display from Experiment

3.2 (**dance x1)** and an adaptation of the original Heider and Simmel display (Nevarez & Scholl, 2000), to complete the experimental display set.

3.3.3 PROCEDURE

During the course of the experiment, each scenario (Social Interaction; Attraction; Clap in Middle; Quick Chase; Walk Single; Walk Jog; dance x2; dance x1) was shown once at each experimental condition (original footage; body silhouette; pulsing block; block), except the Heider and Simmel adaptation, which was always shown in its original animated format. The experiment was run as a between design with each participant being randomly assigned to an experimental condition. Each participant saw all 8 scenarios in their assigned experimental display condition, plus the Heider and Simmel adaptation; for example, a participant in the original footage display condition would only see the eight original displays of the two actors/dancer(s), plus the adaptation of Heider and Simmel movie. The order of the displays was randomised for every participant in each group.

Participants were given two tasks: **(1) Free Response Task:** After viewing a display, participants were instructed to write down "What happened in the display?" Participants were given no indication that answers should be written in the context of human movement; **(2) Post-rating of Self-Propulsion:** After viewing each display once and completing the free response task for each display, participants were then shown each display again and asked to give a rating of self-propulsion for the clip on a scale of 1 to 9. A rating of 1 indicated that an unseen force was controlling the actors/block(s), and a rating of 9 indicated that the actors/block(s) were moving around under there own free will.

For both the free response task and the post-rating of self-propulsion task, viewing order of displays was randomised to reduce any order effects. The same experimental apparatus as Experiment 3.2 was used.

Participants were sat at a distance of approximately 1m, giving a starting visual arc of each dancer in conditions 1 and 2 of 8.6', and of 1' in conditions 3 and 4'. Luminance in the real video displays was measured as averaging 84.2 cd/m^2. In the silhouette condition, luminance of the black background was 0.27 cd/m^2 and the luminance of the actors was 152 cd/m^2, giving a contrast of 0.98 (Michelson, 1927). In the animacy conditions 3 and 4, the luminance of the white squares was 0.3 cd/m^2, with the black background luminance recorded as 150 cd/m^2, giving a contrast value of approximately 1. Displays were shown at 30fps.

3.3.4 RESULTS

The aim of this experiment was to create displays that represented the movement of actors performing common, simple human actions to test if viewing these simpler displays would elicit the use of animate terms. Actors were filmed performing scenarios based on everyday human actions and interactions that are similar to those seen in the original display of Heider and Simmel, such as walking and running. We hypothesised that participants would describe these new displays in animate terms, as they depicted simple movements that people could easily relate to. Furthermore, we hypothesised that, in a post-rating self-propulsion task, participants would rate the shapes in the animacy displays as being high in self-propulsion.

The results of the free response task, as measured by the proportion of participants who described the displays in animate terms, are shown in Figure 3.3.1. A response was coded as animate when words indicative of self-propelled motion were used. Words such as chased, touched, followed, flirted, etc. were taken as indication that participants perceived the shapes as moving by themselves and counted as an animate response. Furthermore, in general participants gave very brief descriptions and therefore we accepted a singular occurrence of any animate word in the descriptions as being an indication that a participants was describing the display in animate terms. The results from the Original Footage condition and the Body Silhouette condition are not shown in the graph due to a ceiling effect where all participants described these displays in animate terms:

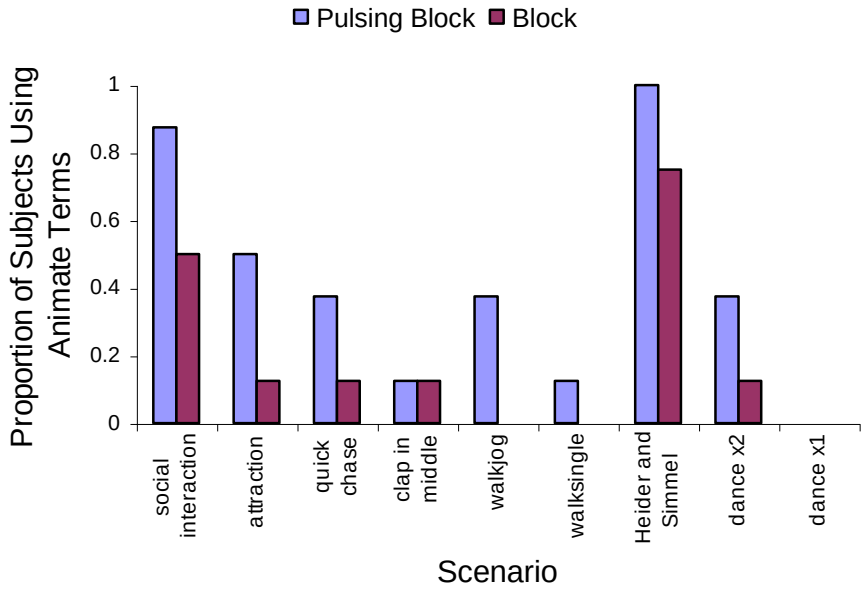

Figure 3.3.1: Proportion of participants using animate terms to describe the scenarios. Only the results for experimental conditions 3 and 4 are shown due to ceiling effects in conditions 1 and 2.

In the graph, though the Heider and Simmel movie was always shown in its original condition, it is coloured to represent how the different experimental groups viewed it, thus the blue bar indicates the proportion of participants in experimental display condition 3 group (Pulsing Block) who described it in animate terms, and the red bar indicates the proportion of participants in experimental display condition 4 group (Block) who described it in animate terms.

The display that was most often described in animate terms was the Heider and Simmel adaptation when viewed by the participants in the Pulsing Block display condition (100%), closely followed by the **Social Interaction** display by the same set of participants (87.5%). Furthermore, the graph suggests a difference in the use of animate terms when describing the Heider and Simmel movie between these two display conditions with the Pulsing Block group more often using animate terms than the Block group (100% Vs. 75%, respectively). Almost every display was described in animate terms by at least one person: the exceptions were Walk Jog and Walk Single in the Block display condition, and the monadic dance display (Dance X1) in both the Pulsing Block and Block conditions. That these 3 displays are poorly described is also indicative that displays involving two actors would appear to be more likely described in animate terms than displays involving one actor, however uneven display numbers makes this suggestion hard to substantiate. Finally, overall there would appear to be a difference in the use of animate terms between these two display conditions presented: the participants that always viewed the Pulsing Blocks used animate terms more often than the participants who only viewed the structurally rigid blocks (42% Vs. 19%). A Mann-Whitney Non-Parametric test confirmed

that there was a significantly larger use of animate terms by the participant group that viewed the Pulsing Block condition, compared to the Block condition (U = 10.5, $p<0.05$, two-tailed). Results would indicate that this method for the production of animacy displays is a valid technique for producing displays that will be described in animate terms.

The results of the post-rating task of self-propulsion can be seen in Figure 3.3.2. The data from all four experimental conditions have been included. In the self-propulsion task, a rating of 1 indicated that the participants thought an external force was controlling the actors/block, whereas a rating of 9 indicated that participants thought the actors/blocks were moving under their own control.

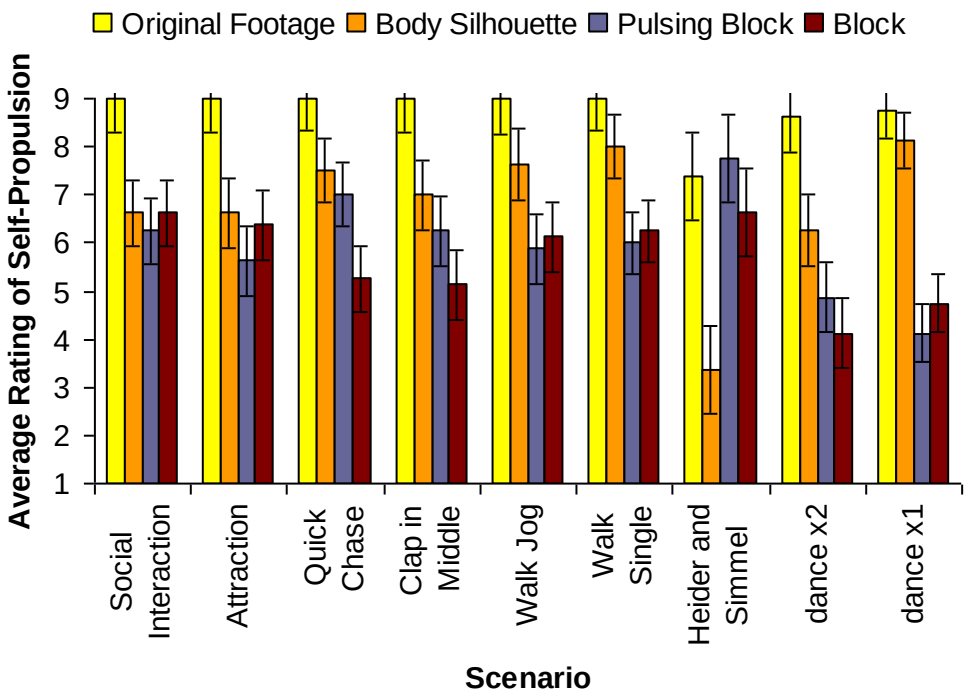

Figure 3.3.2: Average ratings of self-propulsion across scenarios, errors bars indicate standard error of the means. A rating of 9 indicates that participants perceived the agents to be free willed and a rating of 1 indicated that participants perceived the agents to be controlled by an unseen external force.

It is clear from the graph that all original footage displays (excluding the Heider and Simmel display) were given high ratings of self-propulsion: not surprising as the displays showed actors moving freely. However, there would appear to be more variability in the ratings of self-propulsion across scenarios in the remaining display conditions.

A two way mixed design ANOVA was run with Display Condition (Original Footage X Body Silhouette X Pulsing Block X Block) as a between factor, and Scenario as a within factor (Social Interaction X Attraction X Quick Chase X Clap in Middle X Walk Jog X Walk Single X Heider and Simmel X Dancex2 X Dancex1). The ANOVA revealed a main effect of Display Condition: $F(3,28) = 14.12$, $\mathbf{p<0.05}$ with Fischer's LSD showing a significant difference between the Original Footage displays and all other display conditions: 8.8 (Original Footage) Vs 6.8 (Body Silhouette); Vs 6 (Pulsing Block); Vs 5.7 (Block), all at the $p<0.05$ significance level. After correcting for Sphericity using Greenhouse-Geisser, the ANOVA showed no main effect of Scenario. An interaction was found between Display Condition and Scenario, $F(15.2, 141.9) = 2.66$, $\mathbf{p<0.01}$ with post-hoc analysis using Tukey's HSD revealing multiple significant differences. These are summarised in Table 3.3.1 (page 83), with mean ratings for each Scenario by Display Condition shown in Table 3.3.2 (page 82):

	Original Footage	Body Silhouette	Pulsing Block	Block
Social Interaction	9	6.63	6.25	6.63
Attraction	9	6.63	5.63	6.38
Quick Chase	9	7.5	7	5.25
Clap in Middle	9	7	6.25	5.13
Walk Jog	9	7.63	5.88	5.13
Walk Single	9	8	6	6.25
Heider & Simmel	7.38	3.38	7.75	6.63
Dance X2	8.63	6.25	4.88	4.13
Dance X1	8.75	8.13	4.13	4.75

Table 3.3.2: Summary of the mean ratings of self-propulsion for each scenario (rows) at all display conditions (columns).

		Body Silhouette									Pulsing Block									Block								
		S.I.	Att	Q.C	C.I.M	W.J	W.S	H&S	DX2	DX1	S.I.	Att	Q.C	C.I.M	W.J	W.S	H&S	DX2	DX1	S.I.	Att	Q.C	C.I.M	W.J	W.S	H&S	DX2	DX1
Original Footage	S.I.	+	+		+			+	+		+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
	Att	+	+		+			+	+		+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
	Q.C	+	+		+			+	+		+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
	C.I.M	+	+		+			+	+		+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
	W.J	+	+		+			+	+		+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
	W.S	+	+		+			+	+		+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+
	H&S							+										+	+			+	+				+	+
	DX2	+	+					+	+		+	+		+	+	+		+	+	+	+	+	+	+	+	+	+	+
	DX1	+	+					+	+		+	+		+	+	+		+	+	+	+	+	+	+	+	+	+	+
Body Silhouette	S.I.							+											+									+	+
	Att							+											+									+	+
	Q.C							+										+	+			+	+				+	+
	C.I.M							+										+	+								+	+
	W.J							+				+			+			+	+			+	+				+	+
	W.S							+				+			+	+		+	+			+	+				+	+
	H&S								-	-	-	-	-	-	-	-	-			-	-			-	-	-		
	DX2																		+								+	
	DX1											+			+			+	+			+	+	+			+	+
Pulsing Block	S.I.																											
	Att																											
	Q.C																											
	C.I.M																		+								+	
	W.J																											
	W.S																											
	H&S																	+	+			+	+				+	+
	DX2																											
	DX1																			+	+			+	+	+		
Block	S.I.																										+	
	Att																										+	
	Q.C																										+	
	C.I.M																											
	W.J																										+	
	W.S																										+	
	H&S																										+	
	DX2																											
	DX1																											

Table 3.3.1: Summary of all significant differences as found using Tukey HSD. Original Footage self-comparisons are not displayed, as there were no significant differences found. A (+) indicates that the display in the row was rated significantly higher than the display in the column, whereas a (-) indicates that the display in the row was rated significantly lower than the display in the column. Abbreviations for displays are as follows: S.I. = Social Interaction; Att. = Attraction; Q.C. = Quick Chase; C.I.M. Clap in Middle; W.J. = Walk Jog; W.S. = Walk Single; H&S = Heider and Simmel; DX2 = Dance X2; DX1 = Dance X1. All differences are significant at the $p<0.05$ level.

Post-hoc analysis shows that the Original Footage displays were rated significantly higher than a large number of other displays. Between the remaining Display Conditions there were fewer significant differences found, with the poorest rated displays, for self-propulsion, being the dyadic and monadic dance displays in the Pulsing Block and Block conditions, and the Heider and Simmel display in the Body Silhouette condition.

3.3.4 DISCUSSION

This experiment tested the use of common scenarios for the creation of animacy displays that when shown to participants would be described in animate terms such as chase, follow, touch, etc. In order to do this we filmed simple everyday scenarios and interactions that were influenced and inspired by the original display created by Heider and Simmel. It was hypothesised that depicting simple displays would facilitate the use of animate terms, compared to the complex modern dance scenarios that were used in the two previous experiments. We also incorporated a post-rating of self-propulsion and hypothesised that participants would rate the animacy displays of this experiment as being high in self-propulsion.

From the results of the free response task we showed that this technique for the synthesis of animacy displays is efficient at producing displays that, when shown to people, will be described in animate terms. Though it was the adaptation of the Heider and Simmel display, by Nevarez and Scholl, that elicited the most animate terms and the more story-like responses, the **Social Interaction display** was described in animate terms by a similar number of people. This display depicted an actor crouching and another approaching. When the

approaching actor got close enough the crouching actor jumped up and chased the second actor off: a representation of this scenario at each display condition can be seen in Figure 3.3.3:

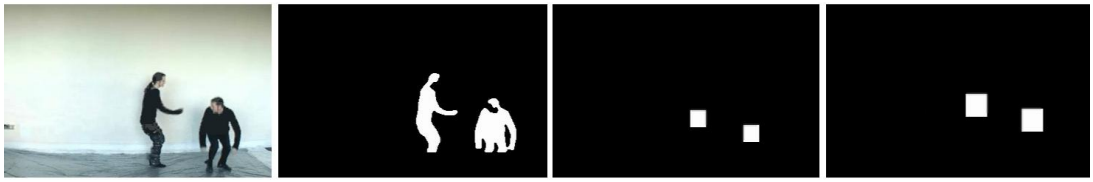

Figure 3.3.3: A frame from each of the experimental display conditions of the Social Interaction display at the point where the actor on the right is jumping out of the crouched position to chase the actor on the left: (A) Original Footage; (B) Body Silhouette; (C) Pulsing Block; (D) Block. The lower legs are missing in the Body Silhouette condition as a mask was placed over image to cover the glare of the floor.

When this display was depicted as animate blocks, it was often described as one block (the crouching one on the right) feigning injury and encouraging the second block to come closer, then eventually chasing them off for one reason or another. All other displays, bar the display involving one dancer performing a modern dance, were described in animate terms, albeit in varying incidence and depth.

The ratings of self-propulsion task showed that the Original Footage displays were rated significantly higher than a large proportion of other displays but less variability was found across the displays of the remaining display conditions: Body Silhouette, Pulsing Block and Block. Apart from the dyadic and monadic dance displays of the Pulsing Block and Block conditions, and the Heider and Simmel display in the Body Silhouette, all displays were rated towards the self-propelled end of the scale. It is interesting that the Heider and Simmel display in the Body Silhouette condition should receive such a low rating (3.38), as in all other display conditions it received high ratings. It may be that

this is an effect of low participant numbers but it would intriguing to examine if this result were due to an inhibition/facilitation effect. It is possible that the participants in the Original Footage rate this display high because they are influenced by their ratings for other displays in the condition, and participants in the Pulsing Block and Block condition rate this display high as it is the best of the animacy displays. This however does not explain why the Heider and Simmel display would get such a low rating in the Silhouette display, but it may be due to some cognitive comparison between the cartoon-esque human representations in the silhouettes and the geometric shapes in the Heider and Simmel display. However, the effects of changes in levels of contrast and luminance can not be ruled out as they were not held constant across the four display conditions, and may account for some variation in the ratings of self-propulsion.

The results of the two tasks, when taken together show that even when a display is not described in animate terms, it may still receive a high rating of self-propulsion. This fact could point one of two ways: either that the free response task is not a particularly sensitive task and another task should be sought, or that some displays contain cues that are better at eliciting animate responses than others. This argument is open to debate, but one criticism of the free response task that has been pointed out by participants during experimental debriefs, was that participants did sometimes think about the displays in a human context but were a bit embarrassed to write it down, so instead described the display in a geometric context.

What can be taken from these results is that in this group of scenarios, some scenarios result in better animacy displays for eliciting animate responses

than others, as shown by displays like Social Interaction and Attraction often being described in animate terms, and the monadic dance display never being described in animate terms. Why one display should work better is unclear, however an important aspect would appear to be attention. Of particular note would be the increase in the use of animate terms used to describe the dyadic dance display from Experiment 3.1 to this experiment. As mentioned in Experiment 3.1 and 3.2, it was suggested that the online slider task might have reduced observers' ability to perceive the displays as animate by dividing their attention between viewing the display and moving the slider. In this experiment, with the removal of the slider task, we find that the dyadic dance display from Experiment 3.1 is now described in animate terms by observers. This would indicate that some aspect of perceiving geometric shapes as animate relies on observers giving the displays their full attention. Of course this could just be as a result of new participants, but it could also point to the importance of observers giving their full attention to the display to understand what is happening.

Understanding the scenario may also be an important aspect in the attribution of animate terms. Zacks et al (2001) suggested that the concept of a goal is required for one to comprehend another's actions. This would seem to offer an explanation as to why these new simple displays depicting everyday scenarios are described in animate terms and more complicated displays that tend not to follow social schemas, in particular, the monadic dance display, would not be described in animate terms.

The results from the free response task may also indicate a slight difference between displays depicting one actor and those displaying two actors. Displays showing just one actor appeared to be described fewer times in animate

terms, compared to those depicting two actors, suggesting that perhaps an animacy display depicting at least a dyadic interaction is better at facilitating animacy responses. Though it must also be pointed out that even if a display showing one actor is not described in animate terms, the shape may still be perceived to be self-propelled, and that this finding is in keeping with the work of Tremoulet and Feldman (2000).

Finally, the results show an increase in the use of animate terms for displays that show non-rigid surface motion of agents as indicated by the increased use of animate words for blocks that changed in size in accordance with the Quantity of Motion. This was noted in the work of Gergely, Nadasdy, Csibra & Biro, (1995) and Csibra, Gergely, Biro, Koos & Brockbank, (1999). These authors tested reciprocal expansions and contractions of the surface of agents and concluded that it was not essential for the perception of animacy. The results of Experiment 3.3 support this finding that it is not essential for the perception of animacy, as displays where the surface is rigid are also described in animate terms, however this experiment does show an increase in the use of terms where the surface of the agents is non-rigid, however it is unclear as to what perceptual cue this non-rigid surface motion is giving. It is possible that the change in surface area of the shapes is perceived in a similar manner to a person moving their limbs, yet on the other hand, it could be just as easily giving a false cue to depth, that the shape is coming near and far and getting bigger and smaller. In Csibra et al and Gergely et al, due to the given horizon, it is probably the former percept, but in this experiment, due to the sparse environment in which the shapes are moving, it is at present difficult to determine without further investigation.

In conclusion, this experiment has shown that this technique to create animacy displays, using human motion as a starting block, can produce displays that will elicit animate terms from people when asked to describe them. We have shown that displays involving simple scenarios and interactions do seem to make better animacy displays involving complex motion patterns. Furthermore, that the use of animate terms used to describe the displays in a free response task appears to be influenced by the number of people in the display, and by non-rigid surface movements.

Experiment 3.4 – The Importance of Viewpoint in Animacy Displays

The final experiment of this chapter looked to introduce a new line of research in animacy literature by exploring differences in perception of animacy displays, depending on the viewpoint from which they are viewed. In animacy research, some studies have used animacy displays from a side view, particularly in child development (Csibra, Gergely, Biro, Koos & Brockbank, 1999; Gergely, Nadasdy, Csibra & Biro, 1995; Kuhlmeier, Wynn & Bloom, 2003) however the majority of studies utilising animacy displays have incorporated displays solely from the overhead (Bassili, 1976; Blakemore, Boyer, Pachot-Clouard, Meltzoff, Segebarth & Decety, 2003; Bloom & Veres, 1999; Blythe et al, 1999; Castelli, Frith, Happé & Frith, 2002; Gelman et al, 1995; Heider & Simmel, 1944; Tremoulet & Feldman, 2000; Zacks 2004). Previous research in this chapter has so far utilised displays that show the motion of the agents from a side viewpoint. With this new method for stimuli, it is possible to systematically compare the same interaction from differing viewpoints. It was our intent in this experiment

to compare ratings of self-propulsion between displays from two viewpoints – an overhead perspective and a side-view perspective. We hypothesised that there would be no difference between ratings of self-propulsion for displays from different viewpoints.

METHOD

3.4.1 PARTICIPANTS

18 new participants[8] were recruited from the undergraduate participant pool. All were naïve to the purpose of the experiment, had normal or corrected to normal vision, and all received either course credit or a monetary incentive for taking part.

3.4.2 STIMULUS PRODUCTION

2 new actors (1 male) of similar stature were filmed on a purpose built 5m square stage with a black floor and three black walls, one on either side and one at the back. The actors were dressed in white body suits, including hoods. Two video cameras were used to film the scenes: 1 camera positioned directly above the centre of the stage at a height of 6 metres; 1 camera positioned on a tripod at approximately 5 metres from the centre of the front edge of the stage. The side camera, a JVC GR-DV700EK with 10X, f/1.2 optical zoom lens (3.8 – 38mm), captured the complete horizontal span of the stage, while the overhead camera, a SONY DCR-TRV950E with 12X, f/1.6 optical zoom lens (3.6 - 43.2mm), fitted with a 0.3x magnification wide-angle lens adaptor, captured the

[8] The number of participants used in each experiment alters with the design of the experiment on the thinking that in a between subject design experiment, more participants are required to perform statistical analysis, and when a within subject design is used, fewer participants are required to perform statistical analysis, given the repeated measures for each display combination.

entire stage. The footage was extracted from the cameras using Windows Movie Maker. The positional co-ordinates of each actor were extracted using Eyesweb, and a combination of Matlab (the Mathworks, Natick) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) were used to create the experimental stimuli. For this experiment only the fourth experimental condition of the previous experiments, Block, was used; white circles of a fixed dimension on a black background represent the actors. The third experimental display condition, where the shape representing the actors changes in size in accordance with their Quantity of Motion, was excluded until further research can be carried out to ascertain if this cue is used by observers as a representation of Quantity of Motion, as intended, or if observers use this cue, incorrectly, as a depth cue. Figure 3.4.1 depicts a frame from a scenario in its original format and its equivalent animacy display, shown from both viewpoints (overhead and side view):

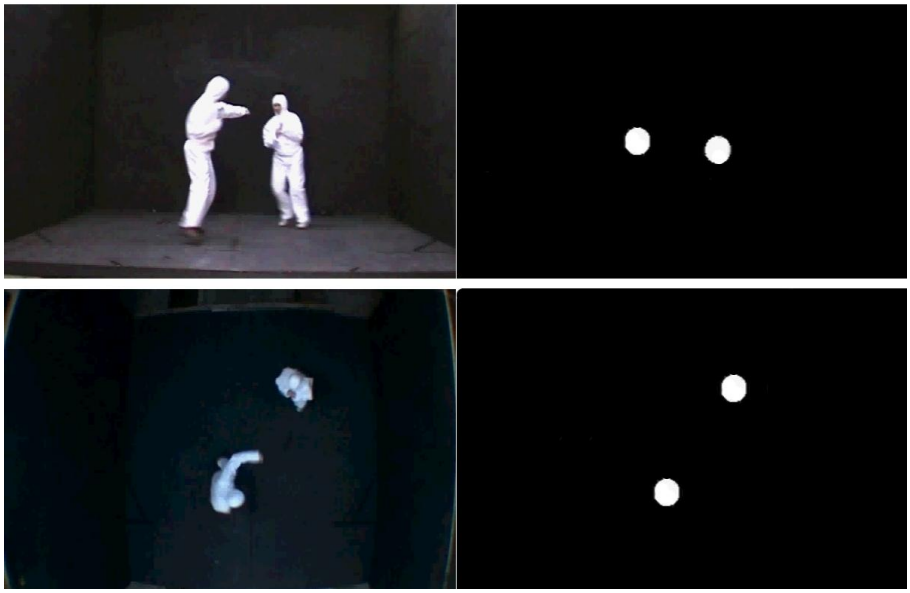

Figure 3.4.1: Boxing1 display shown in its original format (left) and the equivalent animacy displays (right), at both viewpoints: side view (top) and overhead (bottom).

The actors were filmed performing 16 new scenarios involving a wide range of actions. Of the 16 scenarios, 15 were used to create the experimental displays[9]. Within this set of 15 displays, 3 involved one actor, and the remaining 12 showed dyadic interactions. The scenarios were scripted as follows, starting with the three monadic displays: **(1) Walk:** one actor walks from left of stage to right of stage; **(2) WalkLR:** one actor walks from left of stage to right of stage, turns and walks back; **(3) Walk Jog:** one actor walks from left of stage to right of stage, turns and jogs back; **(4) Attraction1:** same as **Attraction** display in Experiment 3.3; **(5) Attraction2:** same as **Attraction1** but both actors walk off side by side instead of one following the other off; **(6) Boxing1:** actors start in diagonally opposite corners, walk out to meet each other in center then move round stage as though in a boxing match, occasionally throwing fake punches, after 30 seconds they return to their original corners; **(7) Boxing2:** same as **Boxing1**, except after 30 seconds, actor on left falls to ground and the remaining actor jumps up and down in celebration; **(8) Chase1:** first actor starts on centre of left wall, second actor starts at centre of back wall, both begin to jog in a clockwise direction, never getting any closer to each other, after two laps they stop in original positions; **(9) Chase2:** same as **Chase1**, except after one lap, second actor gets closer to first actor; by end of second lap the actor who is trailing gets level with first and they stop; **(10) Chase3:** same as **Chase2**, except when second actor is level with first actor, they continue to run side by side for one more lap of stage; **(11) Chase&Evade1:** same starting positions as **Chase1**, after 1 lap first actor stops moving in circles and turns to face second actor; for

[9] It was discovered that the 16th display (Chase&Evade2) was not suitably recorded to be able to extract the positional co-ordinates. This was due to problems with Eyesweb that are discussed in Chapter 2. These problems have now been rectified and the display still exists for future analysis.

next 30 seconds actor one tries to not get touched by second actor by moving anywhere on the stage; **(12) Chase&Evade3:** same as **Chase&Evade1** except after 30 seconds, actor two catches actor one and they run side by side for one lap of the stage; **(13) Clap:** same as **ClapInMiddle** from Experiment 3.3; **(14) SocInt1:** same as **SocialInteraction** from Experiment 3.3; **(15) SocInt2:** same as **SocInt1** except the crouched actor does not chase the other actor away, but instead stays in same spot whilst first actor runs away.

3.4.3 PROCEDURE

The experiment was run on a G4 Apple Macintosh (O.S. 9.2) using Matlab 5 (the Mathworks, Natick), Showtime (Watson and Hu, 1999) and the Psychophysics Toolbox Version 2.5 (Brainard, 1997; Pelli, 1997). Each experiment consisted of 90 trials split into 3 blocks of 30 trials, with 3 practice trials to acquaint the participants with the task. Each block showed all 15 displays from both viewpoints – overhead perspective and side view perspective. Using a 15(Displays) X 2(Viewpoints) within design, participants saw each display three times at both viewpoints. A within design was used in order to obtain numerous responses from each participant, for all display combinations, thus allowing for a reduction in number of participants required for statistical analysis. After each display participants were asked to give a rating of self-propulsion of the shapes in the display: a rating of 1 meant little or no self-propulsion and a rating of 9 meant the shape(s) appeared to be moving under free will.

Participants were sat at a distance of approximately 1m, giving a starting visual arc of each circle in both viewpoints of 0.5'. The luminance of the circles

in both viewpoints was 150 cd/m^2, and luminance of the black background was 0.28 cd/m^2. This gives a contrast between the white circles and the black background, in both viewpoints, of approximately 1 (Michelson, 1927). Displays were presented at 30fps.

3.4.4 RESULTS

This experiment was designed to examine the effect of viewpoint of the animacy display on the perception of self-propulsion of the agents in the display. It was hypothesised that there would be no difference in ratings of self-propulsion between viewpoints. A graph showing mean ratings of self-propulsion in the two viewpoints, overhead and side view, can be seen in Figure 3.4.2. A rating of 9 indicates that the shapes are viewed as being fully self-propelled and a rating of 1 indicates that they are viewed as being controlled by an unseen, external force.

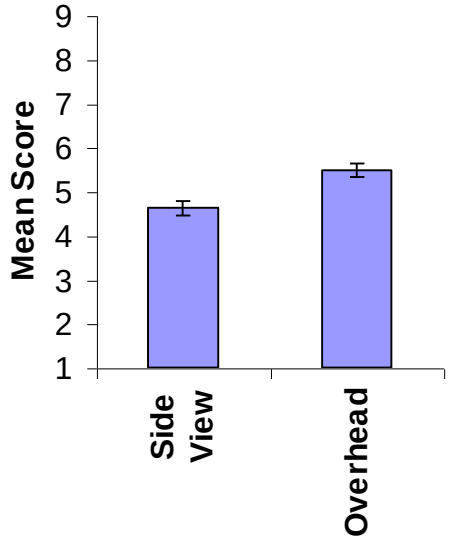

Figure 3.4.2: Average ratings of self-propulsion across viewpoints, showing standard error of the means. A rating of 9 indicates that participants perceived the agents to be free willed and a rating of 1 indicated that participants perceived the agents to be controlled by an unseen external force.

As can be seen in the graph there would appear to be a tendency to rate displays showing agents from the overhead perspective as appearing more self-propelled than displays showing agents from a side view. The mean ratings for the overhead perspective displays and the side view perspective displays are 5.4 and 4.58 respectively. A graph showing the mean ratings in both viewpoints for all displays can be seen in Figure 3.4.3. Again, a rating of 9 indicates that the shapes are viewed as being fully self-propelled and a rating of 1 indicates that they are viewed as being controlled by an unseen, external force.

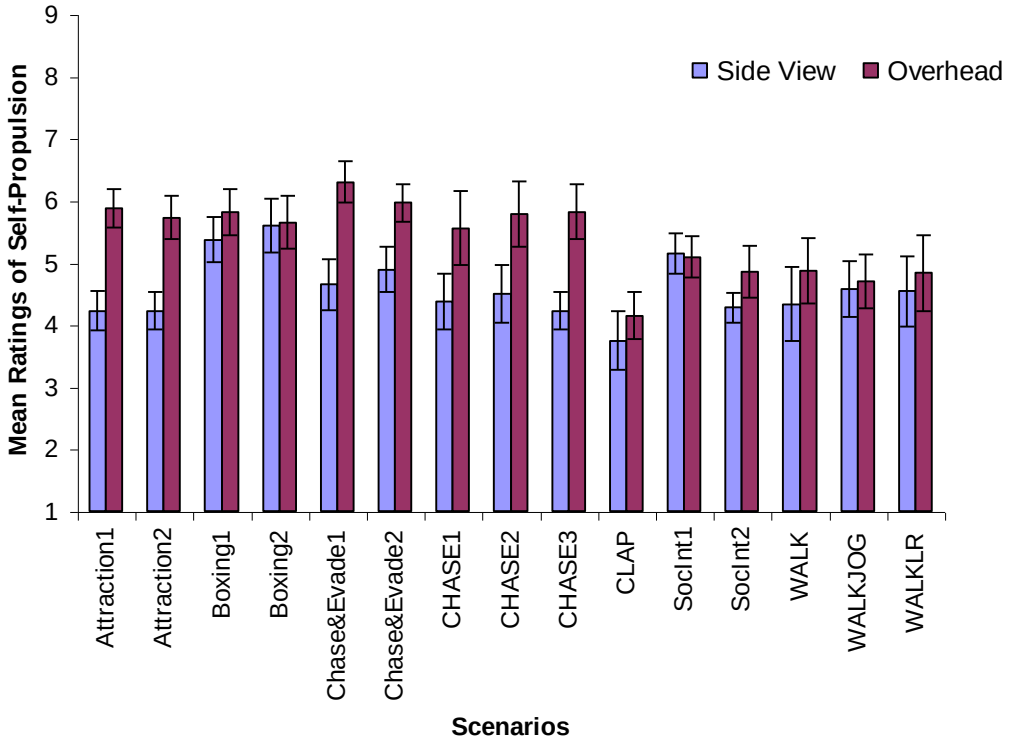

Figure 3.4.3: Average ratings of self-propulsion across scenarios and viewpoint. Error bars indicate standard error of the means. A rating of 9 indicates that participants perceived the agents to be free willed. A rating of 1 indicated that participants perceived the agents to be controlled by an unseen external force.

Figure 3.4.3 shows that there is variation between the displays in mean ratings of self-propulsion, however overall the graph suggests that for all but one

of the displays, SocInt1, it would appear that the overhead displays are rated as more self-propelled than the side view displays.

A two way ANOVA was performed – Viewpoint (overhead X side) Vs Display (Attraction1 X Attraction2 X Boxing1 X Boxing2 X Chase&Evade1 X Chase&Evade3 X Chase1 X Chase2 X Chase3 X Clap X SocInt1 X SocInt2 X Walk X WalkJog X WalkLR). Prior to this, the data were normalised using z-scores in order to compensate for the variances in the use of the scale by different participants. Results showed a main effect of Viewpoint, $F(1,18) = 15.7$, $\boldsymbol{p<0.01}$, with Fischer's LSD post-hoc analysis revealing that displays shown from the Overhead are rated significantly higher for self-propulsion than displays show from the Side view: mean ratings for Overhead Vs Side View are 5.4 Vs 4.6, respectively. No main effect of Display was found. Finally, an interaction between Viewpoint and Display was seen, $F(14,238) = 2.62$, $\boldsymbol{p<0.01}$. As in the previous experiment, using Tukey's HSD post-hoc analysis, multiple significant differences were found and are summarised in Table 3.4.1 (page 98), with the mean ratings of all scenarios at both viewpoints summarised in Table 3.4.2 (page 97):

	Side View	Overhead
ATT1	4.22	5.87
ATT2	4.22	5.72
BOX1	5.37	5.81
BOX2	5.59	5.65
CE1	4.65	6.30
CE3	4.89	5.96
CHASE1	4.37	5.56
CHASE2	4.50	5.78
CHASE3	4.22	5.81
CLAP	3.74	4.15
INT1	5.15	5.09
INT2	4.28	4.85
WALK	4.33	4.87
WALKJOG	4.57	4.70
WALKLR	4.54	4.83

Table 3.4.2: Summary of the mean ratings of self-propulsion for each scenario (rows) at both viewpoints (columns).

		Side View															Overhead														
		ATT1	ATT2	BOX1	BOX2	CE1	CE3	CHASE1	CHASE2	CHASE3	CLAP	INT1	INT2	WALK	WALKJ	WALKLR	ATT1	ATT2	BOX1	BOX2	CE1	CE3	CHASE1	CHASE2	CHASE3	CLAP	INT1	INT2	WALK	WALKJ	WALKLR
Side View	ATT1																													
	ATT2																													
	BOX1										+																			
	BOX2	+	+					+		+	+		+	+																
	CE1																													
	CE3																													
	CHASE1																													
	CHASE2																													
	CHASE3																													
	CLAP																													
	INT1										+																			
	INT2																													
	WALK																													
	WALKJ																													
	WALKLR																													
Overhead	ATT1	+	+			+		+	+	+	+		+	+	+	+										+				
	ATT2	+	+					+	+	+	+		+	+												+				
	BOX1	+	+					+	+	+	+		+	+	+	+										+				
	BOX2	+	+					+		+	+		+	+												+				
	CE1	+	+			+	+	+	+	+	+		+	+	+	+										+				
	CE3	+	+			+		+	+	+	+		+	+	+	+										+				
	CHASE1	+	+							+	+		+	+												+				
	CHASE2	+	+					+	+	+	+		+	+		+										+				
	CHASE3	+	+					+	+	+	+		+	+	+	+										+				
	CLAP			-	-																									
	INT1	+	+								+		+																	
	INT2																													
	WALK																													
	WALKJ																													
	WALKLR																													

Table 3.4.1: Summary of all significant differences as found using Tukey's HSD. A (+) indicates that the display in the row was rated significantly higher than the display in the column, whereas a (-) indicates that the display in the row was rated significantly lower than the display in the column. Abbreviations for displays are as follows: ATT1, ATT2 = Attraction 1 and 2; BOX1, BOX2 = Boxing 1 and 2; CE1, CE3 = Chase&Evade 1 and 3; INT1, INT2 = SocInt 1 and 2; WALKJ = WalkJog.

In summary, the majority of side view displays, except Boxing 1 & 2 and SocInt 1, were rated significantly lower that at least one or more of the dyadic overhead displays. The overhead display Clap was also rated significantly lower than the majority of dyadic overhead displays, and by Boxing 1 and 2 in the side view. None of the monadic displays in either viewpoint were rated significantly higher than any display.

3.4.5 DISCUSSION

We proposed to investigate the relevance of the viewpoint that animacy displays are viewed from. Little or no literature prior to this experiment has covered this topic, either due to not seeing the relevance or not having the facilities. This new method for the production of animacy displays allows for a direct comparison of a display from two viewpoints. We hypothesised that across 15 displays there would be no difference in ratings of self-propulsion between two viewpoints – an overhead perspective and a side-view perspective. The results however have shown a significant difference between these two vantages.

The results show that shapes displayed moving around from an overhead perspective appear to be more in control of themselves than the same shapes moving in the same motion pattern but shown from a side view perspective. This is a surprising result given that we are comparing our natural viewpoint, side view, against an unnatural viewpoint, the overhead, and finding that people rate displays shown from overhead as being more animate than those from the side. It is however possible that information provided uniquely by the overhead

displays cue these higher ratings of self-propulsion, and it would be interesting to discover what information is present in the overhead and not in the side view.

In side view displays, particularly displays in which the actors swap sides, it is maybe hard to track which shape represents which actor and this confusion could stop people from viewing them as individual agents with purposeful motion and instead lead to a percept of just random shapes bouncing off each other. This is probably most evident in the Clap displays where the actors walk in straight lines, meet and interact, and leave in straight lines. This on the side view may appear to be two shapes bouncing off each other in the middle, and as the overhead display gives no more information to the contrary, this could explain the low ratings of self-propulsion for this display in both viewpoints. On the other hand, in displays where the overhead display offers more information as to the distance between the agents or more clarity as to which agent is which, then we may expect these displays to be rated higher in self-propulsion, in comparison to the equivalent side view, as seen in the Chase displays. From this, we are beginning to get a realisation that it may be the removal of relative cues, i.e. the relationship between two agents that explains lower ratings of self-propulsion. Furthermore, a lack of relative cues could partly explain the low ratings of self-propulsion for monadic displays in both viewpoints. This would be consistent with the work of Blythe, Todd and Miller (1999) that showed that judging the intention of a computer-generated ant is reduced when the second ant is removed: thus removing the relative cues and leaving only absolute cues, i.e. the relation between an agent and its background.

In relation to the previous experiments, the overall ratings of self-propulsion and the individual ratings for the side view displays do appear to be

slightly lower in this experiment. This is most likely due to the inclusion of the overhead displays and participants using these displays as a benchmark for rating the side view displays. Furthermore, the poor use of animate terms in the previous three experiments of this chapter could be related to the finding of low ratings of self-propulsion in side view displays. In short, if people do not perceive the agents to be moving under their own free will, they are unlikely to perceive the display as animate and will not describe the displays as such.

In summary, this experiment introduces an interesting and as yet un-reported aspect of animacy displays – it does matter what viewpoint an animacy display is to be perceived from. It has been shown that agents viewed from an overhead viewpoint get higher ratings of self-propulsion than agents viewed from a side view. The reason for this is unclear though most likely related to cues that are visible in the overhead displays but not in the side view displays, such as distal relation of one agent to another, or clarity of which agent is which. The differences in the perception of animacy displays based on viewpoint will form the basis of much of the research carried out in the subsequent experimental chapters.

3.5 Chapter Discussion and Summary

This chapter serves as an explorative introduction to a new technique for the production of animacy stimuli that are truly representative of human motion. The perception of animacy is the phenomenon of viewing moving geometric shapes and attributing human-like qualities to the motion, such as beliefs, desires, feelings, etc. Animacy was originally reported by Heider and Simmel (1944)

using a display created by a key-frame technique, or stopgap animation. Hitherto, methods for creating these displays have ranged from complex algorithms (Zacks, 2004) to clever computer game scenarios (Blythe, Todd and Miller, 1999), with parametric variations of the original Heider and Simmel display in the middle (Berry, Misovich, Kean and Baron, 1992). The results of these experiments have always been analysed in relation to what they tell us about the attribution of social intention in humans, however no display has been made that is a true representation of human motion. In this chapter we presented a technique that takes recorded motion of human actors, and through a series of background subtraction and video processing stages, outputs displays where the motion of these actors is now represented by moving shapes. This technique offers a secondary advantage that it allows for the same interaction/action to be created simultaneously from differing viewpoints. We made use of this advantage to introduce a new line in animacy research, that of the importance of viewpoint from which the display is perceived. To this point, the vast majority of animacy literature makes use of displays that are to be viewed from an overhead viewpoint, as though we are looking down onto the moving shapes (Bassili, 1976; Berry et al, 1992; Bloom and Veres, 1999; Dittrich and Lea, 1994; Gelman, Durgin and Kaufman, 1995; Heider and Simmel, 1944; Stewart, 1982; Tremoulet and Feldman 2006; 2000). The use of animacy displays showing a side-on view, one that we would be more akin to human perception, are generally only found in child development literature (Csibra, Gergely, Biro, Koos and Brockbank, 1999; Gergely, Nadasdy, Csibra and Biro, 1995). Research carried out in this chapter made direct comparisons of displays viewed from both viewpoints to examine the effect of the viewpoint in animacy displays.

In the Experiments 3.1 and 3.2 we tested this new technique for the production of animacy displays using recordings of modern dance. The displays of the first experiment showed two men in the video recordings and two red shapes in the animacy displays. We incorporated an online slider task to measure changing levels of arousal as the visual information became more reduced and we used a free response task to examine if participants would describe the displays in animate terms. The results showed that only when the displays clearly showed actors interacting, i.e., the video recordings and the silhouette displays did the participants describe the displays in animate terms. The displays where the motion was shown as two red blocks moving around a sparse black environment were never described in animate terms, instead being seen as random motion. The online slider task showed that the silhouette displays created greater levels of arousal in participants than any of the other displays. When we compared this to the free responses we found that often the silhouettes would be described as a fight or some form of martial art, whereas the original video footage would be described as a dance and the animacy displays, i.e. the pulsing block and block conditions, were described as shapes moving in random motion. That the silhouettes are viewed as a fight would explain the increased levels of arousal for this display condition, and also shows an interesting effect of a simple background reduction and removal of context on the perception of an event.

The second experiment used a solo dancer as the original footage as it was thought that the complex nature of the interaction of the dyadic dance may have created a confusing display and thus prevented participants from perceiving the shapes as animate. Tremoulet and Feldman (2000) showed that a single

moving object would be rated as alive when it undergoes a change in speed and/or direction but without a visible cause for this change. The resultant animacy displays of the solo dancer footage should therefore have been viewed as animate, as the dancer was continually changing direction and speed, including stopping and starting. The cue of self-propulsion has previously shown to be indicative of animacy, but not essential (Csibra et al, 1999; Gergely et al, 1995). Experiment 3.2 also included an online slider task to measure emotional engagement of the participant, as well as the free response task. The results again showed that only the display conditions where the dancer is clearly seen were viewed as animate, and the conditions showing a moving geometric shape were not. There was no effect of emotional engagement across displays.

It was expected that the animacy displays of these Experiments 3.1 and 3.2 would be described in animate terms as they portray human motion degraded to show just the global motion. Previous animacy literature has shown that the spatial and temporal dynamics of a display, as well as influences of top-down prior knowledge, create the perception of animacy, and not the aesthetic qualities of the displays. It is therefore surprising that these displays were not viewed as animate as they show shapes moving with spatial and temporal properties consistent with what previous animacy literature would suggest would create the perception of animacy. The results from the first two experiments of this chapter raise the issue that perhaps not every human motion can be degraded into its global properties to create animacy displays, and that perception of animacy can only be created using specific motions depicting certain actions. Re-examining the motion of the dyadic dance footage and the monadic dance footage, it is evident that the action follows no social schema and both displays depict

complex movements that are perhaps only meaningful to those that are trained in modern dance or are avid viewers of such actions. It would be of interest to repeat these experiments to obtain the free responses of experts who may have some schema of how dancers move in modern dance routines, and compare these to the free responses of our novices. That a social schema is required to understand an event has been suggested in the literature of event understanding (Zacks and Tversky, 2001) and a lack of social schema for modern dance routine is suggestible as the cause of the lack of perception of animacy thus far. The subsequent experiments in the chapter looked at social interactions and actions that would be more common in general life to people, such as running, chasing, flirting, walking, etc.

In Experiment 3.3 we made use of such everyday actions and again looked for free responses to the animacy displays derived from these actions. We also used a post-rating of self-propulsion, similar to that of Tremoulet and Feldman (2000), as it was thought that the online task might stop people correctly perceiving the motion in the displays as they are concentrating on the slider task. Six new displays were created and were tested along with the displays from the previous experiments and an adaptation of the Heider and Simmel display (Nevarez and Scholl, 2000). The free response data showed that at least one of our displays, the Social Interaction display, was described in animate terms at an occurrence level similar to that of the Heider and Simmel display. This Social Interaction display depicted an actor/block crouched in the corner and being approached by a second actor/block; as the second actor/block gets close to the first actor/block, the first actor/block jumps up and chases the other agent away. Why this display should be perceived as animate motion so

often is unclear but comparing it to the other displays we may begin to get some understanding. Of all the displays used, this display depicts a story; it shows an injured/pretending-to-be-injured person being approached by another and then a chase ensues. Compared to the other displays of walking and running, this is the one that is most rich in coherent actions; that the person who was crouched does not want to be approached. Other displays that are described in animate terms are the attraction display, and the dyadic dance display that had previously never been described in animate terms in Experiment 3.1. Both these displays also depict some form of coherent interaction between two shapes but the purpose is not as clear as in the Social Interaction display. It is also worth noting that the actors never cross over in the Social Interaction display, but do so in the other two mentioned displays, and perhaps the clarity of which actor is which helps facilitate the perception of animacy and the use of animate terms. There is a slight contradiction in this theory as the clarity of agents in the Clap in Middle and Quick Chase displays, in terms of linear motion of the agents is very similar to the Social Interaction, and therefore one might have expected more animate terms being used to describe these two displays. However, it is possible that the difference between the Social Interaction display and the Clap in Middle and Quick Chase displays, is the underlying stories: the basis of Social Interaction may be easier to make a coherent story out of compared to the other two displays, irrespective of the fact that in all displays it is easy to determine which agent is which.

Nearly all of the displays used in Experiment 3.3 were described by at least one person in animate terms. Though this is not a great indicator that this technique for the production of animacy displays from recordings of human

motion always produces animacy displays, it does show that the basis of the technique is valid. Of further interest, displays that were not often described in animate terms sometimes received ratings of self-propulsion that indicated that they had been perceived to be in control of their own movement. This may point to the inadequacies of the free response task for these experiments, a point backed up by post-experimental interviews with participants that indicated they felt self-conscious about writing responses that gave the shapes human qualities, even though this was how they had perceived them. Also, the removal of the online slider task has had an effect on the dyadic dance display from Experiment 3.1. In Experiment 3.3, where the online rating scale had been removed, the dyadic dance display was now described in animate terms, thus pointing to the significance of constant and undivided attention to animacy displays.

Finally, Experiment 3.3 showed a difference in the use of animate terms when comparing displays where the shapes changed in size in accordance with the Quantity of Motion, and where the shape size was rigid. The results showed that the use of animate terms appeared to be more likely used to describe the displays, when the shapes in the displays changed size, and that this is consistent across scenarios. Non-rigid surface structure has been shown to help the perception of animacy by Gergely et al (1995) and Csibra et al (1999), however in the displays of Experiment 3.3, the perception that the changing surface size gives is unclear. It may be that it does show reciprocal changes in size and is indicative of a free willed agent, or it may be that it gives the false perception of depth, thus the shape may appear to be coming towards the observer when it gets larger and may appear to be moving away from the observer when it gets smaller: a false suggestion of depth between the two agents may influence the

perception of animacy by suggesting movement in depth when the agents were actually moving on a linear plane. In the research of Gergely et al and Csibra et al, there is a clearly indicated horizon and thus there can only be the perception of reciprocal changes in structure shape and size. In the context of this thesis, it was decided that further research into this cue is required before continuing to use it as a representation of the Quantity of Motion to ascertain if participants are susceptible to this cue or if they are perceiving the changes in size as a cue to motion in depth.

The final experiment of this chapter, Experiment 3.4, was created to introduce the topic of the relevance of viewpoint in animacy displays, a topic that has thus far received little attention. Prior to this experiment all displays in this chapter, except the Heider and Simmel adaptation, have been recorded and displayed from a side viewpoint. However the majority of animacy research has used displays depicting the overhead viewpoint. This experiment compared the effect of viewpoint on the perception of numerous displays, all of a similar ilk to the previous experiment, i.e. of common occurrence in general life. The free response task was removed and only the post-rating of self-propulsion task was administered. Results showed a consistent preference for displays depicting the motion from the overhead viewpoint. This is a somewhat surprising finding and it is unclear as to why there should be a preference for displays showing an unnatural viewpoint, the overhead, over displays showing our natural viewpoint, the side view.

Closer inspection of the displays reveal differences that may answer why people rate the overhead displays as more animate. The overhead displays show the motion on a two-dimensional plane as opposed to the linear, one-dimensional

plane depicted in the side view. This allows for clarity of identity of agents in the overhead as well as allowing the perception of depth between the agents. The side view can only show the shapes moving left to right and as they are both of the same colour, it is maybe difficult to follow the motion of each agent coherently, especially if the agents cross over and change lateral positions. In the overhead view there is no mix up between the agents when they cross over and therefore it is simpler to track the motion of each agent. Future experiments in this thesis will examine the importance of clarifying which agent is which. Though aesthetic qualities, such as colour of shapes, are generally considered to be unimportant, it is perhaps the case that in side view displays, clarity of identity of agent may facilitate the perception of animacy. Furthermore, in the overhead displays there is a clearer representation of the cues such as the distance between agents and the angle between agents, on top of the cues such as speed and acceleration that are evident in the side view. All these cues have previously been shown to be effective for the attribution of intention (Blythe et al, 1999; Barrett, Todd, Miller and Blythe, 2005) and in event segmentation (Zacks, 2004).

Findings that indicate that displays appear less animate when shown from a side view may point to a reason for the poor free responses that were seen in the first three experiments of this chapter. If the agents are not being perceived as being under their own control, as is suggested by the results of this last experiment, then it is in turn evident that observers would not describe them in terms suggesting such. In other words, it is highly probable that participants have been using animate terms sparsely in the previous experiments to describe the displays, simply because the displays do not appear to be of animate motion. Visual information has ultimately been lost, predominantly in the side view, in

the transition from footage of human actors to the animacy displays, which would indicate that the motion should be perceived as purposeful human motion. Subsequent research in this thesis will continue to examine differences in the perception of animacy displays as a result of viewpoint, and will try to restore cues that will increase the perception of animacy in displays shown from the side view.

In conclusion, this chapter introduces a new method for the production of animacy displays that are true representations of human motion and has subsequently validated this technique by showing that people, on viewing these displays, do perceive these motions to be animate. It has been shown that displays depicting more common interactions such as flirting and chasing, and actions such as walking, when converted into animacy displays are more readily seen as animate than displays involving uncommon complex motions, such as modern dance. It is thought that one reason for this difference maybe due to the familiarity with events and prior knowledge of social schemas. Finally this chapter introduces the topic of viewpoint in animacy displays and shows the surprising result of an advantage for displays depicting the motion from an overhead perspective. The research of this chapter serves as the basis and springboard of the subsequent experimental chapters in which the technique for the production of animacy displays will be further developed, and the question of viewpoint will be examined more thoroughly but via a narrower set of intentions and actions.

Chapter 4 – Animacy from Specific Human Intentions

Chapter Introduction and Overview

The previous chapter focused on developing and validating the technique of tracking positional co-ordinates of actors and using these co-ordinates to create animacy displays that were representative of human motion. In doing so emphasis of the work slipped from researching cues that facilitate or create the perception of animacy and the use of numerous scenarios may have been the cause of this. In this chapter we intended to focus the emphasis onto a specific set of intentions, which are fewer in number but more generic in nature, in an attempt to advance the research into cues that lead to a perception of animacy. We will continue to use the technique for the creation of animacy displays, developed in the previous chapter, to study judgements of intentions from animacy displays and to discuss the cues that define them, whilst continuing to use viewpoint as a variable.

In the animacy literature there are few studies that use displays that focus mainly on specific intentions, instead favouring to use displays involving a variety of intentional movements. Research that does single out intentions is that of Blythe, Todd & Miller (1999). Their aim was to discover the cues that are used to distinguish between action categories, and to investigate cognitive algorithms that may dynamically enhance this process. Following extensive analysis of anthropological, causality and animacy literature, they focused their research on a small group of intentions that they deemed to be the most typical forms of intention driven interactions between two agents. They looked at the intentions of Chasing, Evading, Courting, Be Courted, Fighting and Playing. All of the intentions examined had reproductive and survival qualities, and based on the importance of these intentions to survival, the authors made the assumptions

that people would be able to perceive these intentions in animacy displays and be able to infer the appropriate intention.

To create animacy displays depicting these intentions, Blythe et al (1999) created a computer game scenario where pairs of participants would control animated ants on a computer monitor. Participants were placed in separate rooms with individual monitors on which they could see both ants but controlled only one. Each participant controlled a different ant and was instructed to move their ant in accordance with the intentions specified by the authors. The programme used to create the ants was fitted with algorithms that would satisfy acceleration and deceleration constraints, and a third participant would judge the displays to determine if they did indeed show the pre-determined intention. The trajectories of the ants were recorded for subsequent analysis, and subjective impressions of the trajectories showed differences in various aspects such as speed changes, turns, pauses, etc. A new group of participants was shown the displays and asked to categorize the displays using a six alternative forced choice. An approximate overall success rate of 50% was found, with chance being 16.67%. Hit rates and false alarm rates did vary with the most notable being a high false alarm rate for Play, which the authors suggested was reliant on play being a pre-cursor of all other intentions, and also an underlying belief by participants that play is a more common intention in life. Blythe et al further showed that removing one of the ants reduced hit rates to approximately 30%, suggesting that the attribution of intention in these displays was heavily reliant on the relative properties between the two bugs, e.g. the distal relationship between the two bugs. The authors looked at 7 cues that they deemed useful for the categorization of these intentions: relative distance, relative angle, relative

heading, relative velocity, relative vorticity, absolute velocity and absolute vorticity – with vorticity being defined by the authors as changes in heading, i.e. turning to face different directions. A neural network was trained to check the validity of these cues for the categorisation of intent and a success rate of 82% was achieved. Therefore, these cues do not capture all the information but do capture enough information to have a hit rate significantly higher than human participants. The authors used these cues as the make-up of an algorithm called Classification by Elimination (CBE). CBE is a simple, fast and frugal algorithm that only uses as many cues as it needs to make classification, be it one or all seven. The downside of the algorithm is that it is reliant on the order of the cues entered into the algorithm, and therefore, to be most efficient, the correct order of cues, as regards explaining the variance, must be established prior to using the algorithm. Blythe et al established that absolute velocity was the best classifier followed by relative angle, relative velocity, relative heading, relative vorticity, absolute vorticity and relative distance. Hence this was the order that the cues were entered into the algorithm. CBE, using only half the cues required by other algorithms, was shown to perform equivalent to a more traditional neural network, showing that these cues do indeed have strong diagnostic qualities for classifying these intentions.

In a follow up paper, Barrett, Todd, Miller and Blythe (2006) investigated whether the ability to investigate these intentions was culturally specific or indeed a universal trait of people. They ran a cross-cultural study comparing participants from the hunter-horticulturalist society of the Shuar adult population from Amazonian Ecuador, with a modern German adult population. The authors used four of the intentions from the previous study and two new intentions of

survival importance, giving the six intentions of Chasing, Courting, Fighting, Following, Guarding and Playing. They employed the same computer game scenario to create the displays (Blythe, Todd and Miller 1999), however this time arrowheads were used instead of ants. Results showed no overall significant differences between cultures, both achieving an overall hit rate of around 80%. This hit rate was much higher compared to the original study and it is suggested by the Barrett et al that the lack of specific instructions given to the participants making the displays in this experiment, as opposed to the specific instructions given in the previous experiment, may account for this difference. There was no overall difference in recognition between intentions across cultures. From these results, Barrett, Todd, Miller and Blythe (2005) concluded that these six intentions can be perceived across diverse cultures and are universally important for survival and reproduction.

The six intentions of Barrett et al (2005) will form the basis of the experiments and analysis in the following chapters. This chapter will focus on using these intentions to further study the perception of animacy in displays derived from human actions. Where Blythe et al (1999) and Barrett et al (2005) used computer-generated ants in a game scenario, we will use actors on a stage, performing the intentions of Chasing, Courting, Fighting, Following, Guarding and Playing. From there we will use the positional co-ordinates of the recorded actors to create animacy displays of each of the intentions. We will record the actions from the two viewpoints established in the previous chapter, overhead and side-view, to see if a similar pattern of results is achieved as in Experiment 3.4, where it was shown that displays showing agents from the overhead viewpoint appeared more self-propelled than agents in the side-view displays.

Furthermore, we will compare overall success rates for judging intentions in animacy displays derived from human motion and contrasts these to the hit rates achieved via more conventional techniques of animacy display creation by Blythe et al (1999) and Barrett et al (2005).

Experiment 4.1 – Judging Intentions: A Pen and Paper Task: Pilot

We initially ran a pilot experiment to examine the ability of observers to judge intention in animacy displays created from human motion, using the six intention categories as described by Barrett et al (2005). We hypothesized that participants would be able to correctly categorize intentions at levels above chance for displays showing the overhead view and the side view. Experiment 3.4 in the previous chapter showed that displays depicting agents from an overhead viewpoint tended to be rated as being more self-propelled than agents in displays showing the side view. We therefore hypothesized that differences in perception of animacy displays across viewpoint may affect the ability to judge the intention depicted in the side view displays, and therefore people would be better at categorizing displays showing an overhead viewpoint, though this is not our natural viewpoint.

METHODS

4.1.1 PARTICIPANTS

31 participants from the New York University[3] undergraduate participant pool took part in the experiment. All had normal or corrected to normal vision, were naïve to the purpose of the experiment, and received course credit for taking part in the experiment.

4.1.2 STIMULUS PRODUCTION:

2 actors (1 male) were filmed using the same stage and equipment set up as discussed in Experiment 3.4. The 2 actors were instructed to act out the six following intentions, with the given instructions: **(1) Chasing:** the two actors started a foot apart against the back wall of the stage, both facing in the same direction, they then jogged at a ¾ pace in a clockwise direction doing laps of the stage but never caught each other, every couple of laps the actors ran in a figure of eight across the stage. The actors stopped after 30 seconds; **(2) Courting[4]/Flirting:** both actors stand with bodies facing towards the side camera about five feet apart; actor on the right moved to the actor on left and circled them twice then moved back to their original starting point; second actor moved over to first actor and repeated the circling motion; after second circle, both actors moved off together to top left of stage; **(3) Fighting:** actors started at opposite sides of the stage facing inwards towards each other; they walked towards each other and stopped about a foot apart; they both walked in a clockwise circle facing each other as though measuring each other up for a

[3] This experiment was run whilst on a trip in the summer of 2005 to visit the lab of Dr. Kerri Johnson at New York University, New York, New York, U.S.A.

[4] Though in the original papers of Blythe, Todd & Miller (1999) and Barrett, Todd, Miller and Blythe (2005) this intention was referred to as courting, it was decided, after discussion with Kerri Johnson, NYU, NY, that a more appropriate word for it would be flirting.

battle/fight; after a couple of complete circles actor on right took an exaggerated and aggressive step towards the other actor, who took an exaggerated jump backwards; the actor who jumped back rushed towards the other actor and took an exaggerated and aggressive step towards the other actor who jumped back; action is repeated for 30 seconds upon which actor on right fell back and down to the ground; **(4) Following:** same instructions as **Chasing** except performed at a walking pace; **(5) Guarding:** one actor started on far left and one started on far right; actor on far right had an imaginary ball behind them[5]; actor on left tried to get past actor on right to get to ball; actor trying to get ball never achieved this goal; **(6) Playing:** actors start in middle of stage facing towards each other, two feet apart; actor on right was instructed to try to touch the actor on left, and a game of tag ensued; this lasted for about 30 seconds.

In the experiments of Chapter 3, the experimental displays were created using the positional co-ordinates of the actors, extracted from Eyesweb. However these displays, particularly the side view displays, sometimes contained artefacts from the merging of the positional co-ordinates of the two actors if they got too close together, resulting in a "jump" between shapes that did not exist in the original video recording. For this reason it was decided to filter the positional co-ordinates to remove any "jumps". Prior to filtering it was necessary to manually check and adjust the co-ordinates of the actors that had become switched due to the tracking problem of Eyesweb discussed in Chapter 2, whereby the tracking system in Eyesweb jumped between actors depending on which side of the screen they were in. This problem was readily solved due to switched regions of co-ordinates being identified by large changes in sequential

[5] A couple of practice attempts with a real ball were used to give actors better idea of what they were performing.

co-ordinates. Furthermore, it was decided to filter the co-ordinates as this would eliminate the vertical component of gait motion from the side view so both overhead and side views were equivalent. The co-ordinates were filtered in Matlab using a 4th order Butterworth low pass filter with a cut-off frequency of 0.8Hz. A separate Matlab code was used to create the final experimental movies, outputting the displays as QuickTime Movies with a frame rate of 30 fps, depicting white circles on black backgrounds.

4.1.3 PROCEDURE

Displays were shown to students during 4 drop-in sessions where the number of students present in each session ranged from 3 to 11. The movies were run on an Apple Macintosh Powerbook and displayed on a large screen using a projection system. In each session all intentions were shown once at both viewpoints. Using a 2(Viewpoint) X 6(Intention) design, participants saw a total of 12 displays. Two random orders of display presentations were created prior to the drop-in sessions and were randomly allocated to the four sessions. Participants were administered a pen and paper version of a six alternative forced choice[6]. One display would be shown and participants would be asked to circle the intention that best described the intention in the display.

4.1.4 RESULTS

The purpose of the experiment was to explore the perception of animacy displays derived from human motion using six specific intentions: chasing, fighting, flirting, following, guarding and playing. We examined these animacy

[6] No ratings of self-propulsions were obtained in this experiment or in any following experiment, as the decision over whether an agent is self-propelled or not was removed by forcing participants to choose a human intention to attribute to the agents.

displays from the two viewpoints of overhead and side view that were established in the previous experimental chapter. We hypothesised that people would be able to differentiate between intentions at levels above chance, in both viewpoints. It was also hypothesised that as overhead displays had been shown to contain agents that appeared more self-propelled than those in side view displays, in Experiment 3.4, participants would be poorer at perceiving the side view displays as animate, and this in turn would reduce their ability to categorise the intentions correctly from this viewpoint. A plot of the results can be seen in Figure 4.1.1.

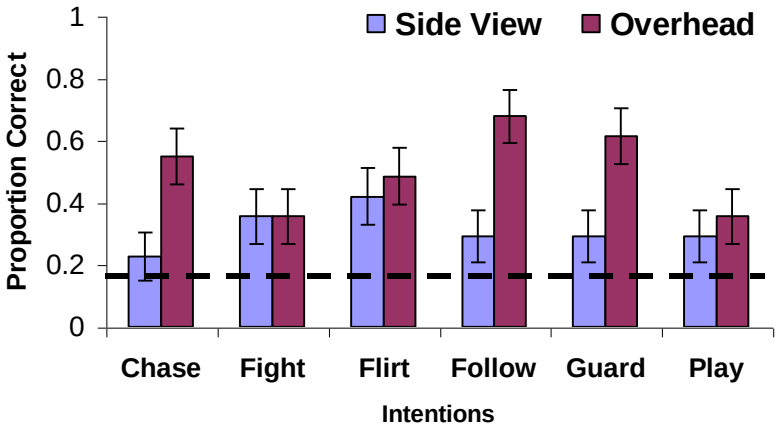

Figure 4.1.1: Proportion correct for observers' judgments of intentions in animacy displays for both Viewpoints, with error bars indicating standard error of means. Bold dash line indicates chance: 0.1667.

As can be seen from the graph, people were generally able to distinguish between the intentions at levels above chance for each intention, with chance being 16.67% for a six alternative forced choice. Furthermore, this was found to be the case for both the overhead perspective and the side view perspective. A two-way ANOVA was run – Viewpoint (Overhead, Side) Vs Intention (Chasing,

Fighting, Flirting, Following, Guarding, Playing) – and revealed a main effect of Viewpoint, $F(1,30) = 20.84$, $\mathbf{p<0.05}$ (50.5% overhead versus 31.2% side view). No main effect of intention was found, nor any interaction of intention by viewpoint.

4.1.5 DISCUSSION

This study was designed to replicate the findings of Blythe, Todd & Miller (1999) and of Barrett, Miller, Todd and Blythe (2005), justifying their findings that these are important intentions that people can identify from the global motion of geometric shapes. Our results have shown that people are adept at distinguishing between these intentions at levels above chance, irrespective of the viewpoint used to depict the interaction.

It would appear that for half of these intentions the ability to recognise the intentions is as good for displays depicting the overhead view as it is for displays depicting the side view, namely Fighting, Flirting and Playing. For the intentions of Chasing, Following and Guarding, people are better at distinguishing the intention from the overhead perspective, though this analysis is only subjective as no interaction was found. It is possible that with more participants or with repeated measures from participants, an interaction may be found.

The results reflect the pattern shown in Experiment 3.4 where people attributed more self-propulsion to displays depicting an overhead viewpoint. In the present experiment, people are better overall at categorizing intentions when the display shows the motion from an overhead viewpoint. This is something that has never been shown before in animacy displays and shows a preference for

animacy displays where the motion is viewed from above as opposed to side on, our natural viewpoint. The reason for this preference is unclear, though perhaps reduced depth information in the side view is responsible for the low performance. It is possible that a depth cue missing in the side views of Following, Chasing and Guarding is required for these intentions to be understood from the side view, but other cues missing in both viewpoints of Fighting, Flirting and Playing result in similar performances of judging intention for these intentions at both viewpoints.

Finally, the success rate of 50% for the overhead view is on a par with that reported by Blythe, Todd and Miller (1999), suggesting that this current method would appear to be an appropriate manner of creating the displays and that these intentions are appropriate categories of human actions to research. Barrett et al (2005) using less rigid instructions for the creation of their animacy displays, achieved higher hit rates than the present experiment and the Blythe et al paper, and it would be interesting to do further research into the role of the instructions for creating these displays.

In conclusion, this experiment has shown that this method is a successful method for creating animacy displays using video recordings of humans as a basis. This is backed up by similar results being achieved in this experiment as found in previous experiments by different authors. Furthermore, this experiment again showed a preference in participants for animacy displays depicting the overhead viewpoint, but that categorizing some intentions may still be independent of viewpoint. Finally, due to the small number of trials per participant this experiment was viewed as a pilot, though the results raise important points for future research.

Experiment 4.2 – Judging Intentions: Repeated Measures: Pilot

This experiment was designed to replicate the findings of Experiment 4.1, but using an experimental method that would allow for more judgements from participants. The study would incorporate a repeated measures design using a five alternative forced choice and presenting the displays in a more lab-based environment. Furthermore, we viewed this experiment as a second pilot to test the experimental method, and decided to use only five intentions: due to the similarity of Chasing and Following, only Following was used. We hypothesised that using a repeated measures design we would obtain results similar to Experiment 4.1, namely that overall people would be better at judging intentions from an overhead viewpoint.

METHODS

4.2.1 PARTICIPANTS

30 participants from the undergraduate population of the University of Glasgow took part in the experiment. All were naïve to the purpose of the experiment, had normal or corrected-to-normal vision, and received no incentive for their participation.

4.2.2 STIMULUS PRODUCTION

The same experimental stimuli from Experiment 4.1 were used.

4.2.3 PROCEDURE

The experiment was run on a G4 Apple Macintosh (O.S. 9.2) using Matlab 5 (the Mathworks, Natick), incorporating Showtime (Watson and Hu, 1999) and the Psychophysics Toolbox Version 2.5 (Brainard, 1997; Pelli, 1997). Each experiment consisted of 20 trials split into 2 blocks of 10 trials, with 3 practice trials to familiarize the participants with the task. Using a 5(Intention) X 2(Viewpoint) design, participants saw each intention twice at both viewpoints. After each display participants, using a five alternative forced choice, selected the intention that they thought had been portrayed in the display.

Participants were sat at a distance of approximately 1m, giving a starting visual arc of each circle in both viewpoints of 0.5'. The luminance of the white circles in both viewpoints was 147 cd/m^2, and luminance of the black background was 0.25 cd/m^2. This gives a contrast between the white circles and the black background, in both viewpoints, of approximately 1 (Michelson, 1927). Displays were presented at 30fps.

4.2.4 RESULTS

The aim of the experiment was to replicate Experiment 4.1 using a repeated measures design in a lab-based environment. The experiment was predominantly viewed as a pilot and therefore only five of the six intentions from the previous experiment were used. It was expected that similar results would be found in this experiment as in Experiment 4.1, which showed that overall

subjects were better at judging intention for displays depicting the overhead viewpoint. A graph of the results can be seen in Figure 4.2.1:

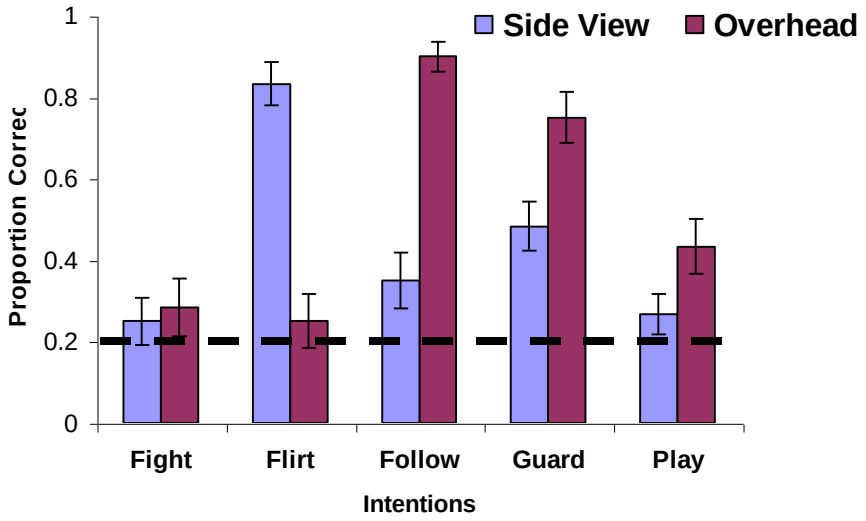

Figure 4.2.1: Proportion correct for observers' judgments of intentions in animacy displays for both Viewpoints, with error bars indicating standard error of means. Bold dash line indicates chance: 0.2.

The graph would suggest that people were able to categorize the intentions at levels above chance for each intention in both the overhead and side viewpoints, with a suggested slight advantage towards the Overhead, except for the Flirt intention where there appeared to be an advantage towards the side view displays. Chance is 20% for a five alternative forced choice.

A two-way ANOVA was run: Viewpoint (Overhead, Side) Vs Intention (Fighting, Flirting, Following, Guarding, Playing). No main effect of Viewpoint was found but there was a significant main effect of Intention, $F(4,116)=16.31$, $p<0.05$. Post-hoc analysis using Fischer's LSD revealed that the intentions of Following (62.5%), Guarding (61.7%) and Flirting (54.2%) are all significantly better categorized than the intentions of Fighting (26.7%) and Playing (35%).

An interaction, as suggested by the graph, was found between Viewpoint and Intention, $F(4,116) = 32.1$, $\mathbf{p<0.05}$. Tukey's HSD analysis showed that the side view of Flirting and the overhead views of Following and Guarding were all significantly better categorized than the side views of Fighting, Playing, Following and Guarding, and the overhead views of Fighting and Playing. The mean percentage of these categorizations are presented in Table 4.2.1:

	Side View	Overhead
Fight	0.25	0.28
Flirt	0.83	0.25
Follow	0.35	0.90
Guard	0.48	0.75
Play	0.27	0.43

Table 4.2.1: Mean proportion correct for categorizing in all Intentions at both Viewpoints.

4.2.5 DISCUSSION

The aim of this experiment was to run a brief pilot that would test a lab-based method of Experiment 4.1, incorporating a 5 alternative forced choice response on a computer, rather than using paper and pen. It was also hoped that similar results to Experiment 4.1 would validate the findings of both experiments. We hypothesised that people would be able to correctly categorize the intentions in both viewpoints at levels above chance and that overall, this ability would be better in the overhead view.

The procedure for the experiment was successful as participants could successfully categorize the displays at levels above chance. In contrast to the previous experiment, there was no main effect of Viewpoint, however a main effect of Intention was found as well as an interaction between Viewpoint and

Intention. Participants were particularly good at categorizing the intention of Flirting in the side view, and the intentions of Following and Guarding in the overhead view.

There is a major change in ability to judge the Flirt intention from the previous experiment to this experiment. In Experiment 4.1, the ability to perceive the Flirt intention appeared to be viewpoint independent, however in the present experiment there is a clear advantage for viewing it in the side view display. It is unclear why this result should be so markedly different from the previous experiment. It is possible that there was a distinct movement in the Flirting display in the side view that allowed people to perceive it easily, but in the overhead view, it may have looked slightly different and could have been perceived as a play or fight. It would also appear that the categorization of Fighting, though low, might be independent of the viewpoint of the display as there is little difference in ability to categorise this intention, across viewpoints. These are suggestions that will be looked at in the next experiment.

Of further note from this and the previous experiments involving viewpoint as a variable, individual debriefing sessions with participants after they had completed the experiments asked them to judge from which viewpoint they had perceived the displays. There appeared to be a striking bias towards perceiving the displays from the overhead perspective. It is possible that people were either not perceiving all the displays from the correct viewpoint, or that they are, but the memory of the overhead displays remains stronger. In the following experiment we will incorporate cues that increase the saliency of the viewpoint of the displays.

Experiment 4.3 – Judging Intentions with added cues to Occlusion and Context

We suggested in the previous experiments that cues to depth may be important for accurate categorization of intentions in the side view displays. This experiment is designed to address the question of the importance of cues to depth and to examine the effect of increasing the saliency of the viewpoint from which the displays should be perceived. In order to test the importance of depth cues we will produce displays where one circle is coloured black and the other circle is coloured white. This will give ordinal depth cues in the side view and will also provide a cue to identity in both viewpoints. Furthermore, we will use displays that show boundaries around the edges of the displays as a representation of a stage, thus providing a cue to viewpoint. It is hypothesised that these cues will increase the ability of participants to categorize the intentions in both viewpoints as it is thought that some of the low hit rates and confusions in judging the intentions in previous experiments, arises due to a lack of depth information in the side view displays, as well as a possible inability to perceive the correct viewpoint.

METHODS

4.3.1 PARTICIPANTS

17 participants from the University of Glasgow undergraduate population were used in this experiment. All were naïve to the purpose of the study, had normal or corrected to normal vision, and were paid for their participation.

4.3.2 STIMULUS PRODUCTION

The same filtered set of co-ordinates from Experiment 4.1 were used to create the experimental stimuli of this experiment, and were adapted to investigate the effect of increasing the saliency of viewpoint and of providing depth information in the Side View. Two cues were added - a contextual cue involving boundaries (C), and an occlusion cue (O) giving ordinal depth information in the side view displays and identity information in both viewpoints. This resulted in four experimental display conditions: (1) no occlusion; no context (NONC) - each actor represented by a white circle on a grey background; (2) no occlusion; context present (NOC) - same as condition 1, except each display has a boundary surrounding it to suggest viewpoint, overhead displays have four surrounding white lines, and side view displays have three surrounding lines: one beneath the circles and one on either side; (3) occlusion present; no context (ONC) - one actor depicted as a white circle and one actor depicted as a black circle on a grey background; (4) occlusion and context present (OC) - same as condition 3 with the boundaries from condition 2. An example of each experimental condition at both viewpoints can be seen in Figure 4.3.1 on page 129.

The filtered co-ordinates, along with the relevant contextual and occlusion cues were rendered as QuickTime Movies at a frame rate of 30 fps, mean duration 33ms.

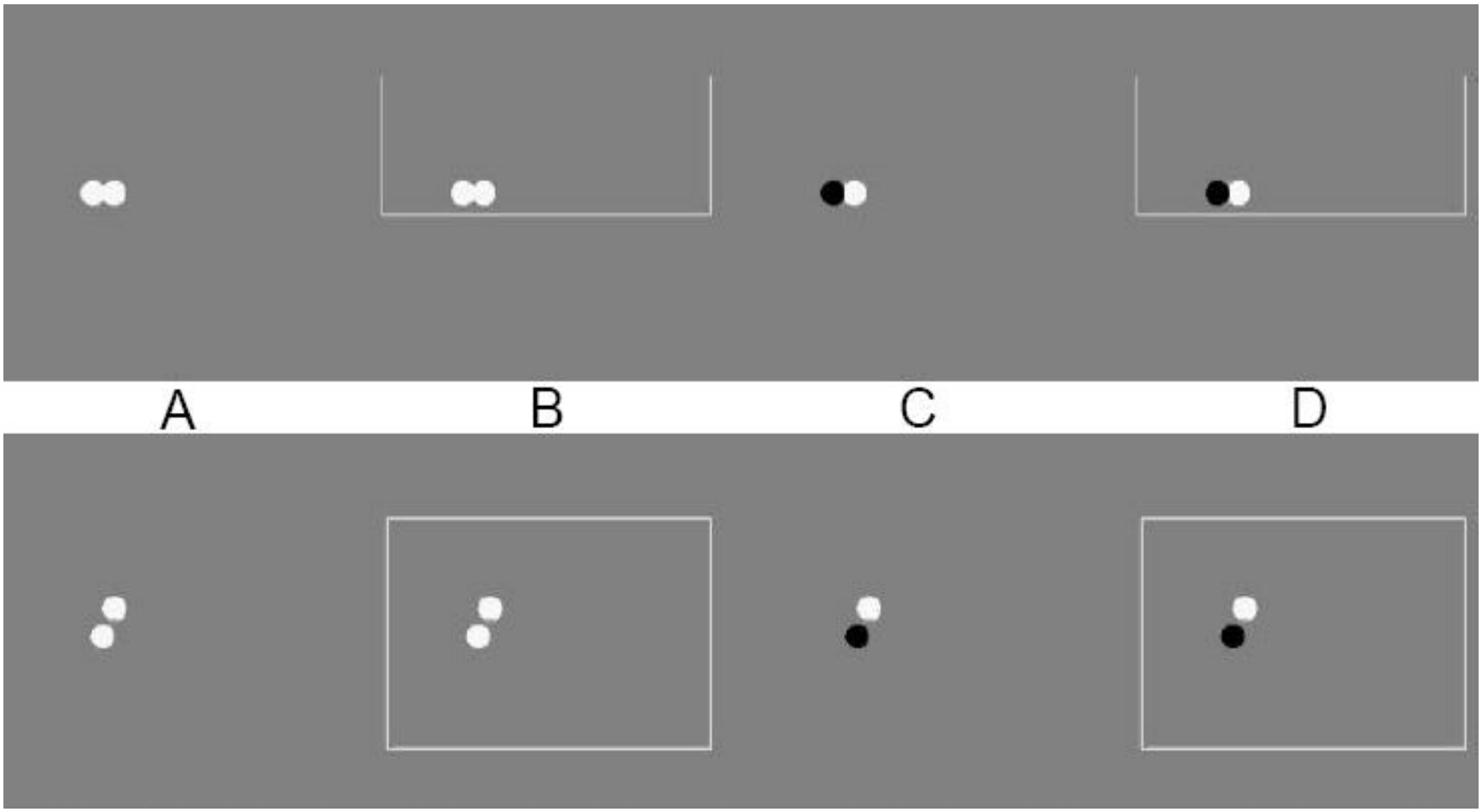

Figure 4.3.1: Four experimental conditions introducing additional visual cues. Top panel shows the side view displays, and the bottom panel shows the overhead displays. The Four conditions are: (A) - no occlusion, no context (NONC); (B) - no occlusion, context present (NOC); (C) - occlusion present, no context (ONC); (D) - occlusion and context present (OC). Though there can be no occlusion cue in the overhead displays, the colour of one ball black and one ball white does give identity to the balls, however this is still referred to as occlusion for clarity.

4.3.3 PROCEDURE

The same experimental apparatus was used as in Experiment 4.2. Each experiment consisted of 96 trials split into 4 blocks of 24 trials, with 3 practice trials to familiarise the participants with the task. Using a 6(Intention) X 4(Display Condition) X 2(Viewpoint) design, participants saw each intention twice at both viewpoints, for the four experimental conditions. After each display participants used a six alternative forced choice to select the intention that they thought had been portrayed in the display.

Participants were sat at a distance of approximately 1m, giving a starting visual arc of each circle in both viewpoints of 0.5'. The luminance of the white circles in both viewpoints was 142 cd/m^2, the luminance of the black circles was 0.58 cd/m^2, and the luminance of the grey background was 43.3. This gives a contrast between the white circles and the grey background, in both viewpoints, of approximately 0.5 (Michelson, 1927); a contrast value between the black circles and the grey background of approximately 1; and a contrast between the white circles and black circles of approximately 1. Displays were presented at 30fps, with an average duration of 30secs.

4.3.4 RESULTS

We investigated peoples' ability to discriminate six social intentions using animacy displays derived from human motion. We proposed that people would be able to clearly differentiate the correct intention for each display from both a

side view and an overhead view, at levels greater than chance. Furthermore, we also examined whether increasing the saliency of the viewpoint, via cues of context and occlusion would increase the ability to differentiate intentions. The overall ability to differentiate intentions for each viewpoint, across the four experimental conditions, is summarized in Figure 4.3.2:

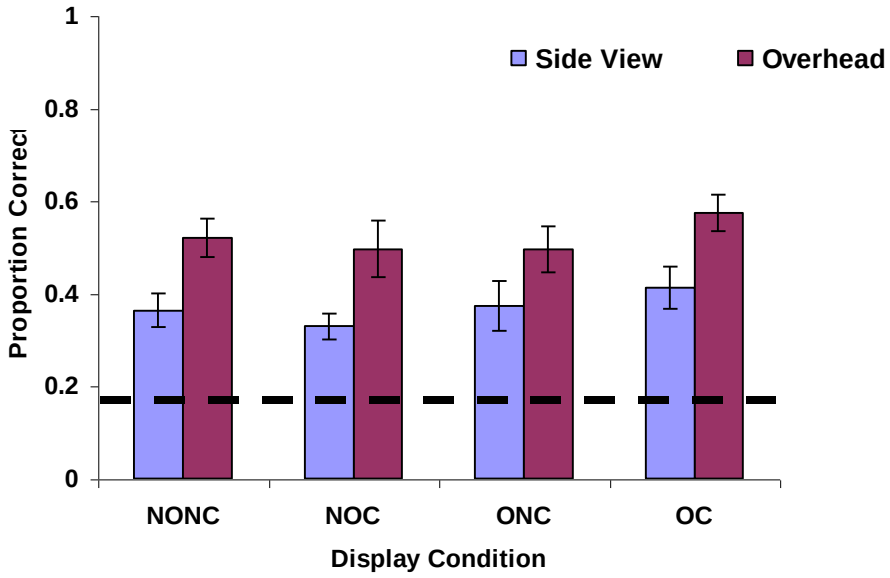

Figure 4.3.2: Proportion correct for observers' judgments of intentions across the four experimental display conditions, at both viewpoints, collapsed across intentions. Error bars indicate standard error and bold dash line indicates chance level: 0.1667.

Participants were clearly able to differentiate the intentions at levels above chance (16.67% correct for a 6AFC) at both viewpoints across all display conditions. It appeared that this ability was better when the displays showed an overhead view rather than a side view.

A three-way repeated measures ANOVA was run – Intention (Chasing, Fighting, Flirting, Following, Guarding, Playing) Vs Display Condition (NONC, ONC, NOC, OC) Vs Viewpoint (overhead, side). The

ANOVA revealed a significant main effect of Viewpoint, *F(1,16)=22.14,* ***p<0.01***, indicating that participants were better at judging intention in overhead displays (52.1%) than in side view displays (36.9%). A significant main effect of Display Condition was found *F(3,48)=3.35,* ***p<0.05,*** with Fischer's LSD revealing that participants are better at categorizing intentions when shown displays of experimental display condition 4 - OC – (49.3%) than the other three experimental display conditions: NONC (44.1%); NOC (41.2%); and finally, ONC (43.9%). A significant main effect of Intention was found, *F(5,80)=3.33,* ***p<0.05,*** with Fischer's LSD revealing that participants were significantly better at categorizing the intentions of Flirting (54.5%) and Following (58.5%) than the intentions of Guarding (35.7%), Playing (37.1%) and Chasing (38.2%). Finally, the ANOVA revealed an interaction between Display Condition and Intention, *F(15,240)=2.37,* ***p<0.05***. Tukey HSD post hoc analysis showed that participants were better at categorizing the Follow intention when shown in experimental display condition 4 (73.5%) – OC – than various other displays: Guarding (29.4%), Playing (36.7%) and Chasing (30.9%) in experimental condition 1 – NONC; Guarding (33.8%), Playing (36.8%) and Chasing (41.2%) in experimental condition 2 – NOC; Fighting (35.3%); Guarding (38.2%), Playing (39.7%) and Chasing (30.8%) in experimental condition 3 – ONC; and finally; Fighting (38.2%), Guarding (41.2%) and Playing (35.3%) in experimental condition 4 – OC. There were no other significant interactions.

The confusion matrices for each intention, collapsed across all four display conditions, can be seen in Table 4.3.1 for overhead and Table 4.3.2 for side view displays:

		Overhead					
		Response					
		Chasing	Fighting	Playing	Flirting	Guarding	Following
Presented	Chasing	0.54	0.01	0.06	0.01	0.01	0.38
	Fighting	0.03	0.4	0.32	0.14	0.1	0.01
	Playing	0.13	0.3	0.43	0.04	0.05	0.05
	Flirting	0.01	0.01	0.14	0.59	0.23	0.02
	Guarding	0.01	0.18	0.14	0.16	0.5	0.01
	Following	0.24	0.01	0.02	0.04	0.03	0.67

Table 4.3.1: Confusion Matrix for categorization of Intentions in Overhead displays collapsed across Display Conditions. Participant responses are shown in the columns with the presented displays shown in the rows.

		Side View					
		Response					
		Chasing	Fighting	Playing	Flirting	Guarding	Following
Presented	Chasing	0.22	0.05	0.21	0.04	0.11	0.38
	Fighting	0.02	0.46	0.35	0.1	0.03	0.03
	Playing	0.07	0.44	0.32	0.1	0.04	0.04
	Flirting	0.01	0.05	0.26	0.49	0.17	0.02
	Guarding	0	0.33	0.29	0.15	0.21	0.01
	Following	0.05	0.05	0.2	0.1	0.12	0.49

Table 4.3.2: Confusion Matrix for categorization of Intentions in Side View displays collapsed across Display Conditions. Participant responses are shown in the columns with the presented displays shown in the rows.

Results from the overhead view show similar confusion patterns to that described by Blythe, Todd & Miller (1999). From Table 4.3.1 it can be seen that for overhead displays we find clear confusions between Fighting and Playing and between Chasing and Following, while the displays of Following, Flirting and Guarding were clearly distinguishable with only slight confusions. The confusion patterns for the side view displays showed greater overall confusion and hence lower percentage correct scores. Similar follow-chase confusions and fight-play confusions were found as in the overhead, though Guarding and

Flirting were also entwined in the latter. Finally, most false alarms were called play suggesting a bias towards this intention.

4.3.5 DISCUSSION

We investigated peoples' ability to discriminate social intentions using animacy displays derived from human motion. We proposed that people would be able to clearly differentiate the correct intention for each display from both a side view perspective and an overhead perspective. Furthermore, we proposed that cues added to clarify viewpoint, i.e. a contextual cue and an occlusion cue, would increase the ability to differentiate intentions.

Participants were indeed able to distinguish intentions at levels greater than chance in both viewpoints, with a clear advantage for displays showing the overhead, (52.1% versus 36.9% correct, collapsed across all conditions). Looking at the confusion matrices, this advantage would appear to be consistent over all intentions except for the fight display where the ability to distinguish intention would appear to be similar for both viewpoints (40% Vs 46%, for Overhead and Side View respectively). It was also found that the context and occlusion cues added to facilitate identification made a small, but non-significant, difference in perceiving the correct intention when presented by themselves, however when the cues were shown in combination, i.e. in experimental display condition 4, they contributed to a significant increase in participants' ability to perceive the correct intention. Furthermore, the overall hit rate in the overhead was similar to that reported by Blythe, Todd and Miller, again validating this technique of animacy display production.

Results from the overhead perspective show similar confusion patterns to that described by Blythe, Todd & Miller (1999). The chase-follow confusion in the current experiment is likely to be resultant of the relative velocities of the shapes in the displays, requiring more investigation solely into the role of velocity in these two actions, and the fight-play confusion in the current experiment may be due to amount of contact involved between the circles in both displays. As increasing the saliency of the viewpoint via boundaries and occlusion produces a small increase in the ability to categorize intention, it would appear that confusions in side view displays cannot be accounted for solely by the misinterpreting of viewpoint, and therefore confusions may result from either an inability to perceive, or a misunderstanding of, the appropriate information that would allow correct interpretation of the intention.

That people are better at attributing intentions to displays shown from an overhead viewpoint is a surprising aspect given that, ecologically speaking, we would have expected people to be better at attributing intention to displays with a side view, our natural viewpoint. One reasonable explanation is that in overhead displays you get more information as to the location of one protagonist to another. This would certainly give an advantage to perceiving motion from overhead, and examples of this can be seen in sports coaches and choreographers plotting the movement of their players or dancers. However, the occlusion cue provided information in the side view displays gave ordinal depth to the circles, yet results showed only a slight increase in ability when this cue was given. This suggests that the additional information provided by the overhead view would exceed ordinal depth information. Further research might explore, in more precise terms, which particular depth or distance relationships are important for

distinguishing an intention. It may be that for some intentions the inclusion of depth information is more relevant than in other displays, which would explain the ability to distinguish the fighting intention being similar across viewpoints, a finding that has been shown in the previous two experiments of this chapter. Analysis that may isolate reasons as to what motion information makes the overhead view better for judging intent in dyadic interactions, compared to the side view, is performed in the following chapter.

In conclusion, we have shown that people are adept at categorizing intentions of human interactions based on the global motion. Results shown in this experiment are consistent with previous findings, validating both this technique and the categories of intentions. It would appear the ability to differentiate these intentions is better in displays showing an overhead viewpoint than a side view. Furthermore, occlusion and contextual cues added to increase the saliency of the viewpoint and thus increase participants' accuracy at judging intentions, had no effect when used separately but produced a slight effect when presented simultaneously.

Experiment 4.4 – Judging the Viewpoint in Animacy Displays

Results from the previous experiments of this chapter have shown that people are adept at distinguishing intentions from the global motions of agents in animacy displays derived from human motion. It has further been shown that this ability is better for displays depicting an overhead view than displays depicting a side view and that the addition of cues to increase the saliency of the viewpoint has only a slight influence on this ability. However, it has also been noted in post-experimental debriefs that participants generally claimed that the

majority of displays had been shown from the overhead perspective in the experiment, where in fact there was an equal number of presentations of both viewpoints. It is likely that an inability to perceive displays from the correct viewpoint would decrease the ability to differentiate intentions, and as participants appear to have a bias to perceiving the overhead display, this may explain the poor performances in the side view displays. The aim of this experiment was to test the ability of participants to differentiate the intended viewpoint of a display using displays showing either an overhead view or a side view.

METHODS

4.4.1 PARTICIPANTS

15 new participants took part in the experiment. All were naïve to the purpose of the study, had normal or corrected to normal vision, and were paid for their participation.

4.4.2 STIMULUS PRODUCTION

The same experimental stimuli from Experiment 4.3 were used in this experiment.

4.4.3 PROCEDURE

The same experimental apparatus was used as in Experiment 4.2. Each experiment consisted of 48 trials split into 4 blocks of 12 trials, with 3 practice trials to familiarise the participants with the task. Using a 6(Intention) X 4(Display Condition) X 2(Viewpoint) design, participants saw each intention

once at both viewpoints for all experimental conditions. After each display participants used a two alternative forced choice to select the viewpoint from which they perceived the display, overhead or side.

4.4.4 RESULTS

The aim of the experiment was to test the ability of participants to differentiate viewpoint between displays showing overhead views and displays showing side views. This is important as an inability to perceive the correct viewpoint may partly explain poorer results in the side view. We tested the null hypothesis that participants could not tell the difference between the viewpoints in the displays. A plot of the results is shown in Figure 4.4.1, collapsed across intentions:

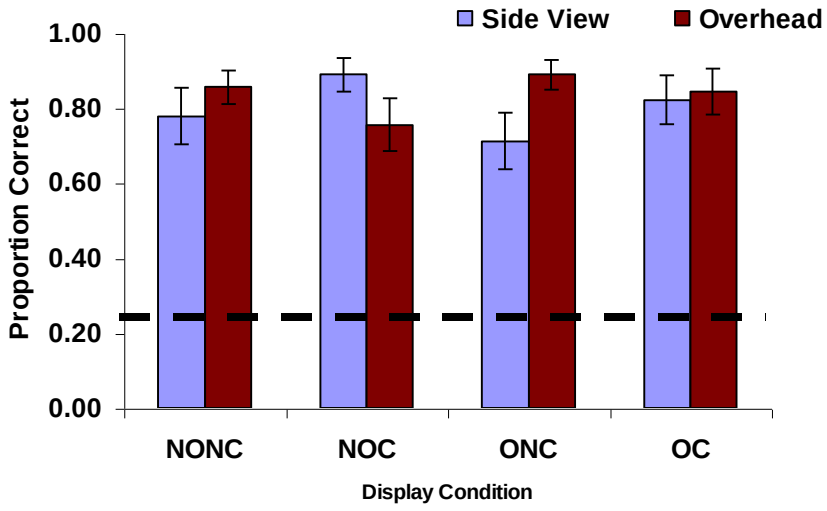

Figure 4.4.1: Proportion correct for observers' judgments of viewpoints across the four experimental display conditions, at both viewpoints, collapsed across intentions. Error bars indicate standard error and bold dash line indicates chance level: 0.25.

Results appear to show that overall people are very adept at distinguishing the correct viewpoint of the displays. High hit rates were scored for the overhead displays and the side view displays, 83.6% and 80% respectively.

A three-way ANOVA was run – Intention (chase, fight, flirt, follow, guard, play) Vs Display condition (NONC, ONC, NOC, OC) Vs Viewpoint (overhead, side). The ANOVA revealed no main effect of Viewpoint, (82% for Overhead and 78% for Side View), no main effect of Display Condition, (NONC - 81.7%; NOC – 82.2%; ONC – 80%; OC – 83.3%), but did reveal a main effect of Intention, $F(5,70) = 3.94$, $\mathbf{p<0.05}$. Fischer LSD analysis revealed that participants are better at judging the viewpoint when the display shows a Playing intention (91.7%) than when the display is Fighting (79.2%), Guarding (73.3%), Following (80.8%) or Chasing (80%). No significant differences between Flirting (85.8%) and any other intention were found.

The ANOVA also revealed an interaction between Viewpoint and Display Condition, $F(3,42) = 4.31$, $\mathbf{p<0.05}$, but no significant differences were found in further analysis using Tukey HSD. Finally, an interaction was found between Display Condition and Intention, $F(15,210) = 1.78$, $\mathbf{p<0.05}$, with Tukey HSD post-hoc analysis revealing that participants were better at judging viewpoint when the display showed a playing intention in experimental display condition 3 – ONC - (97% correct), compared to when the display showed a guarding intention in experimental display condition 2 – NOC - (63%). No other interactions were found to be significant.

The accuracy at judging viewpoint for each intention across the display conditions can be seen in Table 4.4.1 for the overhead displays, and Table 4.4.2 for the side view displays:

	Overhead			
	NONC	**NOC**	**ONC**	**OC**
Chasing	0.87	0.8	0.87	0.87
Fighting	0.73	0.87	0.87	0.8
Flirting	0.93	0.8	0.93	0.93
Following	0.73	0.87	1	0.73
Guarding	0.87	0.47	0.67	0.73
Playing	1	0.73	1	1

Table 4.4.1: Confusion Matrix for proportion correct for judging viewpoint in Overhead displays for all intentions (rows) at each experimental display condition (columns)

	Side View			
	NONC	**NOC**	**ONC**	**OC**
Chasing	0.8	1	0.47	0.73
Fighting	0.8	0.8	0.6	0.87
Flirting	0.73	0.87	0.8	0.87
Following	0.73	0.93	0.67	0.8
Guarding	0.73	0.8	0.8	0.8
Playing	0.87	0.93	0.93	0.87

Table 4.4.2: Confusion Matrix for proportion correct for judging viewpoint in Side View displays for all intentions (rows) at each experimental display condition (columns)

The tables show that, for both viewpoints, proportion correct was above chance of 50% for all intentions at all display conditions, with the exception of Guarding when there is no occlusion cues but boundaries in the overhead

display, and Chasing when there is occlusion cues but no boundary cues in the side view. When viewing these displays, results would suggest that perceiving the correct viewpoint is at chance level.

4.4.5 DISCUSSION

The purpose of this experiment was to test the ability of participants to perceive the correct viewpoint from which to view displays. Overall, results showed that people were able to correctly judge when a display was showing an overhead viewpoint 83.6% of the time, and 80% of the time when it was a side view. This difference was shown to be non-significant. The results therefore suggest that the overall poorer ability to differentiate intentions in displays shown from the side view, as found in previous experiments of this chapter, is not due to an inability to perceive the display from the correct viewpoint. Furthermore, it was found that for only two displays did the ability to judge viewpoint drop below chance, and though this may have slightly affected the ability to perceive intention in these displays, it is unlikely that it would explain the overall reduced ability in side view displays shown in the previous experiment.

In conclusion, this experiment has shown that people are accurate in differentiating viewpoint in the displays. This in turn suggests that poor performances in categorizing intentions in previous experiments, particularly for side view displays, are not due to incorrectly perceiving the viewpoint of the display. It is interesting that people did tend to report in previous experiments

that displays were shown from the overhead viewpoint more often than the side view, and could lead to research into finding the optimal preferred viewpoint for viewing animacy displays and recordings of human motion in general. Finally, this and the previous experiment indicates that poorer ability to judge intentions in the side view displays, is not due to an inability to view the display correctly, nor does the inclusion of boundaries nor ordinal depth increase this ability. We therefore must assume that other depth cues that are missing in the current side view displays are important for judging intention from the side view.

Experiment 4.5 – Free Responses to Animacy Displays of Human Intentions

In the four previous experiments of this chapter, participants have been shown to be successful at judging intentions in animacy displays when provided with multiple options. However, it is possible that participants do not actually perceive the displays to be animate and would not have normally categorised the displays in the manner that they were instructed to. The purpose of this experiment is to use a free response task to look to see if participants do spontaneously describe these displays using animate terms. It is hypothesised that participants will describe the displays as intentional interactions using animate terms in the free response task.

METHODS

4.5.1 PARTICIPANTS

14 new participants took part in the experiment. All were naïve to the purpose of the study, had normal or corrected to normal vision, and were paid for their participation.

4.5.2 Stimulus Production

The same experimental stimuli from Experiment 4.3 were used in this experiment.

4.5.3 Procedure

The same experimental apparatus was used as in Experiment 4.2. Each experiment consisted of 48 trials split into 4 blocks of 12 trials, with 3 practice trials to familiarise the participants with the task. Using a 6(Intention) X 4(Display Condition) X 2(Viewpoint) design, participants saw each intention once at both viewpoints for all experimental conditions. Participants were informed to "Watch the display and after it, describe what happened in the display." Participants wrote down their response. Duration of the experiment varied from participant to participant with an approximate mean of 1hr.

4.5.4 Results

This experiment intended to explore if participants would spontaneously use animate terms to describe the animacy displays used in previous experiments of this chapter. It was expected that participants would use animate terms to describe the displays. Two raters marked the free responses with an overall agreement of 80.1%. The raters marked the responses as being described as animate or not, and then what intention the participant was describing. How the participants described the displays can be seen in Table 4.5.1 for the overhead displays and Table 4.5.2 for the side view displays. The tables show the

intentions collapsed across the display conditions. They also contain columns that show the proportion of times that the displays were not described in animate terms, and when they were described in animate terms that did not match one of the original intentions. Other descriptions included interactions that either did not fit any category or where it was not easily discernable what was the best category to put it in, e.g. dancing.

		Overhead							
		Response							
		Chase	**Fight**	**Flirt**	**Follow**	**Guard**	**Play**	**Other**	**Not-Animate**
Presented	**Chase**	0.46	0.02	0	0.23	0	0.02	0.09	0.18
	Fight	0.02	0.11	0.02	0.02	0	0	0.18	0.66
	Flirt	0.04	0.04	0.13	0.05	0	0	0.09	0.66
	Follow	0.14	0	0	0.54	0	0.02	0.05	0.25
	Guard	0.02	0.05	0.04	0.09	0.16	0	0.05	0.59
	Play	0.25	0.11	0.04	0.11	0	0.02	0.05	0.43

Table 4.5.1: Confusion Matrices for free response task for Overhead displays. Presented displays are shown in rows with participant responses in columns. A response was termed other when participants used a term that did not fit into the other intentional categories, e.g. dancing. A response was termed Not-Animate when participants used no intentional or animate words to describe the motion of the shapes in the display.

		Side View							
		Response							
		Chase	**Fight**	**Flirt**	**Follow**	**Guard**	**Play**	**Other**	**Not-Animate**
Presented	**Chase**	0.21	0.04	0.04	0.34	0	0	0.09	0.29
	Fight	0	0.07	0.02	0.02	0	0	0.14	0.75
	Flirt	0	0.05	0.05	0.04	0	0	0.11	0.75
	Follow	0.11	0.04	0.04	0.34	0	0	0.13	0.36
	Guard	0	0.07	0.07	0.04	0	0	0.16	0.66
	Play	0.05	0.07	0.05	0.04	0	0.11	0.04	0.64

Table 4.5.2: Confusion Matrices for free response task for Side View displays. Presented displays are shown in rows with participant responses in columns. A response was termed other when participants used a term that did not fit into the other intentional categories, e.g. dancing. A response was termed Not-Animate when participants used no intentional or animate words to describe the motion of the shapes in the display.

The tables show high proportions of non-animate descriptions in the side view displays, and to a slightly lesser extent in the overhead displays. In particular, guard in the side view was never correctly categorized, and fighting (7%) and flirting (5%) in the side view, and playing (2%) in the overhead view were all poorly categorized correctly. Only the displays showing chasing and following were correctly described in animate terms in accordance with their original intention, for both viewpoints, at proportions much higher than chance.

A graph for the proportion of displays described as animate, collapsed across display condition, for both viewpoints can be seen in Figure 4.5.1, and a graph for the proportion of displays described in accordance with their original intentions can be seen in Figure 4.5.2:

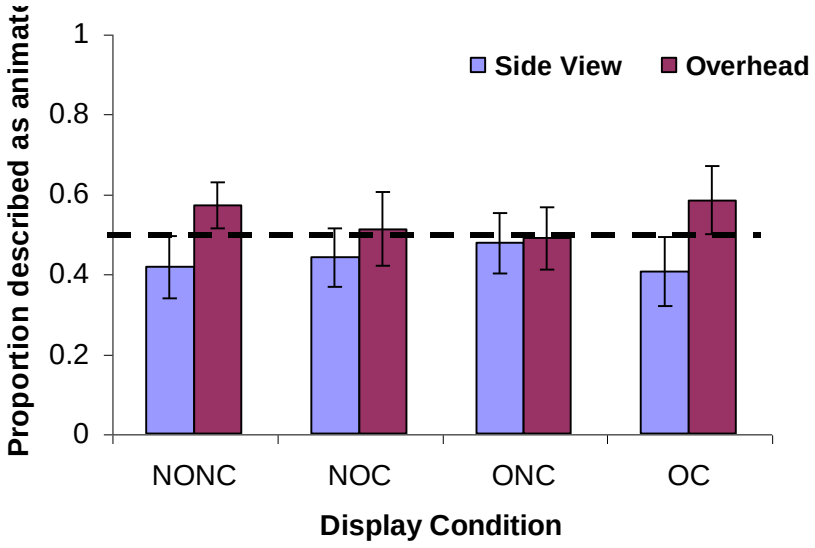

Figure 4.5.1: Proportion of displays described in animate terms collapsed across intentions, including displays that did not fit the pre-determined intentions. Error bars indicate standard error and bold dash line indicates chance of describing any display in animate terms: 0.5.

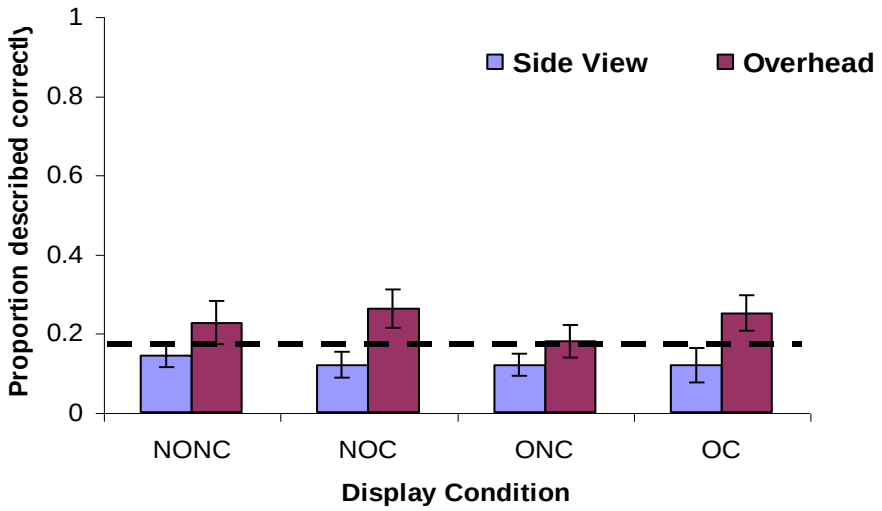

Figure 4.5.1: Proportion of displays described using the correct intention collapsed across intentions. Error bars indicate standard error and bold dash line indicates chance of describing the displays with the correct intention if all displays had been described as animate using only the six pre-determined intentions: 0.1667

Overall people described 54% of the overhead displays as being animate and 43% of side view displays. When looking at whether participants described the displays in accordance with the original intention of the displays, participants used the original intentions 13% of the time on side view and 23% of the time on overhead displays.

A three-way ANOVA was run – Intention (chase, fight, flirt, follow, guard, play) Vs Display Condition (NONC, ONC, NOC, OC) Vs Viewpoint (overhead, side), solely looking at the data of displays described as animate, irrespective if they were described using the correct intention. The ANOVA revealed a main effect of viewpoint, $F(1,13) = 22.98$, $\mathbf{p<0.05}$, with post-hoc Fischer LSD analysis showing that participants were significantly more likely to use animate terms to describe a display if it portrayed the motion of the shapes

from an overhead viewpoint (53.9%) as opposed to portraying the motion form the side view (43.5%). The ANOVA also revealed a main effect of Intention, $F(5,65) = 4.73$, $\mathbf{p<0.05}$, where it was found, again using Fischer's LSD, that participants were significantly more likely to use animate terms when the display showed a chase (77.7%) or a follow (69.6%) than when it showed the intentions of fighting (30.4%), flirting (39.3%), guarding (28.6%) or playing (46.4%). Also, participants were again more likely to use animate terms when the display showed a play intention (46.4%) rather than a fighting (30.4%) or flirting (39.3%) intention. No significant effect of display condition was found: NONC (49.4%); NOC (47.6%); ONC (48.2%); and finally, OC (49.4%). Finally, no significant interactions across any combination of experimental variables were found.

4.5.5 DISCUSSION

This experiment examined the use of spontaneous animate terms to describe the animacy displays created for previous experiments in this chapter, to explore if participants would normally describe the animacy displays using the options given in the forced choice tasks of the first three experiments of this chapter. It was suggested that perhaps participants do not actually perceive the displays as being animate and that this fact is being missed due to the nature of the forced choice task that has previously been employed. It was expected that participants would describe the displays in animate terms but perhaps not always describing the displays in accordance with the original intention used to make them. Results of the experiment showed that participants would describe the overhead displays in animate terms around 54% of the time, and around 43% of

the time for side views. Participants used terms that indicated the original intended intention of the display much less.

The experiment suggests that participants do not always perceive the displays as having intentional movement, especially in the side view. Participants will describe a display in animate terms approximately 50% of the time according to the results of the present experiment. This fact could help explain poor performances in the forced choice tasks. If participants do not perceive any intention in the motion then it will be difficult for them to correctly categorize the different intentions in the displays. Nevertheless, it must be noted that participants can still categorize the intentions of the displays to a level above chance using a forced choice task. Therefore, when given the options and forced to differentiate the intentions, cues in the motion of the agents can still indicate the intended intention even though participants might not have readily described the intention in animate terms, given the situation of a free response task.

Heider and Simmel (1944), using a free response task, reported participants giving detailed stories about the agents in their display and attributing meanings to the interactions. It is worth noting that the displays used in the experiments in this chapter are markedly sparse in regards to detail and story structure compared to the Heider and Simmel movie. Where they used three agents of differing shape, size, etc, and blended one action/intention into another, we have used simple displays that show just one intentional interaction. It would be worth investigating the effect of merging one interaction into another, for example, a chase becoming a fight, to see the effect on the participants' free responses. It is possible that changes in intentions of two agents may indicate different parts of a larger whole, thus eliciting some form of

social schema in the participants. This would be similar to some of the work of Zacks (2004) and Zacks and Tversky (2001), where it is asserted that people need to have some concept of the schema of an event to understand it, i.e. that it is understandable for a chase to turn into a fight, etc.

This experiment appears to be the only direct comparison of the use of alternative forced choice tasks and free response tasks in the animacy literature, and certainly the first that does so using animacy displays derived from recordings of human actions. One major indication of this experiment is that it backs up the conclusions of Experiments 3.1 and 3.2, which show that not all human movement can be automatically turned into displays that will be described as animate, and enhances this by revealing this to be true, irrespective of the viewpoint used to view the display from. In turn, this points to the suggestion that animacy displays though predominantly influenced by the spatial and temporal dynamics of the global motion of the agents, are also influenced by other factors, possibly the merging of one intention into another or prior knowledge of social schemas. Finally, the results of this experiment, in conjunction with results of previous experiments of this thesis, indicate that though participants may not always perceive the displays as animate, when forced to select from limited options, participants can correctly distinguish between intentions in displays using cues available in the global motion.

Experiment 4.6 – Judging Intentions in Short Duration Animacy Displays

The concluding experiment of this chapter diverges from the previous experiments in order to investigate the speed at which participants can make accurate judgements about the intentions in animacy displays. We propose that

in life we can often make quick judgements of intention and with this experiment we intend to test the ability of participants to judge the intentions of the displays used throughout this chapter, over short durations of 1, 5 and 10 seconds. Furthermore, as this experiment looks solely at how efficient participants can be, we will only use the overhead displays as these are the displays that are more often described in animate terms in free response tasks, and participants are better at judging the intentions of shapes within these display when using forced choice tasks. It is hypothesised that participants will successfully be able to distinguish between the intentions of Chasing, Fighting, Flirting, Following, Guarding and Playing at levels greater than chance for all 3 time intervals and that participants' ability to correctly categorize intention will improve as duration of display increases.

METHODS

4.6.1 PARTICIPANTS

13 new participants took part in the experiment. All were naïve to the purpose of the study, had normal or corrected to normal vision, and were paid for their participation.

4.6.2 STIMULUS PRODUCTION

Only the overhead animacy displays from Experiment 4.3 were used in this experiment. Adobe Premier Pro (Version 1.5) was used to create 1 second, 5 second, and 10 second versions of these displays. Each display, for all time intervals, started at the first frame of the full-length displays and ended after the relevant time period, with displays being recorded at 30fps.

4.6.3 PROCEDURE

The same experimental apparatus was used as in Experiment 4.2. Each experiment consisted of 72 trials blocked by time periods. For all time durations, participants saw 2 blocks of 12 trials, with 3 practice trials to familiarise the participants with the task. It was decided that so participants could not base their judgement of shorter duration displays on their judgement of longer duration displays, all participants saw the 1 second display blocks first, followed by the 5 second display blocks, and finally the 10 second display blocks. This may however introduce an order effect whereby participants may improve from one display duration to another due to increased familiarity with the displays, and this point will be considered when reviewing the results of the experiment. Using a 6(Intention) X 4(Display Condition) X 3(Duration) design, participants saw each intention twice at the three time periods for all experimental conditions. After each display participants used a six alternative forced choice to select the intention that they thought had been portrayed in the display.

4.6.4 RESULTS

This experiment tested the ability of people to judge intention in animacy displays shown for a brief duration. We selected 3 incremental time durations of 1 second, 5 seconds and 10 seconds. We only tested on the overhead displays as we wanted to examine how efficient participants could be and these displays have been previously shown to be the displays that people are most accurate at judging the intentions in. It was expected that participants would be able to accurately judge the intention in the displays at all time intervals. A plot of the

data for each time duration at all four experimental display conditions, collapsed across intentions can be seen in Figure 4.6.1:

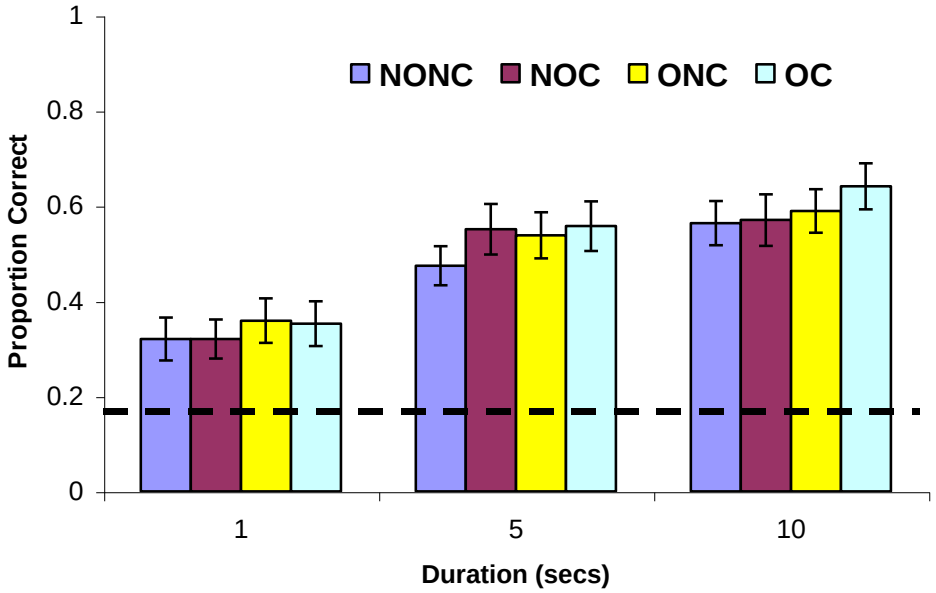

Figure 4.6.1: Proportion Correct for the categorization of intentions across 3 experimental time durations and four experimental display conditions. Error bars indicate standard error and bold dash line indicates chance: 0.1667.

The graph shows that participant can judge intentions in displays at levels above chance for all display conditions, at all time durations. Overall participants had a hit rate of 34%, 53% and 59% for 1 second, 5 second and 10 second displays respectively.

A three way ANOVA was run – Intention (Chasing, Fighting, Flirting, Following, Guarding, Playing) Vs Display Condition (NONC, ONC, NOC, OC) Vs Duration (1 second, 5 second, 10 second). The ANOVA revealed a main effect of Duration, $F(2,24) = 26.7$, $\mathbf{p<0.05}$, with Fischer's LSD analysis revealing that participants were better at judging intention for the 10 second displays (59%) and the 5 second displays (53%) than for the 1 second displays

(34%). No significant difference was found between accuracy in judging intentions between the 5 second and 10 second displays. The ANOVA also showed a main effect of Intention, $F(5,60) = 11.1$, $\mathbf{p<0.05}$, with further analysis, again using Fischer's LSD, showing that participants were better at judging intention in Following displays (68.6%) than in displays showing Fighting (53.8%), Guarding (28.2%), Playing (31.4%) and Chasing (53.5%). Furthermore, participants are better at judging intention in Fighting, Flirting and Chasing displays than in Guarding and Playing displays. In addition, the ANOVA revealed a main effect of Display Condition, $F(3,36) = 3.17$, $\mathbf{p<0.05}$, with Fischer LSD analysis showing that participants were better at judging intention in displays of experimental condition 4 – OC – where boundaries and occlusion cues are present (51.7%) than when neither are present, experimental display condition 1 (45.3%) – NONC. Though adding just boundaries to the displays (experimental display condition 2 – NOC) slightly increased ability to judge intention (48.1%), as did adding just occlusion cues (49.6%) (experimental display condition 3 – ONC), these differences were not significant.

The ANOVA revealed an interaction between Display Condition and Intention, $F(15,180) = 2.16$, $\mathbf{p<0.05}$. Tukey HSD post-hoc analysis showed that participants are better at judging follow displays when the display shows boundaries but no occlusion cues (73.1%) than they are when it is a guard display with no boundaries or occlusion cues (20.5%). No more interactions were found to be significant.

The confusion matrices, collapsed across display conditions, for 1 second displays, 5 second displays and 10 second displays can be seen in Table 4.6.1, Table 4.6.2 and Table 4.6.3 respectively:

		1 second					
		Response					
		Chase	Fight	Flirt	Follow	Guard	Play
Presented	Chase	0.38	0.05	0.04	0.47	0.02	0.04
	Fight	0.01	0.41	0.31	0	0.06	0.21
	Flirt	0.1	0.2	0.26	0.08	0.2	0.16
	Follow	0.17	0.01	0.07	0.64	0.02	0.09
	Guard	0.07	0.35	0.19	0.09	0.13	0.17
	Play	0.11	0.13	0.1	0.03	0.45	0.19

Table 4.6.1: Confusion Matrices for categorizing intentions in displays of 1 second time duration, collapsed across experimental display condition.

		5 second					
		Response					
		Chase	Fight	Flirt	Follow	Guard	Play
Presented	Chase	0.62	0	0.01	0.26	0.01	0.11
	Fight	0	0.56	0.22	0	0.1	0.13
	Flirt	0	0.16	0.72	0	0.06	0.06
	Follow	0.15	0.02	0.09	0.67	0.01	0.06
	Guard	0.01	0.27	0.17	0.02	0.3	0.23
	Play	0.01	0.19	0.05	0	0.43	0.32

Table 4.6.2: Confusion Matrices for categorizing intentions in displays of 5 second time duration, collapsed across experimental display condition.

		10 second					
		Response					
		Chase	Fight	Flirt	Follow	Guard	Play
Presented	Chase	0.61	0	0	0.24	0	0.15
	Fight	0	0.64	0.05	0	0.03	0.28
	Flirt	0	0.02	0.71	0.01	0.16	0.1
	Follow	0.21	0	0.03	0.74	0	0.02
	Guard	0.01	0.22	0.16	0	0.41	0.19
	Play	0.01	0.31	0.07	0	0.18	0.43

Table 4.6.3: Confusion Matrices for categorizing intentions in displays of 10 second time duration, collapsed across experimental display condition.

It can be seen from the tables that in general there is less confusion as to the intention in each display as the duration of the displays is increased. Only the intention of following is consistently well distinguished irrespective of time

duration of display. The ability to spot the intentions of playing and guarding do increase but not to the same level as that of the other intentions. Playing is mistaken with fighting, as is seen in the previous experiments in this chapter using 30-second duration displays. Guarding is often mistaken for fighting, playing or flirting, and was the only intention not to be distinguished at levels greater than chance in the 1 second displays.

4.6.5 DISCUSSION

It was proposed that often we can make quick judgements about the intentions of another's movement and the present experiment examined the ability of participants to distinguish between intentions in animacy displays of brief duration: 1 second, 5 second and 10 seconds. Results showed that participants could accurately judge all but one intention at levels above chance, for all time durations. Only guarding could not be recognised for displays of 1 second duration however, as this intention could be recognised in displays of longer durations, this indicates that the movement of the agents post 1 second, determines the intention as guarding. That all other displays can be distinguished at short durations suggests that the visual system is attuned to recognising these intentions quickly. It is possible that this process has become adapted for survival reasons, validating theories that these intentions are generic in human nature and have in the past served the purpose of maintaining survival and increasing the chances of reproduction (Blythe et al, 1999; Barrett et al, 2005).

The experiment showed that there was a significant difference in ability between 1 second and 5 second displays and between 1 second and 10 second displays, but not between 5 and 10 second displays. It would appear that around

the duration of between 5 and 10 seconds, participants reach a plateau of ability. In comparison, in Experiment 4.3, where the full duration displays were used, the ability to judge intention was equivalent to the ability to judge intentions in the 5 second displays – 52% and 53% for full and 5 second displays respectively. However the ability when shown 10 seconds of the display is higher at 59%. It would be interesting, using displays of lengths intermediate between 10 seconds and the full duration, to examine the ability to judge the correct intention as the display continues to fruition. It is possible that people make a correct quick judgement but as the display plays out to its full duration, participants may begin to rethink their judgements, leading to more mistakes and more confusion. However, we must also consider the possibility that the order in which the displays were shown had an effect on the results of this current experiment.

Participants were shown the displays in order of shortest duration to longest duration in order to prevent participants being able to categorize intentions in short duration displays based on prior knowledge of longer duration displays. This however may create order effects where participants would improve at judging intention just by becoming more familiar and more experienced with the displays. An alternative way to run the experiment would have been to use different participants for different display durations, or to randomise the displays and run more trials, assuming that participants are not connecting one display with another. That said, from the results of the experiment it is suggestible that the order of displays had only a small effect for two reasons. Firstly, the movement in the 1 second displays is very small for all displays and post experimental debriefs reported participants being bemused that they were actually being asked to judge intention from such small movements –

though the results do show participants being above chance for this duration. It is therefore unlikely that participants would have gained much knowledge from these displays that they could have used to make their judgements in later trials. Secondly, the difference in accuracy for categorizing intentions between the 5 second displays and the 10 second displays is small and not significant. Therefore, any knowledge or experience participants gained in judging the 10 second displays by viewing the 5 second displays first, appears to be minimal. These reasons would appear to support the argument that the results may have been only slightly influenced by any potential order effects, however it must be conceded that the optimal way to run the experiment would have incorporated a between subjects design.

Another interesting comparison between this experiment and Experiment 4.3 is the relevance of cues to boundaries and occlusion. In the previous experiment it was found that the combined inclusion of boundaries and occlusion cues had a small significant difference on peoples' ability to judge the intentions, but neither of these cues had an effect when used alone. In this experiment we find again that the people are better at judging intentions when the displays have both boundaries and occlusion cues. It must be noted that the occlusion cues in this experiment offer no cue to occlusion, as there can be no ordinal depth in the overhead viewpoint. They do however provide cues to identity, but are referred to as the occlusion cue to make references to the earlier experiment easier. The results show that both these cues, when combined, can be made use of in displays of short duration, and appear to have the same slight effect of enhancing peoples' ability to categorize intentions of animacy displays, compared to when neither are present.

In conclusion, this experiment has shown that people can make quick and accurate judgements of intentions in animacy displays derived from human motion. It is possible that this ability has lived on from when accurate judgements of intentions had distinct survival purposes. It appears that at short durations people make can make use of cues in a similar manner as they do in displays of longer durations. Finally, this experiment suggests that people can be more accurate when viewing shorter durational displays than when viewing the whole display which could lead to interesting research into how the interpretation of intention builds up over time.

Chapter Discussion and Summary

In this chapter we proposed to continue the work started in the previous experimental chapter on the perception of motion in animacy displays, but to do so using a specific set of actions/interactions. The previous chapter introduced a new technique for the production of animacy displays where the displays are truly representative of human motion, since they are derived from video footage of human actors. It was found that this technique was successful for producing displays that, when shown to observers, the agents within the displays would be described as alive or self-propelled. The final experiment of the previous chapter introduced the topic of viewpoint in animacy displays and showed the surprising result that agents viewed from an overhead position appear more alive and in control of their own movements than agents shown from a side on position. A limitation of the first experimental chapter was possibly that the intentions and goals represented by the displays used were too broad a range to be systematically analysed. This current chapter tried to rectify that problem by

solely using a specific set of dyadic intentions to create the displays. The intentions used, taken from the work of Blythe, Todd and Miller (1999) and Barrett, Todd, Miller and Blythe (2005), were: Chasing, Fighting, Flirting, Following, Guarding and Playing. These papers found that participants could achieve high hit rates when asked to distinguish between these intentions, and Barrett et al (2005) also showed that this was also possible in non-Westernised cultures.

This chapter showed initially, in Experiments 4.1 and 4.2, that it was possible to create displays of the above intentions by first recording human actors and then degrading the displays into animacy displays, and moreover that it was possible to achieve success rates well above chance for distinguishing these intentions in both the overhead and side-on viewpoints. We then, in Experiment 4.3, examined the effect of adding cues for context, i.e. boundaries, and cues to show ordinal depth in the displays, i.e. occlusion, to increase the ability to perceive these intentions, and it was shown that these cues had a small but significant effect when combined in displays. It was then proven, in Experiment 4.4, that participants could perceive the correct viewpoint in displays as it was suggested that an inability to do so would hamper participants' ability to correctly categorize intentions. Experiment 4.5 revealed that participants, in a free response task, would describe some of the displays in animate terms but not necessarily with the intentions intended. Finally, Experiment 4.6 found that people were able to successfully judge the intentions in the displays after very short excerpts of the displays.

A fundamental change was made in the production of displays in this chapter from the previous one. In the displays of the previous chapter it was

noted that if the actors got too close then there co-ordinates would fuse and this would create an unnatural "jump" between the circles in the resultant displays. This was found to be an artefact of Eyesweb. It was decided to filter the co-ordinates in order to smooth this fusing of co-ordinates as well as to remove remnants of vertical motion associated with the gait cycle, seen in the side view displays. A low pass fourth order Butterworth filter with a cut-off frequency of 0.8Hz was used. The filtering removed the "jumps" and gait oscillations, but also removed some motion signal which made the movements appear slightly smooth. It was however decided by the author to use the filtered animacy displays, as they appeared closer to the original human motion of the initial video recordings. It would of course be better to be able to extract the co-ordinates so that the filtering was not necessary, and programming is underway to achieve this, making use of new advancements in video technology such as High Definition recordings. An alternative method would be to film the actors in different colours, for example, red, green or blue, and to use colour filter techniques found in most video manipulation packages to extract one actor at a time and track their co-ordinates. This method would still have a slight flaw when one actor passes in front of the other, the result would be a black screen, but for the majority of scenarios this would only be for a very brief duration.

The first two experiments in this chapter were designed as pilots for the third experiment. They showed that using the technique for production of animacy displays discussed in the previous chapter, as well as the added filtering, it was possible to create animacy displays of the six intentions of Blythe et al (1999) and of Barrett et al (2005) that produced hit rates equivalent to the previous literature. Experiments 4.1 and 4.2 showed hit rates of around 50% for

the overhead displays and around 30% for the side view displays. A hit rate of approximately 50% in the overhead displays is equivalent to that of Blythe, Todd and Miller (1999), validating our technique for the production of displays. A hit rate of approximately 30% for side views is consistent with findings in the previous chapter, Experiment 3.4, that the side view displays are not always perceived as being as animate, and therefore we expected it to be harder to distinguish between intentions in this viewpoint. Both experiments showed that Following and Guarding were much better at being seen in the overhead than in the side view but that ability to judge the intentions of Fighting and Playing was primarily viewpoint independent. As both experiments were designed as pilot studies, the first one to test the displays, and the second experiment to improve the experimental technique, making it more lab-based, no strong conclusions were drawn from the results, though some speculations could be made as regards the judgements of intentions across viewpoints.

The surprising aspect of Experiments 4.1 and 4.2 was the difference in the ability to perceive Flirting between experiments. In Experiment 4.1, Flirting appeared to be viewpoint independent, but in Experiment 4.2 it was markedly better when viewed from the side view. The second experiment was run in a lab-based environment with repeated measures taken from each participant and perhaps the repeated viewings allowed participants to make more accurate judgements. Another possible cause is that Experiment 4.2 only used five displays, chase being removed, and perhaps the fewer options allowed for clearer understanding of the displays.

Of further note from these pilots was an indication in post-experimental debriefs that participants had a tendency to report perceiving the majority of

displays from the overhead viewpoint. A failure to perceive the displays with the correct viewpoint, in particular the side view displays would indeed help explain the poorer performance in this viewpoint.

Experiment 4.3 aimed to test the ability to distinguish the six intentions from both viewpoints, whilst incorporating cues to clarify viewpoint and depth relationships between agents. Two cues were used: a contextual cue that showed the boundaries of the stage that the agents were moving on, and an occlusion cue that showed ordinal depth in the side view. As noted earlier the occlusion cue gave an identity cue in both viewpoints. The results showed hit rates very similar to the previous experiments in this chapter and to that of Blythe et al, namely 52% accuracy in distinguishing intentions in the overhead displays versus 37% accuracy in the side view displays. Again, this validated both the use of these specific intentions and the use of this method for the production of displays. The added cues to clarify viewpoint were shown to have a small significant effect when both were shown in the displays but to have no effect when used separately. A significant effect of viewpoint showed that there was a difference in ability to judge intentions from different viewpoints, with overall ability being better for overhead displays. This difference in judging intentions across viewpoints is similar to the difference in the perception of self-propulsion viewed in the previous chapter, Experiment 3.4, which showed that agents in overhead displays appeared more self-propelled than agents in side view displays. It is possible that the results of Experiments 4.3 and 3.4 are all linked by motion cues that have been degraded and are thus stopping the motions appearing as they should. Experiment 4.3 showed that ordinal depth cues, contextual cues and identity cues have only a small effect in enhancing the

perception of intention in animacy displays and therefore we must look to other cues that are lacking as being the differences in perception across viewpoints. We can make an educated guess to these cues based on the work of previous authors (Blythe et al, 1999; Barrett et al, 2005; Zacks, 2004) that has shown that cues such as the distance between agents and the angles between the agents, as well as speed and acceleration all play important parts in creating the perception of animacy. Furthermore, it is evident that the importance of these cues varies between displays, with some displays, in particular Fighting, being recognised at the same consistency on both viewpoints. The cues available in the displays, and that are useful for intention differentiation, will be given a formal analysis in the following chapter.

Confusions in the overhead displays of Experiment 4.3 were similar to those found in the original research of Blythe et al, with Fighting and Playing being confused, and Chasing being mistaken for Following. The side view, as expected given its lower overall accuracy scores, showed greater confusions, with similar Chasing-Following confusion and Fighting-Playing confusion, but now including confusions with Fighting, Playing, Guarding and Flirting. These confusions would appear to be partly due to the original displays used to make them. The Chasing-Following confusion is likely due to the speed of the actors as the motion was exactly the same. It is thought that if the actors ran faster in the original recording of the chase then this might clarify this confusion. The Fighting-Playing confusion is again due to the original display but harder to improve on. In the original display of the Fighting, no punches were thrown and the fight was more of a heated argument and aggressive stepping rather than an actual fight. The original recording of the Playing intention was a game of tag

where, though playing, did involve a lot more physical contact between the actors. These differences in physical contact between agents within the animacy displays are thought to help explain the confusion. The added confusions of Guarding and Flirting in the side views is most likely due to the confusion in the perception of the actual locations of the agents, and in particular when one agent is actually just passing the other at a distance, in the side view it can look like both agents are touching or pushing each other.

Also, in both viewpoints but more so the side view, there is a tendency to call intentions play, an effect also noted by Blythe et al (1999). They explained this by saying that there is an underlying belief in people that play is a more common intention and furthermore that we often use play as children to learn the other intentions. If we accept the role of play in learning intentions, we are left with a top-down hypothesis where we understand an intention because we have acted it or have previously been witness to it. As a child playing, is probably one of the few occasions in life when we are in a position that is equivalent to an overhead animacy display, where we have a direct viewpoint to the motion below. If play is how we learn to determine intentions, then it is possible that our ability to distinguish intentions from an overhead viewpoint is learnt at a young age during normal development. This would be consistent with research that shows that children of young ages can distinguish intentional motion (Csibra, Gergely, Biro, Koos and Brockbank, 1999; Gergely, Nadasdy, Csibra and Biro, 1995; Kuhlmeier, Wynn and Bloom, 2003) and would fit with a theory that the brain has the components to judge intentions from birth, however it requires development at a young age (Frith and Frith, 2001).

Post-experimental debriefs of Experiments 4.1, 4.2 and 4.3 showed that participants had a tendency to report the majority of displays as being viewed from the overhead perspective, even after cues to facilitate the viewpoint had been added. The fourth experiment of this chapter looked to test this informal observation. It was found that participants could tell an overhead display from a side view display at hit rates of 83.6% and 80% respectively. In only two displays did the success rate drop below 50% and though this may explain poor performance in judging intentions in these two displays, it does not explain the overall poorer performance in the side view compared to the overhead view. Furthermore, that participants could successfully determine viewpoint in sparse environments would be consistent with the cues added to facilitate this ability, in Experiment 4.3, only having a slight effect.

The penultimate experiment of this chapter tested participants' spontaneous use of animate terms to describe the animacy displays of Experiment 4.3 using a free response task. Displays in Experiments 3.3 showed poor free responses even though results from ratings of self-propulsion tasks indicated that people do perceive the agents to be alive. As we had already ascertained that people could judge the intent of the agents within the display, we tested whether participants would give the same intentions to the displays when not prompted with possible answers. The results were slightly disappointing, indicating that approximately only 50% of the displays were seen as animate in the overhead displays, with a lower percentage for the side view displays. The use of the terms that were consistent with the pre-determined intentions of the displays was much lower for both viewpoints. These results are consistent with previous findings of this thesis, that displays created using the new technique do

not freely elicit animate responses from participants when asked to do so. As previously indicated in this thesis, it is possible that it is the sparseness of the displays used in this thesis that may hinder the use of animate terms. In comparison, the display of Heider and Simmel (1944) had different shapes of varying sizes with the intentions and beliefs of each agent changing and altering to react to current situations in the display. Research has shown that the aesthetic qualities of the displays do not have an effect on the perception of animacy but Tremoulet and Feldman (2006) stated that the perception of an agent as animate is influenced by the motion patterns and by its movement in context with its environment. It would be interesting to test if these displays would elicit more animate terms if some context or purpose was given to the movement of the agents, or if one intention was streamed into a second one, for example a chase into a fight, or a follow into a flirt. This would give meaning to the display and distinct events allowing for more scope for a story to be made from the movement.

Experiment 4.5 also adds support to a statement made in the last chapter, based on the results of Experiment 3.1, 3.2 and 3.3, where displays made from human motion were not described in animate terms, that not all human motion can be simplified using this technique to produce animacy displays. This statement is probably not entirely correct and should be readdressed to say that not all human motion can be simplified to create displays that will elicit animate responses. It is evident from earlier experiments in this chapter that show the displays to be perceived as self-propelled and people able to judge the correct intentions from the global motion of the displays, when given the appropriate options. The research of this thesis would suggest that the difference in

perceiving animacy and freely describing a display as such is yet to be uncovered, and that the sparseness of displays and the changing of intentions by an agent to react to its environment may have a crucial role in this difference.

The concluding experiment of this chapter introduced research looking at the judgement of intention in animacy displays at short intervals. Blythe et al (1999) stipulated that the intentions had reproductive and survival qualities, and therefore the ability to spot them must be universal and efficient. The experiment tested the ability to judge the intentions at 1 second, 5 seconds and 10 seconds. Results showed that overall people were above chance at all time intervals emphasising the efficiency of the human perception system to distinguish these intentions quickly. A significant difference was found between the 1-second displays and the 5 and 10 second displays but not between the 5 and 10 second displays. This would indicate that there is an improvement around the 5-second time interval and a plateau of sorts is reached. As mentioned earlier in this chapter, the ability to judge intention at 5 seconds is the same as it was for the full displays in Experiment 4.3, but is lower than ability to judge intention in the 10-second displays. It is possible that people peak then start to second-guess their judgement once the initial need to judge the intention quickly has passed, which would lead to interesting research into unravelling the time course of perceiving intentions.

Finally, across this chapter we have seen hit rates of around 50%, which is comparable to that of Blythe et al (1999) but is lower than Barrett et al (2005) who achieved hit rates of approximately 80%. These authors concluded that the choreography of the displays resulted in this difference, where one was highly controlled and the other was not. This reasoning would be consistent with hit

rates for intentions in the displays used in this chapter, as the displays were initially highly choreographed and perhaps higher hit rates could be achieved by creating displays from video footage where actors have solely been told the intention to portray and have been given no specific instructions of how to do so.

In conclusion, using animacy displays created from human motion we have shown that people, solely using whole body motions, can distinguish between specific, generic intentions, and that this ability appears to be better for displays depicting an overhead perspective. Furthermore, this ability is not due to a failure to determine the veridical viewpoint of the displays. Results are consistent with previous research on animacy, showing that motion patterns are key to the understanding of intention, however it is suggested that contextual factors may facilitate animate descriptions of displays, akin to those reported by Heider and Simmel (1944). Regarding viewpoint, at present it is unclear if a predisposition forms the advantage of the overhead viewpoint or if it is a learnt ability. Furthermore, what information is required to reduce confusion in displays shown from the side view is still unclear, and will be analysed in the following chapter. Finally, results have shown that people are adept at judging intentions quickly and efficiently, making use of available cues to do so.

The findings of this chapter and the previous chapter serve as the basis for an investigation into the application of animacy displays created from human motion in a study of Autistic Spectrum Disorders reported in the Chapter 6. First, we will present a comprehensive analysis of the kinematics and motion patterns of the displays used in this chapter, allowing for a fuller understanding of the motion properties that participants use to differentiate between intentions.

Chapter 5 – Kinematics of Human Intentions Displays

Via the 3-D co-ordinates of the recorded human interactions, we can attempt to determine what information observers use to categorise different interactions. Though the data obtained in this thesis is not yet vast enough to determine what information is used in general, we can probe what data we have to better our understanding of the information available in our images of natural interactions. Hitherto, only verbal descriptions, such as in Chapter 4, of the human motions have been discussed. Though this may give some understanding of how the displays appeared, it gives little insight as to the kinematic qualities of the movements of the agents. Previous research has attempted to perform a similar analysis via providing descriptions of kinematic properties of animacy displays (Barrett, Todd, Miller and Blythe, 2005; Blythe, Todd and Miller, 1999; Zacks, 2004).

Barrett et al (2005) and Blythe et al (1999), examined what cues were required to distinguish the intentions of Chasing, Courting, Fighting, Flirting, Following, Guarding and Playing. The authors determined seven cues useful for intention categorization: Relative Distance, Relative Angle, Relative Heading, Absolute Velocity, Relative Velocity, Absolute Vorticity and Relative Vorticity. Blythe et al (1999) entered the data of these cues into a "frugal" algorithm called Classification by Elimination (CBE): by entering the cues in a pre-determined sequential order, this algorithm uses the minimal number of cues required for accurate judgement of intentions. It is essential to enter the cues in an appropriate order to acquire optimal classification performance. Analysis

revealed that absolute velocity, i.e., the forward velocity of an agent with respect to its' background, was the best classifier for intentions followed by, in decreasing order, relative angle, relative velocity, relative heading, relative vorticity, absolute vorticity and relative distance. It was concluded that these cues have strong diagnostic qualities in the classification of basic dyadic interactions. Behavioural studies showed that participants achieved an overall classification performance of approximately 50% when both agents were shown, with performance reducing to approximately 30% if an agent was removed: thus highlighting the importance of the relative cues between the agents.

Zacks (2004) explored motion properties in animacy displays to establish how people segment sequences of actions into events, determining how motion properties relate to action segments. He examined what information participants used when informed the movement was intentional vs. random, and when told to use short vs. long parsings to segment displays. Zacks found that observers rely more on motion properties to understand intentions when segmenting sequences into the smallest, understandable parsing, yet rely other information such as prior knowledge, when parsings are larger. A set of motion cues were examined including position, speed, acceleration, distance, relative speed, relative distance, and the maximum and minimum values of these cues. Zacks used a stepwise regression to establish that the highest proportion of the variance was accounted for by the relative distance of the agents, i.e. how far apart the agents were, and the acceleration of the agents. In accordance with previous authors, Zacks concluded that observers use motion properties of displays to determine the intentions portrayed by the agents, as well as prior knowledge regarding context, experience, etc.

From reviewing previous research, the motion properties that seem most pertinent in understanding intentions relate to the absolute velocity and acceleration of an agent, and to the relative distance, velocity and acceleration between agents.

5.1 Motion Properties in Animacy Displays from Human Motion

We will examine various motion properties of the displays used in Chapter 4 to find patterns that differentiate the intentions of Chasing, Fighting, Flirting, Following, Guarding and Playing. While our database of recorded movements is not yet large enough to draw firm generic conclusions, we will consider the data presented in the previous chapter when discussing the current analysis of motion properties. This analysis could, in future work, be applied to the experiments of Chapter 3, determining motion properties that predict animacy in general.

Considering previous research (Barrett et al, 2005; Blythe et al, 1999; Tremoulet and Feldman, 2000; Zacks, 2004) we focus our analysis on speed, acceleration and relative distance. The kinematics of the displays will be calculated for both viewpoints examined: Side View and Overhead. We expect our analysis to reveal possible explanations for the confusions between judgements of intentions, primarily between Chasing and Following, and between Playing and Fighting. This analysis should also highlight differences between displays viewed from the Side View and from the Overhead. We will first present analysis of full duration displays, followed by analysis of shortened Overhead displays with time frames of 1, 5 and 10 seconds. Presenting analysis of these short duration displays from Experiment 4.6 shows how motion properties change as the duration increases, and may explain why participants'

judgements substantially improve when the duration is increased from 1 to 5 seconds, but not from 5 to 10 seconds. Finally, we will perform a stepwise regression using an array of motion predictors and the behavioural data from Experiment 4.3, to establish the parameters that best predict each intention. This regression only allows for tentative conclusions due to the concise set of stimuli used, but serves as an indicator for future work on the topic.

The motion properties examined are adapted from Zacks (2004) and described in that paper as follows:

Speed – The speed of each agent or the magnitude of the agent's instantaneous velocity, where velocity of an agent, in both the x and y directions, is calculated by numerical differentiation of the position of an agent, i.e. its' planar 'x' and 'y' location.

Acceleration – the magnitude of each agent's acceleration.

Relative Distance – the distance between two agents.

Relative Speed – the speed at which agents were moving towards or away from each other, as calculated by numerical differentiation of the relative distance.

Relative Acceleration – the acceleration at which agents were moving towards or away from each other, as calculated by numerical differentiation of the Relative Speed.

We will present the range and average for each of these measures. The stepwise regression will make use of the mean and standard deviations of these and other parameters to establish the best predictors. Furthermore, the distance travelled by each agent and the differences in these distances will be examined.

As a starting point for understanding the motion patterns and properties of different intentions, we present 3-D plots of the agents' co-ordinates.

5.1.1 Plotting the Motion Patterns

The positional co-ordinates of each actor are extracted from the original video displays of human motion using Eyesweb (Camurri, Trocca and Volpe, 2002). The co-ordinates are filtered via Matlab 5 (the Mathworks, Natick) and the Pschophysics Toolbox Version 2.5 (Brainard, 1997; Pelli, 1997), as discussed in Experiment 4.1, using a 4^{th} order low pass, Butterworth filter. The filtered co-ordinates are the data set used to explore the kinematic properties and motion trajectories. Initial exploration of the 3-D plots and of histograms of the accelerations and speeds, revealed some unusually high values. It is likely these "jumps" in co-ordinates arise from the tracking problems of Eyesweb, discussed in Chapter 2. As these "jumps" were not true representations of the gestures, we removed them by trimming 10% of the highest and lowest values of each motion property examined was performed.

The 3-D plots, Figure 5.1.1 – Figure 5.1.6, provide an indication of differences in the movement in each intention across viewpoint. The figures represent the positional co-ordinates of each actor across time, for the Overhead and Side View, of Chase, Follow, Play, Fight, Guard and Flirt respectively. The main confusions witnessed in the behavioural data were: between Follow and Chase, and between Play and Fight. The plots are orientated to highlight the '*y*' co-ordinate of the actors as it is in this dimension that displays should show greater difference across viewpoints. The horizontal axes of the plots show the 'x' and 'y' positional co-ordinates, and the vertical axes show the time.

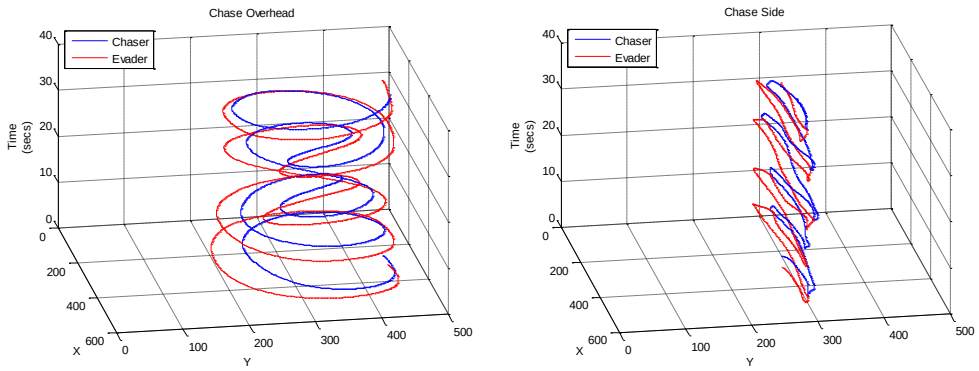

Figure 5.1.1: 3D plots representing the positional co-ordinates of actors performing Chase for both viewpoints: Overhead (left plot) and Side View (right plot). Vertical axis shows time (secs). Actor/Agent being Chased is shown in red.

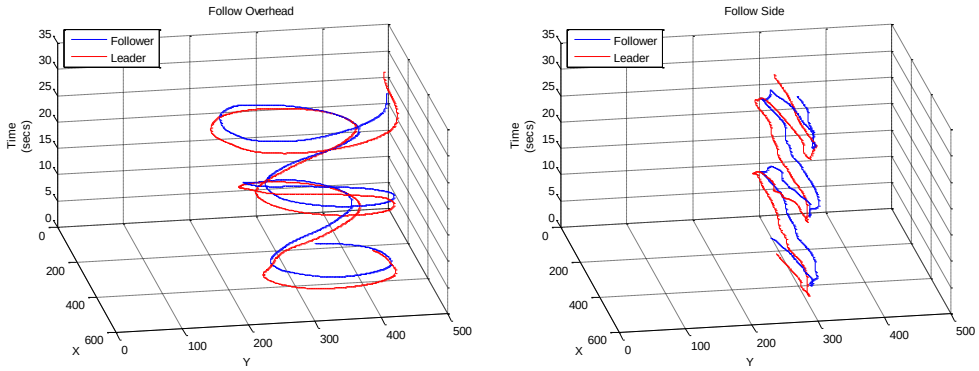

Figure 5.1.2: 3D plots representing the positional co-ordinates of actors performing Follow for both viewpoints: Overhead (left plot) and Side View (right plot). Vertical axis shows time (secs). Actor/Agent being Followed is shown in red.

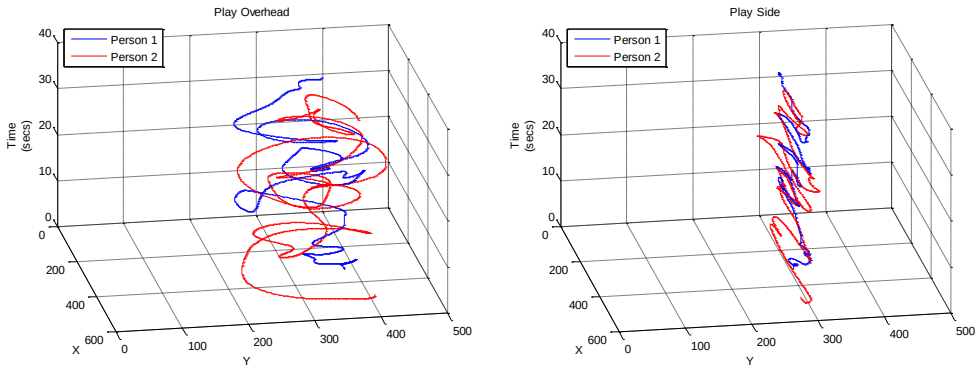

Figure 5.1.3: 3D plots representing the positional co-ordinates of actors performing Play for both viewpoints: Overhead (left plot) and Side View (right plot). Vertical axis shows time (secs).

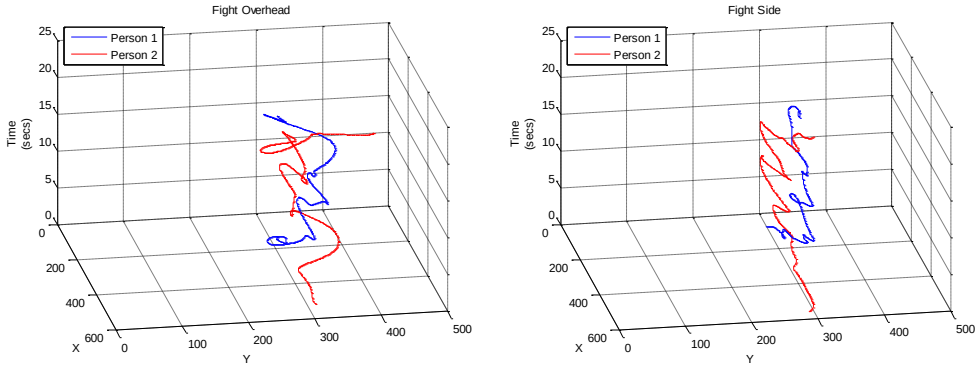

Figure 5.1.4: 3D plots representing the positional co-ordinates of actors performing Fight for both viewpoints: Overhead (left plot) and Side View (right plot). Vertical axis shows time (secs).

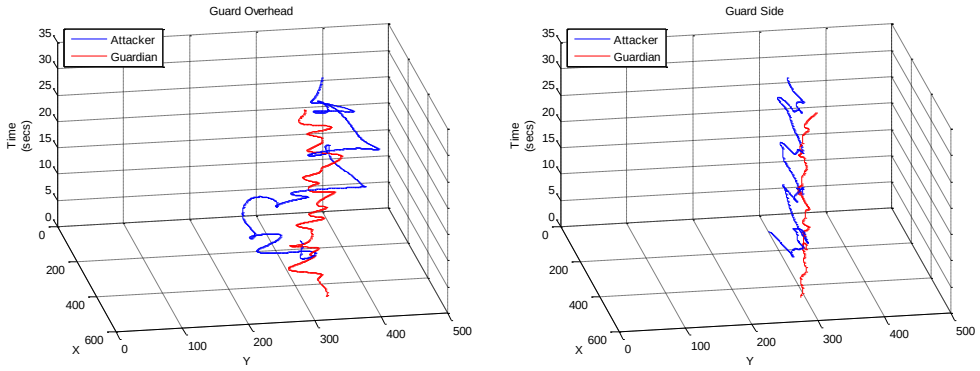

Figure 5.1.5: 3D plots representing the positional co-ordinates of actors performing Guard for both viewpoints: Overhead (left plot) and Side View (right plot). Vertical axis shows time (secs). Actor/Agent Guarding is shown in red.

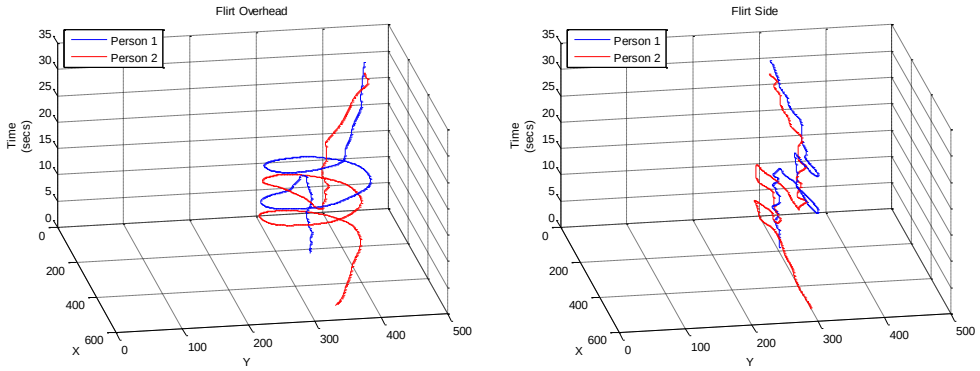

Figure 5.1.6: 3D plots representing the positional co-ordinates of actors performing Flirt for both viewpoints: Overhead (left plot) and Side View (right plot). Vertical axis shows time (secs).

5.1.2 Discussing the Motion Patterns

An important finding of the experiments carried out thus far has been that participants are better at categorising intentions when the displays are shown from the Overhead. Furthermore, agents in Overhead displays are perceived as more animate or self-propelled than their Side View equivalents. The proposed reason for this discovery is that the Overhead displays carry more information regarding the relationship between agents. Side View displays only show the motion in a horizontal plane, with depth and distance information not being made available to observers, making it more difficult to make accurate judgements regarding the intent. The 3D plots confirm this.

The Overhead plots show both agents travel in varying directions around the 'x' and 'y' space. The plots are made up of loops and curves, indicating agents moving in a circular motion, as well as straight lines, indicating agents moving in a direct manner. The Side View plots only show movement in the 'x' dimension. This explains why agents in the Side View appear to move only left and right in straight lines. As a result of these differences in appearance, we would expect participants to be better at judging intentions in the Overhead due to more available information for differentiating displays. The behavioural data does indeed show poorer overall ability to recognise intentions, with greater levels of confusion, in the Side View.

The plots show a large difference in the motion between viewpoints. However the Overhead and Side View plots of the Fight displays look fairly similar. The same is true for the Guard displays. It is possible that large differences in the appearance of the motion plots between viewpoints may indicate differences in accuracy, and perhaps vice versa. If this were true then

we would expect similar accuracy for intention judgements of Guard and Fight across viewpoints. Results of Experiment 4.3 show this is the case only when viewing the Fight displays: participants show no great advantage of viewpoint (40% vs. 46% for the Overhead and Side View displays respectively). For Guard displays there was a clear advantage for viewing the display from the Overhead (50% vs. 21% for the Overhead and Side View displays respectively).

The movement by the actors in the displays of Fighting and Guarding is similar: in both, the actors continually come together and repel apart, with the actors moving in straight, direct paths for the majority of the duration. The main difference being that in the Fight displays both actors are deliberately moving towards each other in the shortest route, whereas in the Guard displays the attacker is trying to get past the Guardian, resulting in slight deviations from a direct path when the agents/actors are close together. These deviations in the Guard displays can cause the attacker to go in front of or behind the Guardian, thus occluding one agent in the Side View but not in the Overhead. It is probable that this fine movement in depth gives the advantage to viewing the Overhead displays. As the actors' paths are linear in the Fight displays there is no advantage in the Overhead display. Another difference between these intentions is the limited movement by one actor in the Guard displays, which participants may use to determine between these intentions. However, a stationary agent is not solely indicative of Guard, as stationary agents are also found in Flirt.

For the other intentions, the Overhead and Side Views show clear differences. As the Side View plots are similar across intentions, and tell us little about the movement in the intentions, we will focus on the Overhead plots. The plots of Chase and Follow look similar. Both have smooth circular motion, with

one agent's motion pattern being almost an exact copy of the other. The person who is following or chasing has a tighter graph, indicating that they were moving in smaller circles than the person being followed or chased. This makes ecological sense as the person following/chasing sees where the other is heading, and can alter their direction accordingly, i.e. showing predictive ability. The Overhead plot of Flirt shows periods of no motion by one agent, whilst being circled by the second agent, and then the reversal of roles. This reciprocal action may be used by participants to distinguish flirting. Finally, the Overhead plot of Play shows various types of movements, including loops and curves, and quick bursts in straight lines. The movements that make up Play consist of a mixture of elements from other intentions: the looped and curved motion seen in Follow and Chase, and the quick movements in straight lines seen in Fight and Guard. The bias observer towards Play may result from this mixture of elements from other intentions. Blythe, Todd and Miller (1999) argued participants are biased to Play as via play we learn all other intentions. Our results suggest participants may incorrectly categorise a display as Play because the movements of the agents involved crossover between intentions.

While the 3-D plots provide some indication of differences between different intentions and viewpoints, the quantitative analysis of the data which follows will further our understanding of these differences.

To quantify these patterns in motion, we shall perform a kinematic analysis of the positional co-ordinates of the displays to look for patterns in the motion properties that indicate intentions. As mentioned in the introduction of this chapter, we will focus on the speed and acceleration of agents, both absolute and relative, and the distance between the agents. These properties have been

shown to be best for classifying intentions (Barrett, Todd, Miller and Blythe, 2005; Blythe, Todd and Miller, 1999; Zacks, 2004).

5.2 Kinematics of Human Intentions in Animacy Displays

Tables below show the average (AV) and range (R) of motion properties we considered as factors that allow participants to classify the intentions (Barrett et al, 2005; Blythe et al, 1999; Tremoulet and Feldman, 2000; Zacks, 2004). Columns in the table correspond to: the speed of Person 1 (Speed 1), the speed of person 2 (Speed 2), the acceleration of Person 1 (Acceleration 1), the acceleration of Person 2 (Acceleration 2), the relative distance between the two agents (Rel. Dist.), the relative speed between the two agents (Rel. Speed), and the relative acceleration between the two agents (Rel. Acc.). Tables are shown with values for both viewpoints, Overhead and Side View, with rows in the tables corresponding to the six intentions: Chase, Fight, Flirt, Follow, Guard and Play. For Chase and Follow, Person 2 is leading/being chased. For Guard, Person 2 is guarding. Values are calculated in pixels for distance, pixels/sec for speed, and pixels/sec/sec for acceleration.

Full Duration			Speed P1	Speed P2	Acceleration P1	Acceleration P2	Rel. Dist.	Rel. Speed	Rel. Acc.
Overhead	Chase	AV	104.8	127.2	93.7	117.9	128.8	114.1	118.5
		R	81 : 127.6	97.1 : 153.7	52.5 : 129.8	54.7 : 165.2	89.9 : 171.5	52 : 155.2	72.6 : 162.2
	Fight	AV	55.1	63.6	92.6	101.0	153.3	96.2	168.4
		R	17.1 : 103.3	18.6 : 112.4	28.5 : 225.2	36 : 210.1	112.3 : 235	32.9 : 155.6	70.3 : 324.1
	Flirt	AV	37.8	47.3	34.9	37.8	86.6	71.9	62.4
		R	2.8 : 89.6	2.2 : 89.3	9 : 102.6	9 : 85.4	54.5 : 145.7	37.2 : 89.7	20.1 : 108.7
	Follow	AV	80.0	89.3	55.7	61.8	110.9	69.3	64.5
		R	65 : 93.9	75 : 104.2	25.5 : 99.4	25.5 : 100.6	90.4 : 135.4	27.4 : 95.4	27 : 127.9
	Guard	AV	56.6	23.6	92.3	60.5	144.9	53.1	91.4
		R	20.4 : 101.9	5.2 : 47.6	36 : 183.4	12.7 : 140.6	103.1 : 236.8	19.9 : 100.9	28.5 : 207.8
	Play	AV	70.4	111.8	131.1	179.3	141.1	125.1	202.8
		R	18.9 : 132.2	47.3 : 180.1	36 : 270	74.2 : 303.6	81.5 : 195	54.6 : 193.7	91.8 : 390.4

Full Duration			Speed P1	Speed P2	Acceleration P1	Acceleration P2	Rel. Dist.	Rel. Speed	Rel. Acc.
Side View	Chase	AV	93.3	108.5	82.8	107.0	112.0	92.3	154.5
		R	18.3 : 153.6	27.6 : 185	20.1 : 189.9	28.5 : 230.7	20.3 : 209.1	17.9 : 193.3	45.9 : 282.6
	Fight	AV	60.4	66.9	119.9	120.0	173.1	97.6	209.8
		R	12.5 : 116.4	14.4 : 138.3	28.5 : 262.4	37.1 : 272.4	79.9 : 279.1	19.2 : 186.9	57.6 : 361.8
	Flirt	AV	35.8	44.7	53.3	51.7	69.9	60.0	99.7
		R	3 : 101.1	2.7 : 100.5	9 : 171.2	12.7 : 145.1	12 : 149.5	12.1 : 112.4	27 : 201.2
	Follow	AV	72.6	76.7	68.9	69.3	96.5	47.3	124.3
		R	21.2 : 122.7	19.1 : 139.1	20.1 : 192.4	18 : 179.3	8.1 : 148.7	11.7 : 152.3	27 : 334.1
	Guard	AV	53.9	13.2	87.0	39.4	165.4	55.3	101.0
		R	10.2 : 118.4	3 : 42.8	20.1 : 234.2	9 : 117	95.6 : 275	11.5 : 134.2	25.5 : 291.5
	Play	AV	58.8	98.2	123.7	165.9	121.1	103.8	206.8
		R	11.7 : 139.6	25.2 : 198	28.5 : 297	40.2 : 352.8	16 : 218.6	21.8 : 210.6	64.9 : 432.1

Figure 5.2.1: Tables showing the Average values (AV) and Range (R) of various motion properties that are thought to be important factors in determining the intention between two agents. The top table shows the data for the Overhead displays and the bottom table shows data for the displays shown from the Side View. Distance is recorded in pixels, Speed in pixels/sec, and Acceleration in pixels/sec/sec. 100 pixels equates to approximately 1 metre. In the displays of Chasing and Following, Person 2 is always the person being Chased or Followed. In the Guard display, Person 2 is always Guarding.

After recording a square cube at the initial recording sessions, we equate approximately 100 pixels on the monitor to be 1m on the stage. Results are discussed systematically, taking each intention in turn, with general conclusions being drawn at the end.

For agents in the Chase Overhead display we see high average speeds of approximately 100 pixels/sec, with narrow ranges indicating the agents maintained these speeds throughout the display. Agent 2 shows a slightly larger range (Speed P2, top table) indicating they were going slightly faster at points then slowing down: perhaps attempting to maintain a lead from the pursuer, or deciding what route to take. The Relative Distance range indicates actors maintained a distance between themselves of approximately 90 pixels (0.9m), with this distance increasing at points, indicative of the person being chased trying to get away. Similar values are seen in the Side View display, though with slightly lower, average values. The main differences between viewpoints are in the ranges of these properties. Overhead displays only have large ranges in individual agents' accelerations. Ranges in all properties of the Side View are large, particularly in Relative Distance and Relative Acceleration. The problem with the Eyesweb tracking system, discussed earlier, may account for large fluctuations in acceleration. Large fluctuations in distance are likely due to agents passing each other and agents running towards or away from the camera. Differences in ranges across viewpoints indicate how altered a display will appear across viewpoints. Similar patterns are expected in the other intentions, when comparing motion properties across viewpoint: except in Fight where no difference in ability to recognise intention across viewpoint was found.

Analysis of the Fight Overhead display shows agents moved fairly slowly, with moments of large accelerations. The original instructions of Fight told agents to aggressively step towards each other, causing large accelerations when the attacker moves towards the other agent who is then repelled. Low average speeds are due to agents always facing each other, negating any running away at high speeds. Interestingly, the Relative Distance between the two agents is high (mean = 153.3 pixels, range = 112.3 to 235 pixels). This suggests agents were never closer than 1m. Such a distance may help explain reduced classification of Fight, as usually fighting would involve being close to your opponent. The problem of Fight appearing as such is discussed again in the general discussion of the thesis. The Side View motion properties show very similar results to the Overhead data: maintained slow average speeds; moments of large accelerations. Furthermore, a large relative distance is again maintained. In reality, actors may have come closer to each other than suggested and this has been removed by the data trimming. However, trimming only affects brief incidents of coming together. If they have been removed, there was probably few of them: supporting the lack of contact between agents. Motion properties indicate Fight looked similar across viewpoints, clarifying why similar levels of successful categorisation were found.

In the Flirt Overhead display, both agents have low average speeds and accelerations, with moments of very low speeds, possibly when one was stationary watching the other move. The averages and ranges of Relative Speed and Relative Acceleration are low, suggesting both agents to be moving at and maintaining the same speed. This could possibly indicate reciprocal movement. The Relative Distance is towards the low side of its range indicating agents to be

near each other for the majority of the display. The Side View data shows similar average speeds and accelerations, but with higher maximum values in the ranges. These differences are again probably due to the tracking problems of Eyesweb, altering to appear as though they may be running away from each other, instead of slowly apart. Furthermore, the Relative Distance has a smaller average but a larger range, than in the Overhead display: again likely due to agents appearing to crossover in the Side View, instead of circling each other, as in the Overhead. These comparisons suggest in Flirt, agents moving faster apart and coming closer together in the Side View than in the Overhead. The clarity of the speed of and distance between agents in the Overhead view may account for improved intention recognition in this viewpoint: 59% vs. 49%, for Overhead and Side View respectively.

Experiment 4.3 showed Follow was correctly recognised 67% of occasions in the Overhead, and 49% in the Side View. It was also evident that Chase was often called Follow, but not so much vice versa. In the Follow Overhead displays, agents kept average speed of approximately 85 pixels/sec with little fluctuation, indicated by narrow ranges. Furthermore, the average acceleration of each agent, the Relative Speed and Relative Acceleration are low, suggesting agents moved at similar speeds, not accelerating away from each other. The average Relative Distance was 110.9 pixels, which is near the middle of a small range, suggesting the agents maintained a constant distance between themselves throughout. This contrasts with Chase Overhead where large fluctuations in Relative Acceleration, Relative Distance and Relative Speed were found, indicating an agent trying to get away. Though there are clear differences in the properties of these two displays, the behavioural data shows Chase

confused as Follow. Perhaps the average speeds of agents in Chase were slow enough to make participants think they were viewing a Follow display. The problem of the Chase scenario being too slow is discussed in detail in the general discussion of this thesis.

Follow in the Side View shows similar averages as in the Overhead display, in regards to absolute and relative speeds, however, absolute and relative accelerations are elevated. As in other side displays, for Follow, ranges for all the motion properties are larger than in the Overhead displays. Furthermore, the Side View Relative Distance of Follow has a smaller average than in the Overhead display, but a larger range. Agents in the Side View Follow appear to accelerate and decelerate more, with greater speeds, than in equivalent Overhead display. Also, the Relative Distance, maintained in the Overhead, fluctuates more in the Side View: again most likely when the agents come towards or away from the camera. Finally, the Relative Acceleration in Side View Follow has a very high maximum value, signifying bursts of acceleration. These increased accelerations in the Side View may partly explain the confusion between Chase and Follow.

The motion properties of Guard appear very similar across viewpoints, yet behavioural results reveal better classification in the Overhead displays (50% vs. 21%). In both viewpoints, the attacker moves more than the guardian, and at faster speeds. However, the acceleration and speed of the guardian are larger in the Overhead displays. It is unlikely that these small differences would account for the large discrepancies in the behavioural data. It is perhaps the limited movement of the guardian, in terms of physical displacement and changes in speed and acceleration, which serve as indicators of Guard. Furthermore, the

additional information given by the Overhead display of the attacker attempting to get past the guardian, information not seen in the Side View display, will play a part in helping to perceive this intention.

Similarly to the other Side View displays, Side View Guard is confused as Play by approximately one third of participants. This confusion/bias is reduced for Overhead displays, with only Fight being confused as Play regularly: again by approximately 30% of participants. Play in the Overhead is perceived correctly by 43% of participants, and by 32% in the Side View displays. It has been suggested that the bias to Play is because we learn the other intentions via Play (Blythe et al, 1999). Therefore, Play can perhaps be thought as an all-encompassing super-category, containing elements of the other intentions. In the kinematic analysis we would expect Play to have wide ranges with averages approximately in the middle, showing that the speeds, distance and accelerations regularly fluctuate between large and small. The tables in Figure 5.2.1 confirm this. We find averages falling almost in the centre of the ranges, with both extremely high and low accelerations and distances. The only measure in Play that does not follow this pattern is the speed of person 1, which would indicate that this person was not moving as fast as the 2nd person. Furthermore, we again see larger ranges for all categories in the Side View than in the Overhead view, partly explaining the greater levels of confusion in the Side View.

Motion properties, such as the speed and acceleration, are coupled in reality with the movements performed by the actors. By combining the trajectories and motion properties, we can get an understanding of how each intention is performed. For example, in the Overhead displays, in Follow and Chase both agents take very similar trajectories, maintaining a steady relative

distance, but the speed and acceleration of each agent is faster in Chase than in Follow. It has been mentioned, however, that agents' speeds in the Chase display were not significantly faster than in Follow, hence resulting in some confusions. We can also see that Fight and Guard are similar with both sets of agents taking direct paths at low average speeds. However, both agents in Fight have large accelerations, but in Guard, only the attacker has so. Furthermore, Flirt is also characterised by low speeds and accelerations, but the agents take looped, circular trajectories, with periods of reciprocal activity. The trajectory of Play contains both looped, circular motion and direct motion, and the motion properties have large ranges with approximately centred averages. This suggests that Play contains elements of all the other intentions and that large variations in motion indicate Play.

Overall, variations are evident in the motion properties and trajectories that can enable participants to differentiate intentions in the Overhead displays. Furthermore, when similarities occur between intentions in either the kinematics or in the trajectories, participants may use the differences in the other cue to determine intent. This ability would appear to be reduced in the Side View displays. In the Side View displays, the trajectories all contain straight, direct paths. This similarity may account in part for the greater confusion in this viewpoint. In addition, in the kinematic analysis of the Side View displays, though some intentions show similarities with the equivalent Overhead displays, they are perhaps not distinct enough from other Side View displays to aid accurate recognition. This would explain the reduced accuracy in intention recognition in the Side View, resulting in increased levels of confusion.

For additional understanding of the displays, using the kinematics, it is possible to obtain a rough approximation of the distance travelled by each agent, and compare how this varies within and across intentions, and across viewpoints. The distance travelled by each agent, calculated via the product of the average speed and duration of display, for all six intentions in Overhead displays is shown in Figure 5.2.2, and the distance travelled by agents in the Side View displays is shown in Figure 5.2.3:

Distance Travelled				
Full Duration		P1	P2	DIFF (P2 > P1) (%)
Overhead	Chase	4113.3	4993.6	21.4
	Fight	2162.2	2498.5	15.6
	Flirt	1486.2	1859.2	25.1
	Follow	3140.6	3507.5	11.7
	Guard	2220.7	925.2	-58.3
	Play	2762.8	4391.5	60.0

Figure 5.2.2: Distance travelled in pixels by each agent in the six intentions of the Overhead displays. In Chase and Follow, P2 is the person being Chased or Followed. In Guard, P2 is Guarding. DIFF represents the distance travelled more by P2 than P1, as a percentage of P1's distance.

Distance Travelled				
Full Duration		P1	P2	DIFF (P2 > P1) (%)
Side View	Chase	3661.8	4259.4	16.3
	Fight	2370.9	2625.7	10.7
	Flirt	1406.3	1756.1	24.9
	Follow	2851.3	3012.0	5.6
	Guard	2118.0	518.8	-75.5
	Play	2309.3	3857.0	67.0

Figure 5.2.3: Distance travelled in pixels by each agent in the six intentions of the Side View displays. In Chase and Follow, P2 is the person being Chased or Followed. In Guard, P2 is Guarding. DIFF represents the distance travelled more by P2 than P1, as a percentage of P1's overall distance.

Overall, agents travel further in the Overhead displays: not surprising given that in this viewpoint the agents move in two planes, where as they only move in one plane in the Side View displays. The differences in distance travelled, across viewpoint, again highlights changes in displays as we alter the viewpoint from which they are perceived: information showing how far each agent travels is lost in the Side View. Fight is the exception as agents in the Side View display cover more distance during the display. This is interesting as Experiment 4.3 showed Fight was the only display in which intention recognition was aided by a Side View viewpoint (40% vs. 46%, for Overhead and Side View respectively). In future research it may be worth exploring in greater depth, if there is a relationship between distance travelled and successful intention recognition at different viewpoints.

In five of the intentions, Guard being the exception, P2 covers more distance than P1. From the 3-D plots it was noted that in the intentions Chase and Follow, the agent who was following/chasing, P1, took a tighter path inside that of the leading/pursued agent, P2. It was suggested that this was due to the Chaser/Follower being able to see what was happening in front of them and adjusting their course accordingly. The data in Figures 5.2.2 and 5.2.3 fits in with this theory. However we must err on the side of caution because P1 always travelled slower on average, not covering as much distance. This is perhaps a limitation of the current data set, created using only two actors. If with more actors, thus more examples, similar findings are shown, we could perhaps be more conclusive in our findings. That said, this data would suggest that a person trailing has the advantage of being able to adjust their trajectory accordingly with respect to the person in front, saving energy.

5.2.3 Summary of Analysis of Full Duration Animacy Displays.

This thesis has shown that animacy displays can be created via the extraction of co-ordinates from videos of human activity. A main finding has been that agents shown from an Overhead viewpoint are perceived as more animate/self-propelled than if shown from the Side View. Chapter 4 used animacy displays to examine intention recognition in participants, looking at the intentions of Chasing, Fighting, Flirting, Following, Guarding and Playing. Results showed people to be better at intention recognition when they viewed displays from the Overhead than from the Side View. However confusions were found between displays such as, Chase being called Follow, Play and Fight being confused, and also a general bias towards Play. Furthermore, the confusions were heightened in the Side View displays. This current chapter has explored these displays to give a fuller understanding of how the interactions looked, and to look for differences and similarities in the trajectories and motion patterns that may explain the behavioural data.

3-D plots of the 'x' and 'y' co-ordinates of each agent across the duration of the displays were examined. Across viewpoint, differences were found between the general appearances, with the Overhead plots giving more information. Between intentions, intentions in the Overhead displays could be characterised by different types of trajectories: Chase and Follow are characterised by circular, looped trajectories; Fight and Guard appear to have more direct trajectories. Play appears to contain parts of all of these intentions, and it is hypothesised that this maybe a reason as to a bias towards calling displays Play. A lack of variation in trajectories of displays in the Side View is thought to result in more confusion in this viewpoint: in the Side View,

characteristics of intentions, such as how far apart the agents are and whether they are moving around each other or are in contact with each other is lost. We performed a kinematic analysis of the co-ordinates, focusing on properties that previous research had shown to be beneficial in intention recognition: speed, acceleration and distance (Blythe et al, 1999; Zacks, 2004). Intentions were shown to vary in terms of the average speed and acceleration of the agents within the display. For example, agents in Chase move at consistently faster speeds than the other displays. The kinematic analysis highlighted possible differences in the motion properties that participants may use to discriminate the intentions. Furthermore, the variation between the motion properties of the Side View displays is reduced, with the intentions all containing similar characteristics. It is thought that similarities in motion properties and motion trajectories across intentions in the Side View result in greater confusion in this viewpoint. Conversely, differences in motion properties and trajectories are what aid successful intention recognition in Overhead displays.

5.3 Analysis of Short Duration Animacy Displays.

Experiment 4.6 showed that participants were able to judge intentions accurately after very brief durations: durations of 1 second, 5 second and 10 second. A significant improvement from viewing 1 second displays (34% accuracy) to viewing 5 second displays (53% accuracy) was found, but no so when duration was increased from 5 second to 10 seconds (59% accuracy). Furthermore, the ability to accurately discriminate displays at 5 seconds appears as good as when viewing the full displays (52.1% accuracy), with ability at 10 seconds better still.

We examined the motion trajectories and properties, such as speed, acceleration and distance, of these short duration displays, to explore how the information changes over time. As Experiment 4.6 only tested the Overhead displays, we analysed only these displays. Plots of the motion trajectories for all six intentions and a table of the motion properties is shown in Figure 5.3.1 for 1 second displays; Figure 5.3.2 for 5 second displays, and Figure 5.3.3 for 10 second displays:

1 Second			Speed P1	Speed P2	Acceleration P1	Acceleration P2	Rel. Dist.	Rel. Speed	Rel. Acc.
Overhead	Chase	AV	46.5	89.2	61.5	100.7	69.7	43.8	57.5
		R	29.4 : 72.6	48 : 131.4	27 : 90	63 : 126.3	55.5 : 88.8	20.7 : 58.5	32.4 : 76.9
	Fight	AV	58.6	40.7	47.5	101.4	347.1	101.1	136.7
		R	44.8 : 70.2	13.7 : 82.7	27 : 57.6	56.9 : 131	300.6 : 379.8	46.7 : 157.2	90.4 : 171.2
	Flirt	AV	3.2	66.3	12.6	28.3	262.7	64.2	31.3
		R	1.8 : 4.9	60.2 : 70.7	9 : 20.1	18 : 38.2	240.2 : 284	58.1 : 65.7	12.7 : 45.9
	Follow	AV	81.0	93.2	37.9	27.6	95.7	59.8	29.5
		R	78.9 : 82.5	91.9 : 94.8	27 : 45	12.7 : 37.1	95.2 : 96.6	53.7 : 64.1	18 : 40.2
	Guard	AV	73.9	4.7	44.4	16.1	288.4	72.5	38.4
		R	57.1 : 88.7	3.2 : 5.53	32.4 : 54.7	9 : 28.5	259.6 : 313.9	60.1 : 86.7	28.5 : 48.5
	Play	AV	9.3	70.2	35.5	108.7	177.7	72.2	119.9
		R	6 : 12.4	58.4 : 81.6	12.7 : 65.5	65.5 : 146.5	158.2 : 189.4	63.5 : 81.6	60.4 : 152.4

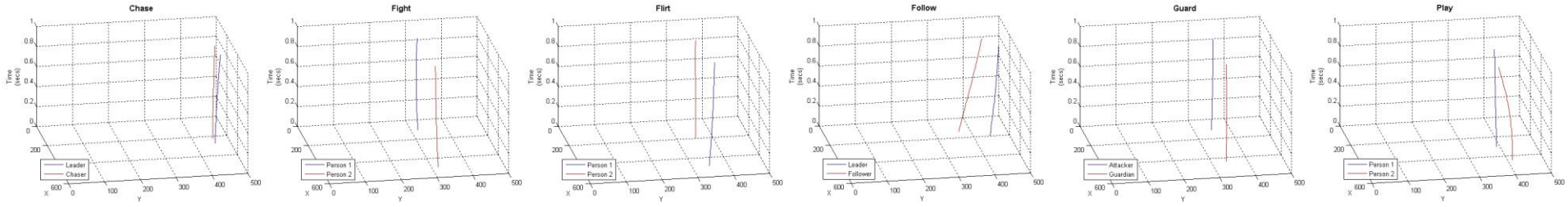

Figure 5.3.1: 1 Second Displays - Tables showing the Average values (AV) and Range (R) of various motion properties that are thought to be important factors in determining the intention between two agents. Distance is recorded in pixels, Speed in pixels/sec, and Acceleration in pixels/sec/sec. 100 pixels equates to approximately 1 metre. In the displays of Chasing and Following, Person 2 is always the person being Chased or Followed. In the Guard display, Person 2 is always Guarding. Graphs below table show the motion trajectories of both agents in each display.

5 Second			Speed P1	Speed P2	Acceleration P1	Acceleration P2	Rel. Dist.	Rel. Speed	Rel. Acc.
Overhead	Chase	AV	97.1	143.5	81.6	127.9	118.9	93.2	89.7
		R	45.6 : 124	93.1 : 182.1	45 : 108.4	81.5 : 162.5	69.8 : 177.8	40.1 : 127.8	54.7 : 146.5
	Fight	AV	55.3	60.9	62.5	76.6	171.5	115.1	130.7
		R	7.3 : 81.3	20.3 : 86	20.1 : 100.6	45 : 117	111.4 : 347.4	40.9 : 168.5	72.6 : 180.2
	Flirt	AV	4.1	81.3	12.3	35.0	138.1	78.0	35.3
		R	1.8 : 6	66.9 : 92	0 : 20.1	12.7 : 68.5	64.1 : 261	65.7 : 88.9	12.7 : 70.3
	Follow	AV	82.4	93.1	38.3	53.4	107.6	52.5	41.3
		R	76.4 : 89.7	85.8 : 102.4	25.5 : 52.5	20.1 : 108.7	95.6 : 113.7	34.3 : 80	20.1 : 94
	Guard	AV	56.3	8.0	53.5	19.9	171.1	51.3	55.3
		R	24 : 87.7	3.8 : 37.9	25.5 : 92.6	0 : 63.6	122.9 : 286.8	24 : 86.9	27 : 92.7
	Play	AV	26.6	108.2	42.1	105.8	184.0	126.4	117.4
		R	9.8 : 50.2	71.2 : 126.1	18 : 72	45.9 : 147.6	167.2 : 200	73.1 : 169.6	81.5 : 167.4

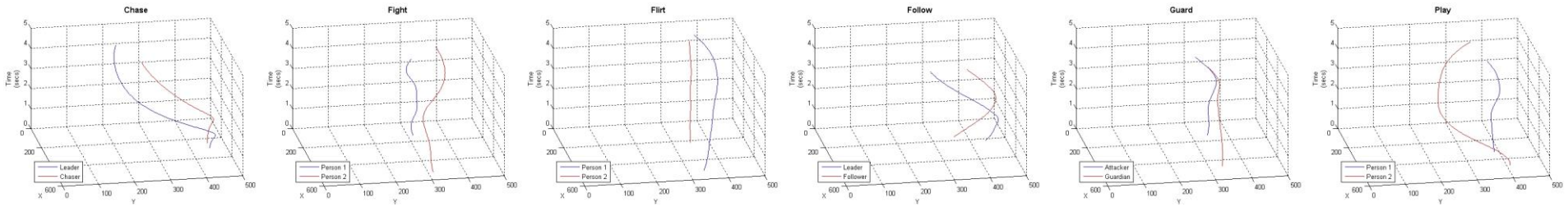

Figure 5.3.2: 5 Second Displays - Tables showing the Average values (AV) and Range (R) of various motion properties that are thought to be important factors in determining the intention between two agents. Distance is recorded in pixels, Speed in pixels/sec, and Acceleration in pixels/sec/sec. 100 pixels equates to approximately 1 metre. In the displays of Chasing and Following, Person 2 is always the person being Chased or Followed. In the Guard display, Person 2 is always Guarding. Graphs below table show the motion trajectories of both agents in each display.

10 Second			Speed P1	Speed P2	Acceleration P1	Acceleration P2	Rel. Dist.	Rel. Speed	Rel. Acc.
Overhead	Chase	AV	105.1	137.4	71.4	125.2	141.7	113.2	103.4
		R	81.1 : 126.6	125.1 : 168.4	50.9 : 81.5	63 : 165.5	97.7 : 182.7	52 : 156.2	65.5 : 157.2
	Fight	AV	58.4	59.0	53.7	101.1	159.6	106.1	173.0
		R	11.5 : 109.1	18.4 : 96.9	32.4 : 72.6	38.2 : 240.2	112.8 : 272.5	38.8 : 160.2	76.9 : 330.9
	Flirt	AV	4.3	78.5	8.0	56.1	98.9	74.8	54.9
		R	2.1 : 6.2	68.5 : 91	0 : 9	20.1 : 102.6	66 : 229.4	65.7 : 88.5	18 : 100.6
	Follow	AV	82.4	92.4	33.3	63.4	109.1	63.7	67.5
		R	65.1 : 91	78.3 : 104	25.5 : 40.2	27 : 117	80.8 : 130.4	30.5 : 102	20.1 : 173.1
	Guard	AV	46.0	17.3	40.3	42.2	135.2	45.4	64.9
		R	17.9 : 78.7	4.2 : 44.7	25.5 : 52.5	9 : 97.3	103.9 : 244.8	17.5 : 81	25.5 : 149.9
	Play	AV	32.5	106.4	38.4	123.4	176.9	129.6	156.5
		R	11.1 : 54.7	68.8 : 127.5	25.5 : 52.5	60.4 : 209.9	136.3 : 196.7	68.3 : 171.9	90 : 298.2

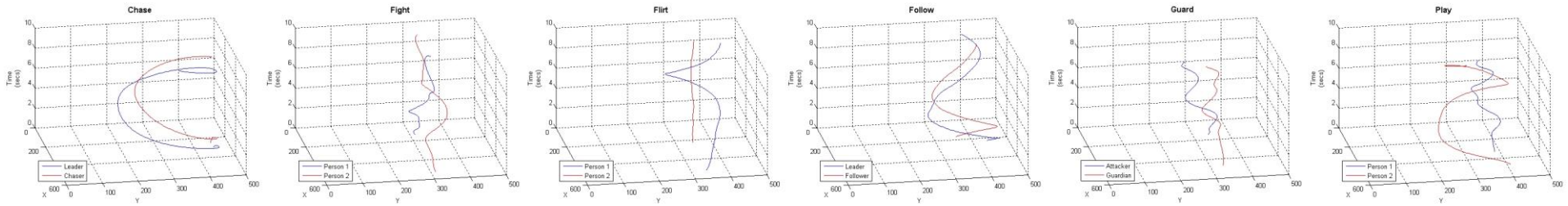

Figure 5.3.3: 10 Second Displays - Tables showing the Average values (AV) and Range (R) of various motion properties that are thought to be important factors in determining the intention between two agents. Distance is recorded in pixels, Speed in pixels/sec, and Acceleration in pixels/sec/sec. 100 pixels equates to approximately 1 metre. In the displays of Chasing and Following, Person 2 is always the person being Chased or Followed. In the Guard display, Person 2 is always Guarding. Graphs below table show the motion trajectories of both agents in each display.

Across time duration, the movement of agents in the 1 second displays is sparse compared to the longer displays. The motion trajectories of the 1 second displays all appear as reasonably straight vertical lines indicating that the movement by each agent was minimal, with only slight deviations forward/backward or to the side. The trajectories of the 5 second displays show more structured movement, with the agents beginning to move more: either towards each other, as in Flirt, Guard and Fight, or away from each other, as in Chase, Follow and Play. The circular trajectories of Chase and Follow, seen in the full duration displays are evident after 5 seconds. Furthermore, we see some movement by one agent towards the other in Guard and Flirt, and of both agents towards each other in Fight. The trajectory of Play at 5 seconds is similar to Chase and Follow except that only one of the agents appears to be taking a circular route, with the second agent taking a straighter route. The behavioural data shows that subjects are significantly better at judging intentions after 5 seconds than after 1 second. A subjective comparison of the motion trajectories at these durations, highlights the increase in information in regards to changes in location. The 5 second displays contain more information regarding agents' direction and route, accounting for the greater ability to discriminate between intentions. However, similarities between displays exist at 5 seconds, partly explaining the confusions witnessed: e.g. Chase and Follow; Fight, Guard and Play.

The 10 second displays' motion trajectories contain more information helping to clarify some of the confusions. Fight and Flirt are no longer confused: the trajectories now show one agent being circled in the Flirt display, and both agents moving towards each other in the Fight display. Confusion between

Chase and Follow has also become reduced, though they have similar trajectories: differences are therefore expected in the motion properties regarding speed and acceleration that indicate the agents staying together, or getting further apart. We suggested that an increase in information in the motion trajectories over time, from 1 second to 5 second, enabled the increase of intention recognition. However, no great increase in accuracy is found between 5 second and 10 second displays. It is proposed that the changes in motion trajectories between these durations only allow for some disentanglement of intentions, and that the motion properties may still be conflicting, allowing for the remaining confusions. The slight reduction in ability between viewing 10 second displays and the full duration displays may be due to movements later in the displays that mislead the participants. The set of displays analysed are limited, with only one example of each intention, and it is possible that movements within these displays are not what observers would associate with that intention.

In the motion properties of these short duration displays, we expect to find useful information in the 1 second display, and that this information becomes clearer as duration increases. We examined properties including the speed and acceleration of agents, and the relative speed, acceleration and distance between the agents. At 1 second, the motion properties show greater variation than in the full duration displays, with agents in the same display having markedly different average speeds and accelerations. For example, in Chase, person 2 has a higher average speed and acceleration than person 1, though this may be because person 2 is trying to flee and person 1 has yet to react. We see clear evidence in the Guard and Flirt of one agent doing the majority of the movement. This is also the case in Play, however this is

inconsistent with the remainder of the display, and may be cause of some confusion. In Fight, after 1 second, both agents are moving at approximately the same speed, though the relative distance is very high, suggesting that they are not moving towards each other yet. Looking at solely the motion properties of these 1 second displays does not reveal many differences that participants may be using to differentiate intentions. It is likely that a combination of the motion properties and trajectories enables participants to discriminate the intentions. It is perhaps testament to the development of the visual system that participants can discriminate these intentions over short durations, though it is too soon to rule out the unknown effects of cognitive reasoning: i.e. participants may be able to recognise a couple of intentions such as follow and play, and discriminate others by default such as chase and fight.

Increasing duration to 5 and 10 seconds, we see clearer patterns in the motion properties, similar to the full duration displays. In Follow, agents maintain their speed and distance from each other, where as in Chase, agents have a slightly faster average speed, more acceleration, and greater fluctuation in their relative distance. Furthermore, in Fight the relative distance is small, and average speeds are low, but with moments of high acceleration. At short durations, Flirt and Guard have similar motion properties, with one agent having a low average speed and the other having a high one, coupled with low accelerations in both agents. Therefore, these two intentions are likely discriminated by the motion trajectories showing how the two agents are interacting. Finally in Play, there are large differences between the agents' speeds and accelerations, with one agent moving and reacting faster. These large

ranges, coupled with the agents moving differently, may explain some of the confusion as to what intention is to be perceived in this display.

Figure 5.3.4 shows the distance travelled by each agent, in all intentions for all time duration.

Distance Travelled				
1 Second		P1	P2	DIFF (P2 > P1) (%)
Overhead	Chase	45.0	86.2	91.7
	Fight	56.6	39.4	-30.5
	Flirt	3.1	64.1	1946.9
	Follow	78.3	90.1	15.0
	Guard	71.4	4.6	-93.6
	Play	9.0	67.9	651.7

Distance Travelled				
5 Second		P1	P2	DIFF (P2 > P1) (%)
Overhead	Chase	482.1	712.9	47.9
	Fight	274.6	302.5	10.1
	Flirt	20.5	404.0	1869.5
	Follow	409.1	462.2	13.0
	Guard	279.5	39.9	-85.7
	Play	132.1	537.5	306.8

Distance Travelled				
10 Second		P1	P2	DIFF (P2 > P1) (%)
Overhead	Chase	1047.8	1369.1	30.7
	Fight	582.2	588.3	1.1
	Flirt	43.1	782.7	1715.1
	Follow	821.7	920.4	12.0
	Guard	458.5	172.6	-62.3
	Play	324.1	1060.2	227.1

Figure 5.3.4: Distance travelled in pixels by each agent in the six intentions, across 3 time durations: *Top:* 1 Second; *Middle:* 5 Second; *Bottom:* 10 Second. In Chase and Follow, P2 is the person being Chased or Followed. In Guard, P2 is Guarding. DIFF represents the distance travelled more by P2 than P1, as a percentage of P1's distance.

Figure 5.3.4 clearly shows the lack of movement by each agent in the 1 second duration displays, highlighting the importance of the direction of motion, and the speed. The distance covered by each agent greatly increases as time goes on, except in Flirt and Guard, where one agent does not move a great deal, in accordance with the instructions given. There is a large difference between the distances covered by Person 2 in the Play intention compared to that covered by Person 1. In the behavioural experiments of this thesis, and of previous research (Barrett et al, 2005; Blythe et al, 1999) a bias towards Play was noted. It is possible that Play is characterised by turn taking, so at times one person is moving more than the other. However, previous researchers did not record such differences between the agents in play, and therefore, the differences found in the current analysis could again show the importance of having multiple data sets when performing such research. This does not detract from the comprehensive analysis of the motion properties and trajectories carried out in this chapter. We have found patterns in the motion trajectories and the motion properties that may indicate, when taken together, how participants discriminate intentions. Furthermore, it has been shown, wherever possible, proposed causes of confusions between intentions, such as when motion paths look similar, or where agents in different intentions move at similar speeds.

5.4 A Stepwise Regression to fit Motion Properties to Intentions

The preceding analysis provides insight into what information is available in the animacy displays for participants to base their judgements of intention upon. However, the conclusions drawn from this analysis have relied on subjective impressions based on pre-determined properties of the motion. An

objective approach would be to employ a general regression model, or to train a neural network, to establish what motion properties drive the perceptions of intent. Ideally, implementation of these approaches requires multiple instances of each intention, from both viewpoint, and unfortunately this data set only contains one example of each intention. What follows, is therefore, an attempt to use a forward stepwise regression model to explain the behavioural data. This preliminary analysis will help to determine if the properties that we and other authors have examined, are indeed appropriate for the classification of intentions in animacy displays.

Previous analyses have shown velocity and distance, both absolute and relative, are important in classifying intentions within animacy displays (Blythe et al, 1999). The following analysis is similar to that applied to the motion properties involved in event segmentation by Zacks (2004)[7], and was performed with the kind assistance of Dr. Jeffrey Zacks. Various predictor variables were established: the position, speed and acceleration of each agent were calculated, as well as the relative speed, acceleration, distance and position. Maximum and minimum values were calculated for the acceleration and relative acceleration of each agent. For all of the mentioned parameters, the mean and standard deviations (SD) were calculated, and normalized to obtain scale-free parameters. This was performed for all six intentions (full duration) at both viewpoints. As a result of this we obtain a total of 26 predictors, which can be broken into four categories for ease of viewing:

[7] Thanks to Dr. Jeff Zacks, Washington University in St. Louis, for his assistance with this analysis.

1. Position Parameters: X1Mean; Y1Mean; X2Mean; Y2Mean; X1SD; Y1SD; X2SD; Y2SD; RelXMean; RelYMean; RelXSD; RelYSD

2. (Relative) Distance Parameters: DistanceMean; DistanceSD

3. Speed Parameters: Speed1Mean; Speed1SD; Speed2Mean; Speed2SD

4. Relative Speed & Acceleration Parameters: RelSpeedMean; RelSpeedSD; RelAccMean; RelAccSD

These predictors were entered into a forward stepwise regression paradigm continually until changes in the Akaike Information Criterion (AIC) (Akaike, 1974) were no longer significant, i.e., the addition of further predictors into the model did not aid the model in explaining more variance, as the model is saturated. The behavioural data (dependant variable) for the regression model were the percentage of participants that identified a display as a specific intention in Experiment 4.3, e.g. how many times any display was called chase. Data is collapsed across presentation conditions: whether occlusion or contextual cues were present, as these made only small differences to the ability to recognise intentions. In brief, the regression paradigm works by calculating which predictor explains the most variance by itself and enters that into the regression equation. The paradigm then compares the effect of adding the other predictors to the first predictor, adding to the equation the predictor that when combined with the first parameter gives the greatest level of explained variance. This procedure continues until the model is saturated, using the fewest number of predictors. As a general rule of thumb, predictors added into the regression model higher than the fourth position explain very little variance (Howell, 1997). Figures 5.4.1 and 5.4.2 show the first 5 predictors in the regression models, in the order that they are entered, for Overhead and Side View displays respectively.

The + and – indicates whether the correlation between the predictor and the categorisation data was positive or negative, and the values in parenthesis indicate the Beta value for that predictor in the model.

Full Duration	Overhead
Chase	+Speed1Mean (+22.46) -RelAccMean (-11.56) -X1Mean (-4.33) -RelAccSD (-1.94) -Y1Mean (-0.393)
Fight	+RelSpeedSD (+31.79) -Speed1SD (-11.05) +Y2Mean (+6.98) -RelSpeedMean (-2.63) +X1Mean (+1.09)
Flirt	-X1Mean (-22.44) -RelYMean (-14.44) -Y2SD (-10.42) -RelSpeedMean (-3.34) +Y1Mean (+2.95)
Follow	-DistanceMean (-27.05) -RelSpeedSD (-15.64) +RelYSD (+3.62) +X1SD (+1.19) +X1Mean (+0.13)
Guard	-RelXMean (-19.22) -Speed2Mean (-8.49) -Y1Mean (-6.1) -RelSpeedMean (-1.39) -X1Mean (-0.6)
Play	+RelSpeedSD (+32.41) +Y2Mean (+19.45) +RelAccMean (+8.71) +Y1SD (+2.81) -X1Mean (-0.7)

Figure 5.4.1: Table showing the 1st 5 predictors entered into a regression equation to explain the variance of the intention recognition data of Experiment 4.3, collapsing across presentation conditions, for Overhead displays. Column on left indicates the intention shown; column on right indicates parameters. + and – indicate whether the correlation between the parameter and behavioural data was positive or negative. Values in parenthesis indicate Beta values

Full Duration	Side View
Chase	+Speed2Mean (+8.56) -X1Mean (-5.68) +Speed1Mean (+4.59) -Y2Mean (-1.28) +Y1Mean (+ 0.31)
Fight	-X1SD (-33.03) +RelXMean (+16.96) +RelSpeedMean (+3.9) -RelYMean (-1.97) +X1Mean (+0.05)
Flirt	-Speed1Mean (-20.9) -X1Mean (-12.05) +Speed1SD (+4.46) +Y2Mean (+4.4) +Y1Mean (+0.18)
Follow	+Y2SD (+46.69) +Y1Mean (+11.68) +Speed2SD (+64.19) -Speed1SD (-52.02) +X1Mean (+0.02)
Guard	-RelXMean (-10.9) +X2SD (+4.99) +Acc1SD (+56.92) -Acc1Mean (-55.34) -X1Mean (-0.05)
Play	+RelAccMean (+5.97) -RelYSD (-4.83) -RelYMean (-2.27) -Speed1Mean (-0.52) +X1Mean (+0.04)

Figure 5.4.2: Table showing the 1st 5 predictors entered into a regression equation to explain the variance of the intention recognition data of Experiment 4.3, collapsing across presentation conditions, for Side View displays. Column on left indicates the intention shown; column on right indicates parameters. + and – indicate whether the correlation between the parameter and behavioural data was positive or negative. Values in parenthesis indicate Beta values

Considering the 1st 3 predictors entered into the regression equation of each intention we get an understanding of which motion parameters best predict the successful recognition of intentions.

Chase in the Overhead tends to be successfully recognised if person 1 is moving fast, i.e. the person chasing, and if the two people are not accelerating relative to each other. Accurate Fight classification is reliant on the Speed of person 1 showing little variation, and large variations in the speed of each agent increasing relatively to the other. The recognition of Flirt in the Overhead is reliant on positional predictors, namely the lateral position of person 1, not changing, i.e. the person who was stationary for the first moments of the displays. Follow is categorised by the mean distance between the two agents remaining low and the relative speed between the two agents not fluctuating. Guard, similar to Flirt, is determined by positional predictors, but also if the speed of the 2nd person remains low: in this case, this was the person who was Guarding. Finally Play would appear to be successfully recognised by large variations in the relative speed between the two agents, and by large relative accelerations.

Many of these descriptions would appear to make intuitive sense, such as Follow being characterised by small mean distances between the agents and few changes in the relative speed between the agents. Furthermore, these descriptions would appear to fit in with the subjective analysis of the motion properties mentioned earlier in the chapter.

Figure 5.4.2 shows the predictors that best predict the intentions when viewing the Side View displays. Similar to Chase in the Overhead, a Side View

display is likely to be called Chase if the speed of person 1 is large, i.e. the person chasing, and furthermore, if the speed of person 2 is also high, i.e., the person being Chased. Fight is categorised by small changes in the lateral position of person 1 and if agents' move at fast speeds relative to each other. Flirt is predicted in the Side View if the speed of person 1, who is almost stationary at the start, is low and if there are large changes in the speed of the same person. A successful categorisation of Follow is based on large changes in the horizontal position of the person being followed, person 2, and by changes in the horizontal position of person 1. Furthermore, Follow is predicted if there are large changes in the speed of person 2: the person being followed. Guard is characterised by changes in the lateral position of person 2, the person guarding, and by large changes in acceleration by person 1. Finally, Play in the Side View is characterised by large accelerations by each agent relative to the other and minimal changes in the horizontal dimension of the agents.

The results of the regression modelling for the Side View displays would appear to be based more on the positioning of the agents, with only a couple of displays being characterised by changes in speed or acceleration, such as Chase and Flirt. There are a couple of confusing predictors such as changes in the Y dimension: it is unclear as to why this predictor would have such an effect given that the changes in the vertical dimension are minimal in the Side View.

Overall, the stepwise regression modelling shows that in general the main predictors of the intentions within these animacy displays are speed and acceleration, both absolute and relatively, with distance and positional predictors also contributing. The results of the current research are broadly consistent with previous findings (Barrett et al, 2005; Blythe et al, 1999; Tremoulet and

Feldman, 2000; Zacks, 2004). However, we find that the speed of an agent is more predictive than changes in acceleration.

Chapter Summary

Chapter 4 showed that participants were capable of recognising human intentions from the movement of agents in animacy displays, derived from video recordings of human actors. Six intentions were studied, taken from previous research by Blythe et al (1999) that suggest that these intentions were generic to human motion. These intentions were Chase, Fight, Flirt, Follow, Guard and Play, and were examined across two viewpoints, Overhead and Side View. Participants were better at recognising intentions when the motion was viewed from above the actors. Confusions were found between displays, such as Chase being called a Follow, and a bias towards Play. Furthermore, greater levels of confusion were witnessed in the Side View displays. The aim of this chapter was to offer an analysis of the movements' trajectories and motion properties, to give a better understanding of the displays, highlighting what factors subject may use to differentiate intentions.

In the motion trajectories, patterns were found in the intentions such as the agents in Chase and Follow taking circular motion paths, and those in Fight and Guard taking more direct routes. The agents in Flirt and Play seemed to contain both moments of circular paths and direct paths: Play having much more variation, whereas Flirt contained extended moments of no movement by agents in a reciprocal manner. The large variation and range of movements within Play is hypothesised to contribute to the confusion/bias towards Play, in that the large fluctuations in the movement could make the display appear to look like another.

These descriptions of intentions are based on a subjective analysis of the Overhead displays. The equivalent intentions in the Side View displays generally appear as agents moving in straight, direct lines. The lack of variation in the Side View displays is thought to contribute to the displays being harder to differentiate in the Side View, thus reduced overall ability to correctly discriminate the displays by participants. The motion paths of the agents in Fight, however, appeared very similar in both viewpoints. Ability to recognise this intention was similar in both viewpoints, suggesting no overhead advantage. The Overhead displays have the advantage of showing clearly the distance between two agents, and give better understanding of moments when, in the Side View, the agents are masking each other.

Motion properties were analysed, relating to the speed and acceleration of the agents. In addition, the distance between the two agents and the overall distance covered, were examined. The mean values and ranges of these properties were calculated. In the Side View displays we found few differences that would assist differentiation of one intention from another. The ranges of each display were very large, with means approximately in the middle, suggesting that the agents moved at a large variety of speeds. Furthermore, large accelerations were witnessed which were suggested to have come from problems with the method of extracting the positional co-ordinates from Eyesweb. The motion properties of the Overhead displays showed more variation between displays. It was shown that the agents in the Chase display maintained high speeds, with the distance between the agents changing, as the chasing actor got closer to the actor being chased. The agents in the Follow display appeared to be maintaining constant speeds as well as keeping the distance between each other

constant. The speed of each agent in Fight was low without much fluctuation, but there were moments of high acceleration when the agents came together. Guard and Flirt both seemed to have one agent moving at a higher speed than the other, with the slower agent moving less overall. Finally, it was shown that the motion properties of the agents within Play seemed to vary more, which is possibly why, in conjunction with varying motion paths, Play could look like other displays, and vice versa. Overall, it is suggested that differences between motion properties between intentions allows for better intention recognition in the Overhead display, and that reduced variation between displays leads to more confusion in the Side View displays. It is also proposed that when motion properties appear similar across intentions, such as in the case of Flirt and Guard, then participants use the motion trajectories for successful discrimination.

The motion trajectories and properties were analysed for short duration displays to examine how the available information changes over time. The behavioural data and experimental stimuli from Experiment 4.6 were analysed. At 1 second duration, motion trajectories did not give much information, as the actors only moved a couple of paces. The motion properties showed large variation within each intention, in terms of speed and acceleration of the agents. This lack of information would explain why there are large levels of confusion between intentions at this time duration. At 5 second duration, the motion trajectories carried more information as to the motion paths of the agents. These paths matched, to an extent, those seen in the full duration displays, in terms of whether the agents took circular routes or more direct paths. The motion properties in the 5 second displays showed values and ranges that were similar to those of the full duration displays. The analysis of the 10 second displays was

similar to both the 5 second displays and the full duration displays. It is proposed that participants can recognise the intentions within a very early time frame and that extended displays can be problematic in that they may contain features similar to other intentions, which may cause participants to rethink their initial, accurate judgement. The ability to recognise intentions rapidly does make ecological sense, and perhaps a further line of research could be proposed to explore the optimal duration for the accurate recognition of intent.

A regression analysis was performed to explore what predictors indicate each intention. A large number of predictors were calculated, including motion, positional and distance predictors: other predictors used by previous authors, such as changes in directions, angle of movement, could have been but were not explored. A forward stepwise regression paradigm was used, entering into the equation the best combination of predictors that would saturate the model using the fewest number of predictors. Overhead displays were predicted by the speed of individual agents, and the speed of agents relative to each other. Side View displays were predicted predominantly by changes in positional predictors. The predictors of the Overhead displays make more ecological sense than those of the Side View displays, and are broadly consistent with previous findings that suggest motion and distance predictors to be best predictors of intentional motion. However, whereas changes in acceleration have been found to be the main predictors (Blythe et al, 1999), the above analysis would suggest that speed is a better predictor of intentions.

It is unclear if this difference in dominance of predictors, i.e. from those involving acceleration to those involving speed, is due to the current displays being derived from actual human motion. It is possible that this subtle change is

because of the difference in methods from the creation of previous animacy displays, i.e. via computer animation, and the creation of the current animacy displays, derived from video recordings of human motion. However, it must be remembered that the analysis of this chapter is based on a concise data set, and to obtain a more conclusive set of results would require the recording and production of numerous examples of each intention by various actors. Furthermore, it is difficult to distinguish if the results of the above analysis are unbiased, or have been influenced by prior knowledge of the instructions given to actors. This question would support the filming of movements where actors have only been told the intention to act, and given no further instructions on how to perform it.

This chapter solely focuses on the data from Chapter 4, as the displays used make up a large proportion of the experiments of this thesis. This in depth understanding of the motion properties involved is useful for interpreting the results of the previous chapter, and the following chapter, looking at the perception of these displays by people with Autism Spectrum Disorders (ASDs). The same analysis could be performed on the displays used in Chapter 3, however the large variation in intention within the displays and of movements would have to be reduced. An alternative idea for a follow-up study would be to use the narrow set of intentions discussed above, and obtain ratings of animacy for them. Using the animacy ratings as the dependant variable would allow us to obtain the predictors of how people determine that these agents within the displays are animate.

In conclusion, a comprehensive study of the motion properties and trajectories of the animacy stimuli of Chapters 4 and 6 was performed.

Furthermore, a regression analysis was performed indicating that speed and acceleration cues are predictors of intentions in Overhead displays, and positional and speed cues are predictors of intentions in Side View displays. Conclusions are tentative due to the concise set of experimental stimuli, yet the analysis serves to give a good understanding of each display, and to give indication of what motion predictors drive the perception of intention when viewing displays of human interactions.

The co-ordinates and movies used in this analysis are available at

www.psy.gla.ac.uk/~phil/co-ordinates.html

www.psy.gla.ac.uk/~phil/movies.html

Chapter 6 – Animacy and Autistic Spectrum Disorders

Introduction and chapter overview

An important aspect of research is highlighting the relevance of a laboratory study to a wider population. The final chapter of this thesis presents a single experiment, which aims to show the application and value of animacy displays derived from human motion in furthering the research and understanding of Autistic Spectrum Disorders (ASD).

Autism, first diagnosed by Kranner (1943) and Asperger (1944), is a developmental disorder that is diagnosed when an individual is shown to have abnormalities in a triad of behavioural domains: social development, social communication, and repetitive behaviours and obsessive interests (APA, 1994; WHO, 1994; Wing and Gould, 1979). The prevalence of autism is currently estimated to range from between 0.3% to 0.7% of the population, with a male to female ratio of three to one (Frith, 2001). The underlying basis of the condition is still unknown but various theories have been put forward ranging from perceptual and cognitive problems in understanding the thoughts and intentions of others (Baron-Cohen, 1995; Baron-Cohen, Wheelwright, Griffin, Lawson and Hill, 2002; Frith, 1989) to neurological abnormalities in various areas of the brain (Abell, Krams, Ashburner, Passingham, Friston, Frackowiak, Happé, Frith and Frith, 1999; Bachevalier, 1991; Baron-Cohen and Belmonte, 2005; Bauman and Kemper, 1994; Brothers, 1990; Frith and Frith, 2001). Certain main theories of Autistic Spectrum Disorders are discussed below.

The Empathizing-Systemizing (E-S) theory (Baron-Cohen, 2002) is a two-fold theory that proposes that a disadvantage in empathizing may explain

social and communication problems, and a superiority in systemizing may explain traits such as repetitive behaviours, obsessive interests and islets of abilities (Baron-Cohen, Wheelwright, Lawson, Griffin, Ashwin, Billington and Chakrabarti, 2005). Baron-Cohen et al (2005) state that empathizing involves both the ability to attribute appropriate intentions and thoughts to another, and to be able to react emotionally in a way that is appropriate to another's actions or feelings. The ability to attribute mental states to explain and predict the behaviour of others is referred to as having a "Theory of Mind" (ToM), (Premack and Woodruff, 1978), and impairment in this ability in people with ASD has been shown in various tasks, including false-belief tasks (Baron-Cohen, Leslie and Frith, 1985) and picture sequencing tasks (Baron-Cohen, Leslie and Frith, 1986). False-belief tasks are characterised by participants being shown a series of events involving two or three characters and being asked questions such as what would person one think person two is thinking, i.e. a first order false-belief task (Baron-Cohen et al, 1985), or what would person one think person two thinks person three thinks, i.e. a second order false-belief task (Baron-Cohen, 1989). Picture sequencing tasks involve participants sorting picture cards depicting characters, into an order that tells a cohesive story. Baron-Cohen et al (1985, 1986) concluded that people with ASD showed cognitive deficits in a ToM capacity, irrespective of general intelligence, thus not being able to infer the appropriate intentions and thoughts of the characters in the displays they were shown. Furthermore, in ToM tasks, people with ASD were shown to be more inclined to use causal terms or simple descriptive terms of behaviour, suggesting one person caused the action, i.e. not an interaction, as opposed to neurotypicals who are more inclined to use terms suggesting mental states and desires.

The Systemizing part of the E-S theory is described as an underlying need or desire to understand the rules that govern a system, i.e. anything that follows the sequence of input-operation-output (Baron-Cohen, 2002). Systemizing is thought to explain why some individuals with autism are seen to have special abilities in maths, music, memory, etc. (Baron-Cohen and Bolton, 1993). Baron-Cohen (2002) argues that Empathizing and Systemizing are two separate dimensions of the brain with empathizing being more prominent in females and systemizing being more prominent in males. The argument is therefore that autism can be thought of as a bias towards the male dimension of the brain and the E-S theory is sometimes referred to as the Extreme Male Brain theory (Asperger, 1944; Baron-Cohen, 2002).

A bias to systemizing is also a foundation of the Weak Central Coherence (WCC) theory (Frith 1989) which suggests that individuals with ASD have an inability to bind local details of a scene or situation into a global percept, indicated by their attention to finer details and disregard for the whole. The WCC theory is supported by research that has shown people with autism to be superior in an Embedded Figures task (Witkin, Oltman, Raskin and Karp, 1971) where participants have to find a shape within a global picture (Jolliffe and Baron-Cohen, 1997), and to be disadvantaged in a Hooper Visual Organisation test (Hooper, 1983), where participants have to re-organise a shape that has been cut into pieces, resembling a jigsaw without the fitted edges (Jolliffe and Baron-Cohen, 2001). Furthermore, Happé (1996) showed that children with autism are not as susceptible to visual illusions as children without autism and Hill and Frith (2003) showed evidence of people with autism only being able to recall specific words in a story and not the gist, where as typically developing people will be

more likely to recall the gist of the story. This evidence taken together would support the theory that people with ASD appear to be more inclined to focus on finer aspects of objects and tasks, and do not establish a global understanding. A difference between the role of systemising in the E-S theory and the WCC theory is that it in the E-S theory, systemizing makes people with autism better at abilities where good knowledge of the fine rules is beneficial, where as in the WCC theory the focus on systemizing and the finer details will prevent an individual with autism from grasping the whole system, thus never fully understanding the system or becoming good at it. It is proposed that this distinction has, as yet, never fully been examined (Baron-Cohen et al, 2005).

Both the Empathising-Systemizing theory and the Weak Central Coherence theory address cognitive or perceptual differences between people with ASD and typically developing people, however there is also evidence to suggest that there are differences at a neurological level. The Social Brain theory (Brothers, 1990) suggests a neurological deficit in areas of the brain that have been shown to be involved in social understanding, for example, the amygdala, the orbito-frontal cortex, the medial frontal cortex and the superior temporal sulcus and gyrus (Castelli et al, 2000; Frith and Frith, 2001). Support for the Social Brain theory comes from reports that have shown increased cell-density, but not increased volume, in the amygdala, thought to be involved in emotion-recognition, in post-mortem studies in autism (Bauman and Kemper, 1994; Rapin and Katzman, 1998). Furthermore, support for this theory comes from animal lesion studies in the amygdala of rhesus monkeys where the monkeys become poor at social integration and are reported to show characteristics similar to people with ASD (Bachevalier, 1991).

Additional evidence for a neurological basis of ASD comes from neuro-imaging studies that have made use of animacy displays similar to the original Heider and Simmel movie, which show differences in brain activity between people with ASD and neurotypicals. Castelli, Frith, Happé and Frith (2002) instructed participants to watch animacy displays of triangles performing (a) random motion, e.g., bouncing, (b) goal-directed motion, e.g., chasing, fighting, and (c) Theory of Mind motion, e.g., deception, coaxing, whilst they performed Positron Emission Tomography (PET) scans on the participants. In a previous study (Abell, Happé and Frith, 2000) it had been shown that in a free response task, people with ASD used fewer words and inferred less appropriate intentions when viewing the ToM displays than typically developed people. Castelli et al (2002) compared adults with autism to typically developed adults and again showed the use of fewer and less appropriate words in the free responses of the autistic population when instructed to describe the ToM displays. Results from the PET scan showed reduced activation in areas previously associated with social understanding (Castelli et al, 2000): in the superior temporal sulcus (STS) at the temporoparietal junction and in the medial prefrontal area. Castelli et al (2002) also showed reduced functional connectivity between the STS and the extrastriate cortex, though the extrastriate cortex showed no reduction in levels of activation when viewing the Theory of Mind displays, compared to the typically developed adults, as measured by cerebral blood flow. As the STS has previously been shown to be involved in the attribution of intention (Castelli et al, 2000) it was suggested that reduced connectivity may result in a lack of transfer of information resulting in a poor mentalizing of the agents in the displays.

Baron-Cohen and Belmonte (2005) offer a review of the main theories of autistic spectrum disorders as well as advancing the argument for a neurological basis for autism. Baron-Cohen and Belmonte suggest, after reviewing results of previous research, that a combination of sensory hyperarousal and abnormal attentional selectivity, along with increased connectivity in local neural networks and a reduction in connectivity between distal neural networks leads to irrelevant information being stored and a subsequent reduction of capacity for useful information. This combination of hyperactivity and inability to focus attention appropriately would be consistent with people with ASD misunderstanding a situation or action.

Highlighted in the above theories and research, an underlying aspect of autism is the reduced ability to attribute appropriate intentions to the actions of others. Abell et al (2000) and Castelli et al (2002) showed people with ASD to be poorer at attributing intentions to agents depicting actions where the movement of one agent is based on the thoughts of another, such as in displays showing coaxing, deception or mocking, with people with ASD generally describing these displays in fewer words and attributing wrong intentions. This chapter will look at further understanding the attribution of intentions by people with ASD, making use of animacy displays derived from human motion. Predominantly, animacy displays used in the study of autism have been cartoon representations of moving shapes and, similar to the animacy literature in general, there has been no use of animacy displays that have their basis in human motion. Therefore, this chapter will use animacy displays derived from human motion to explore if people with ASD are poor at judging the intentions in displays based on human motion, or whether it is an artefact of the cartoon

displays that prevents them from being able to judge the intention. This technique for animacy production also allows a direct comparison of participants' ability to judge intentions when seeing the original video recordings of the intentions and compare it to the ability when viewing the animacy representation. This would appear to be the first time that this comparison has been possible. Furthermore, this chapter will also make use of the ability to compare an intention from two viewpoints, as in previous experiments in this thesis, in order to test the effect of viewpoint of displays on the ability of people with ASD to judge intentions. The results from this chapter will be related to the theories of autism discussed above.

We will compare an ASD population with a neurotypical population on the ability to judge six intentions pinpointed by Blythe, Todd and Miller (1999) and Barrett, Miller, Todd and Blythe (2005): Chasing, Fighting, Flirting, Following, Guarding and Playing. We will compare abilities across two presentation conditions, altering the level of available visual information: the original footage displays, showing all available visual information, and the animacy displays derived from these recordings, showing only global motion. The third variable of the experiment will contrast ability to judge intentions across two viewpoints: an overhead and a side view. It is hypothesised that overall, compared to a neurotypical control population, people with ASD will be poorer at judging the correct intentions from all displays. Furthermore, this difference is expected to be greater in the animacy displays than in the original recordings, due to the reduction of available visual information. Finally, we propose that, in regards to viewpoint, in the original footage displays, there will be no difference in ability to judge intention across viewpoint within groups, and

in the animacy displays, we propose that the ASD population will perform in a manner similar to that found in previous experiments in this thesis involving neurotypicals: that ability to judge intention is reduced in side view displays.

METHODS

6.2 PARTICIPANTS

6 male participants, obtained via advertisements and word of mouth, were screened for High Functioning Autism (HFA). HFA was defined as people who have abnormalities in a triad of behavioural domains, i.e. social communication, social development, and repetitive behaviours and obsessions, and whose IQ is approximately 100 (Baron-Cohen, Wheelwright, Lawson, Griffin, Ashwin, Billington and Chakrabarti, 2005). Participants were administered with the Autistic Spectrum Quotient (AQ) (Baron-Cohen, Wheelwright, Skinner, Martin and Clubley, 2001) and the Social Responsiveness Scale (SRS) (Constantino, 2003; Constantino and Todd, 2002) to determine if they were suitable for a diagnostic interview by a clinician. If participants scored greater than 14 on the AQ and greater than 80 on the SRS, the clinician (J. Piggott) established whether the participants were suitable for the experiment using the Autism Diagnostic Interview (ADI) (Le Couter, Rutter, Lord, Rios, Robertson, Holdgrafer and McLennan, 1989). After this screening process it was determined that 5 of the 6 participants met the criteria for High Functioning Autism and were suitable for the experiment.

Five neurotypical male participants were selected from the University of Glasgow undergraduate population. All 10 participants were aged between 18

and 25, had normal or corrected to normal vision and received a monetary incentive for participation in the experiment.

The ASD population and control population were age-matched, and all fell within the range of 18 – 25 years old. The populations were unfortunately not IQ matched, however all participants had received some form of education at university level or equivalent.

6.3 Stimulus Production

The original footage from both viewpoints, side view and overhead, used to create the animacy displays in Experiment 4.3, were used in this experiment. Furthermore, the resultant animacy displays of this original footage, as used in display condition NONC (no occlusion, no context) of Experiment 4.3, were used. In the original footage displays, actors appeared dressed in white body suits and hoods on a black stage. In the animacy displays, white circles on an empty grey background represented the actors. Example frames from the overhead and side view, original footage and animacy displays can be seen in Figure 6.1 (Page 221).

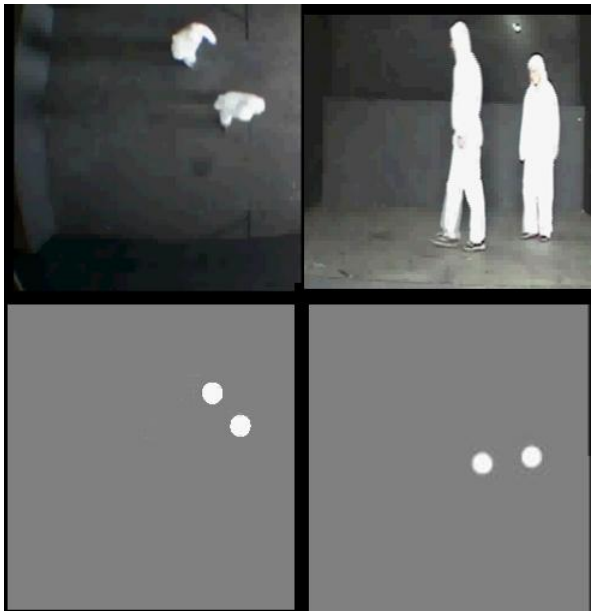

Figure 6.1: Examples of experimental displays showing the Original Footage on upper half of the diagram and Animacy displays on the lower half of the diagram, with Overhead displays on left of the diagram and Side View displays on the right of the diagram.

6.4 PROCEDURE

The experiment was run on a G4 Apple Macintosh (O.S. 9.2) using Matlab 5 (the Mathworks, Natick), Showtime (Watson and Hu, 1999) and the Psychophysics Toolbox Version 2.5 (Brainard, 1997; Pelli, 1997). Viewpoint was counterbalanced throughout the experiment with participants either seeing all the animacy displays first, followed by all the video displays, or vice versa; with participants being randomly allocated to an order. In total, participants saw 48 displays with a running time of approximately 30 minutes. Each experiment consisted of 48 trials split into 4 blocks of 12 trials, with 3 practice trials to familiarise the participants with the task. After each display participants used a six alternative forced choice to select the intention that best described the motion in the display.

The ASD group ran this experiment in conjunction with a battery of tests, not reported in this thesis that investigated the perception of biological motion and affect in people with ASD (McKay, Mackie, Piggott, Simmons & Pollick,

2006). In total, the ASD population ran four experiments (present experiment and three others) with a runtime of approximately 3 hours, though a proportion of this time was used as breaks. Neurotypicals only ran the one experiment described above.

Participants were sat at a distance of approximately 1m, giving a starting visual arc of each circle in both viewpoints of 0.5', and a visual arc of the actors of 1.1' in the Overhead video displays, and of 2.9' in the Side View video displays. The luminance of the white circles in both viewpoints of the animacy displays was 142 cd/m^2 and the luminance of the grey background was 43.3, giving a contrast value of approximately 0.5 (Michelson, 1927). The average luminance of the Overhead video displays was 60.6 cd/m^2, and the average luminance of the Side View video displays was 77.9 cd/m^2. Displays were presented at 30fps.

6.5 RESULTS

This experiment aimed to compare people with Autistic Spectrum Disorders with a neurotypical population on the ability to differentiate between intentions displayed in video recordings of human interactions, and in the animacy displays derived from these recordings. Ability to differentiate intentions was compared over the two viewpoints of overhead and side view. It was hypothesised that in general the ASD population would be poorer at distinguishing intentions between the displays compared to the neurptypicals, and that this would be consistent across viewpoints and presentation conditions: original recordings and animacy displays.

Although run as one experiment, for the purpose of clarity, graphs showing the results have been split into the two presentation conditions of animacy displays and original footage displays. These graphs, collapsed across intentions, can be seen in Figure 6.2 for the animacy displays and in Figure 6.3 for the original footage displays:

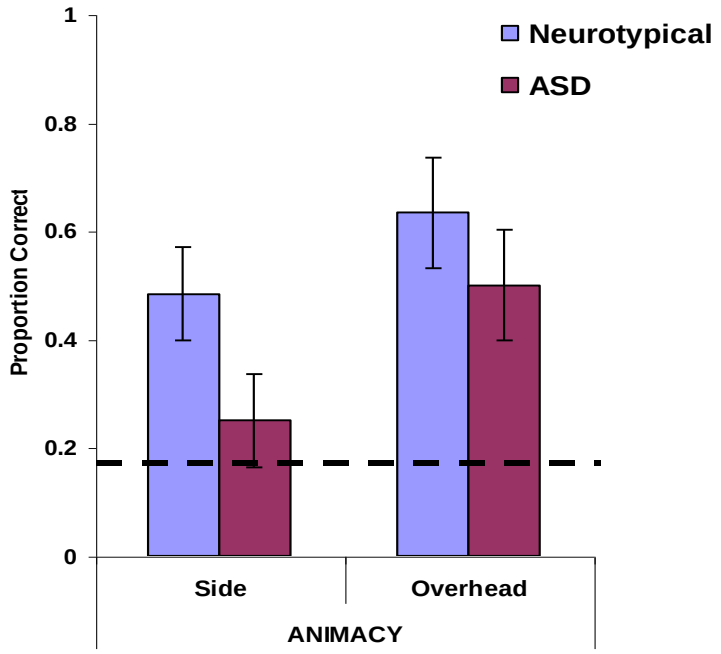

Figure 6.2: Proportion correct at categorising intentions in Animacy Displays derived from Human Motion for Neurotypical (blue) and ASD (red) groups. Error bars indicate standard error and bold dash line indicates chance: 0.1667

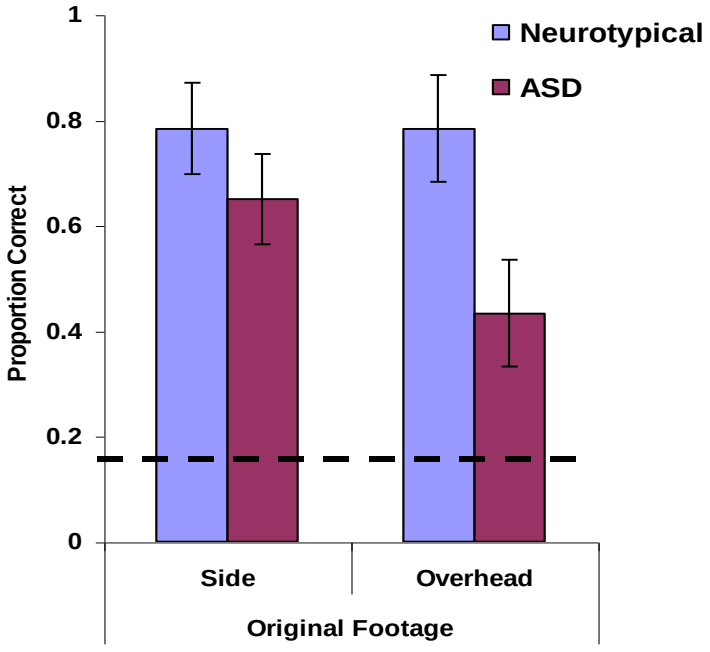

Figure 6.3: Proportion correct at categorising intentions in Original Footage Displays for Neurotypical (blue) and ASD (red) groups. Error bars indicate standard error and bold dash line indicates chance: 0.1667

The graphs appear to show a general reduced performance in ability to distinguish intentions in the ASD population, across viewpoint and presentation condition. In the animacy presentation condition, Figure 6.2, the ASD population appears able to distinguish intentions in both viewpoints at levels greater than chance, but to a degree reduced from that of the neurotypicals (25% for side view and 50% for the overhead view for the ASD population, and 48% and 63% respectively for the neurotypicals). Examination of the variable of viewpoint shows that the general pattern of results is the same in the ASD population as in the neurotypicals, with poorer performance on the side view than in the overhead view. This pattern is consistent with results shown throughout this thesis in which viewpoint has been a variable.

A different pattern is seen in the original footage presentation condition. Overall, both populations appear to have increased performances as would be expected with visual information being increased. However it now appears that where the neurotypicals are consistent in ability across viewpoint (78% for both side view and overhead), the ASD population appear to be better at judging intention in the side view than in the overhead view (65% and 43% for the side view and the overhead). This reversal of ability across viewpoint for the ASD group, from the animacy presentation to the original footage presentation (25% and 50% for side view and overhead in animacy displays and 65% and 43% for side view and overhead in original footage displays) would be expected to be shown in a significant three-way (possible four-way) interaction as the same trend is not found in the Neurotypical group.

These results were analysed using a four-way repeated measures, mixed design ANOVA. The variables included - Intention (Chasing, Fighting, Flirting,

Following, Guarding, Playing); Presentation Condition (animacy, original footage); Viewpoint (overhead, side); and Group (Neurotypical, ASD). For clarity of results we will present interactions first, from highest order to lowest, then present the main effects, and finally the confusion matrices of the intention categorisations. The ANOVA revealed no four-way interaction, nor the three-way interaction between Group, Viewpoint and Presentation Condition that we predicted from the graph, $F(1,8) = 4.66$, $\mathbf{p=0.063}$. The power for this interaction was fairly low (power = 0.476), and given the low number of participants in each population and the pattern of the graphs, post-hoc analysis was performed, using Tukey HSD. It was found that with more participants it is probable that we may find a significant difference in judging intentions between the ASD population and neurotypicals when shown the video displays from the overhead viewpoint. No other three-way interactions were significant.

The ANOVA revealed a two-way interaction between Group and Presentation Condition, $F(1,8) = 0.346$, $\mathbf{p<0.05}$. Post hoc analysis performed using Tukey HSD revealed that the Neurotypical group was better at judging intention in the original footage displays (78.3%) than the ASD group at judging intentions in the animacy displays (37.5%). In addition, the ANOVA revealed an interaction between Viewpoint and Intention, $F(5,40) = 2.82$, $\mathbf{p<0.05}$. Tukey HSD analysis revealed that participants are better at judging intention when the display shows a Following intention from the overhead position (95%) than overhead displays of Chasing (32.5%) and Playing (40%) and side view displays of Guarding (42.5%), Playing (40%) and Chasing (40%). Furthermore, the ANOVA showed an interaction between Viewpoint and Presentation Condition, $F(1,8) = 17.66$, $\mathbf{p<0.05}$, with Tukey HSD analysis determining that judgments of

intentions were significantly better in side views of the original footage displays (71.7%) compared to side views of the animacy displays (36.7%). Finally, an interaction between Presentation Condition and Intention, $F(5,40) = 3.13$, $\boldsymbol{p<0.05}$, which showed participants are better at judging intention from the video display of Following (100%) than the animacy displays of Playing (27.5%) and Chasing (22.5%). No other two-way interactions were significant.

The ANOVA showed no main effect of Viewpoint, $F(1,8) = 1.56$, $\boldsymbol{p>0.05}$, (54.1% for side view displays and 58.8% for overhead displays) nor a main effect of groups, $F(1,8) = 3.81$, $\boldsymbol{p>0.05}$, (67% and 46.7% for neurotypicals and the ASD population respectively). The ANOVA revealed a main effect of Presentation Condition, $F(1,8) = 15.61$, $\boldsymbol{p<0.05}$. Fischer's LSD analysis revealed that participants are better at judging intention overall when shown in original footage (66.3%) than when it is shown in animacy displays (46.7%). Finally, the ANOVA also revealed a significant main effect of Intention, $F(5,40) = 3.13$, $\boldsymbol{p<0.05}$, with Fischer's LSD analysis revealing participants to be better at categorizing the intentions of Following (77.5%), Fighting (68.9%) and Flirting (66.3%) Vs Playing (40%) and Chasing (36.3%), and at categorizing Following over Guarding (50%).

The confusion matrices for both populations can be seen in: Table 6.1 for Animacy displays viewed from the Side View; Table 6.2 for Animacy displays viewed from the Overhead view; Table 6.3 for Original Footage displays viewed from the Side View; and finally, Table 6.4 for Original Footage displays viewed from the Overhead.

		Animacy Displays - Side View					
		Response					
		Chase	Fight	Flirt	Follow	Guard	Play
Presented – Neurotypicals	Chase	0.4	0.1	0	0.3	0	0.2
	Fight	0	0.8	0	0	0.1	0.1
	Flirt	0	0.1	0.8	0	0	0.1
	Follow	0.1	0	0	0.2	0	0.7
	Guard	0	0.3	0	0	0.4	0.3
	Play	0	0.7	0	0	0	0.3
Presented – ASD	Chase	0	0	0.4	0.3	0	0.3
	Fight	0	0.5	0.1	0	0	0.4
	Flirt	0	0.1	0.5	0.2	0	0.2
	Follow	0.1	0	0.1	0.2	0.1	0.5
	Guard	0.1	0.3	0.1	0	0.2	0.2
	Play	0.1	0.6	0.2	0	0	0.1

Table 6.1: Confusion Matrices for Animacy Displays shown from the Side View for both experimental groups: Neurotypical group shown in upper half of table and ASD group shown in lower half of table. Presented displays are shown in rows with participant responses shown in columns.

		Animacy Displays – Overhead					
		Response					
		Chase	Fight	Flirt	Follow	Guard	Play
Presented – Neurotypicals	Chase	0.3	0	0	0.7	0	0
	Fight	0	0.5	0	0	0.3	0.2
	Flirt	0	0	0.8	0	0.1	0.1
	Follow	0.1	0	0	0.9	0	0
	Guard	0	0.1	0	0	0.9	0
	Play	0.1	0.2	0	0	0.3	0.4
Presented – ASD	Chase	0.2	0	0.1	0.7	0	0
	Fight	0	0.4	0	0.1	0.2	0.3
	Flirt	0	0	0.7	0	0.3	0
	Follow	0.1	0	0	0.9	0	0
	Guard	0	0.3	0	0	0.5	0.2
	Play	0.1	0.4	0	0	0.2	0.3

Table 6.2: Confusion Matrices for Animacy Displays shown from the Overhead for both experimental groups: Neurotypical group shown in upper half of table and ASD group shown in lower half of table. Presented displays are shown in rows with participant responses shown in columns.

		Original Footage Displays - Side View					
		Response					
		Chase	Fight	Flirt	Follow	Guard	Play
Presented – Neurotypicals	Chase	0.5	0	0	0.4	0	0.1
	Fight	0	0.9	0	0	0.1	0
	Flirt	0	0	1	0	0	0
	Follow	0	0	0	1	0	0
	Guard	0	0.3	0	0	0.7	0
	Play	0.2	0.2	0	0	0	0.6
Presented – ASD	Chase	0.7	0	0	0.3	0	0
	Fight	0	0.8	0	0	0	0.1
	Flirt	0	0	0.4	0.2	0	0.2
	Follow	0	0	0	1	0	0
	Guard	0	0.3	0	0	0.4	0.2
	Play	0.1	0.2	0	0	0.1	0.6

Table 6.3: Confusion Matrices for Original Footage Displays shown from the Side View for both experimental groups: Neurotypical group shown in upper half of table and ASD group shown in lower half of table. Presented displays are shown in rows with participant responses shown in columns.

		Original Footage Displays - Overhead					
		Response					
		Chase	Fight	Flirt	Follow	Guard	Play
Presented – Neurotypicals	Chase	0.6	0	0	0.4	0	0
	Fight	0	1	0	0	0	0
	Flirt	0	0	0.8	0	0	0.2
	Follow	0	0	0	1	0	0
	Guard	0	0.3	0	0	0.7	0
	Play	0	0.4	0	0	0	0.6
Presented – ASD	Chase	0.2	0	0	0.7	0	0
	Fight	0	0.6	0.1	0	0.1	0
	Flirt	0	0	0.3	0.1	0.3	0.2
	Follow	0	0	0	1	0	0
	Guard	0	0.5	0	0	0.2	0.2
	Play	0.2	0.3	0	0.1	0.1	0.3

Table 6.4: Confusion Matrices for Original Footage Displays shown from the Overhead for both experimental groups: Neurotypical group shown in upper half of table and ASD group shown in lower half of table. Presented displays are shown in rows with participant responses shown in columns.

The rows of the confusion matrices for the ASD population do not always sum to 1 as some of the participants refused to give answers to some of the displays. This was not noticed until after the completion of the experiments, however there appeared to be no pattern between the displays that were not categorized by the ASD group. The programme for the experiment was designed to show the difference between when a participant gives an erroneous answer, hits a wrong button or moves on without giving an answer. As the programme showed that participants had moved on without answering, these lack of responses were classified as wrong answers but could not be included in the confusion matrices. Future experiments should be run under the condition that an answer must be obtained before seeing the next display.

From the confusion matrices we can see generally more confusion in the animacy displays than in the video displays, and in the animacy displays, as expected, there is more confusion for the side view displays than for the overhead displays. For the animacy side view displays, Neurotypicals seem to have a slight tendency to call intentions play, and appear to be poor at recognising following. The ASD group shows a similar pattern but confusion is split into calling displays playing and fighting – again following is poorly recognised, as is chasing. For the overhead animacy views, the neurotypicals are good at distinguishing all the intentions except chasing which is mistaken for following, and playing which is mistaken mainly for fighting and guarding. The ASD population show only good recognition for the intentions of following and flirting, with fighting being mistaken for playing, guarding being mistaken for fighting and playing, and playing being mistaken for chasing, fighting and guarding.

In the original footage displays we see less confusion. In the side view displays, neurotypicals are good at recognising fighting, flirting and following, with playing sometimes being mistaken for fighting, and chasing being mistaken for following. The ASD population show good recognition of chasing, fighting, following and playing, with guarding being mistaken for playing and fighting, and flirting being mistaken for following and playing. In the overhead view displays neurotypicals show good distinction between all intentions with only slight confusions with playing being mistaken for fighting, and chasing being mistaken for following. The ASD group on the other hand show poor recognition of all intentions except fighting and following. Here, chasing is mistaken for following, flirting is mistaken for guarding and playing, guarding is mistaken as fighting and playing, and finally, playing is categorized partly as all other intentions except flirting.

Chapter Discussion

Autistic Spectrum Disorders are developmental disorders diagnosed when an individual is shown to have abnormalities in a triad of behavioural domains, including social communication and development, as well as obsessive interests and repetitive behaviours (APA, 1994; Wing and Gould, 1979). The underlying basis of the conditions is still unknown with various theories pointing to perceptual and cognitive problems in understanding the world and people in it (Baron-Cohen, 1995; Frith, 1989) as well as neurological abnormalities in various areas of the brain associated with social understanding (Abell et al, 1999; Baron-Cohen and Belmonte, 2005). One regularly reported trait of ASD is a poor Theory of Mind (Woodruff and Premack, 1978) or an inability to attribute

mental states to explain and predict the behaviour of others (Abell et al, 2000; Baron-Cohen et al, 1985; Baron-Cohen et al, 1986). The experiment presented in this chapter looked to further the understanding of the attribution of intention by people with ASD using animacy displays derived from human motion.

We examined the ability of people with ASD to differentiate between intentions displayed in video recordings of human interactions, and in the animacy displays derived from these recordings, and compared this to a neurotypical population. Ability to judge intentions was compared over the two viewpoints of overhead and side view. Taking into consideration previous research by other authors (Abell et al, 2000; Baron-Cohen et al, 1985; Baron-Cohen et al, 1986; Castelli et al, 2002) which showed people with ASD have reduced ability in correctly interpreting the intentions of others, it was hypothesised that in general the ASD population will be poorer at distinguishing intentions between the displays compared to the neurptypicals, and that this would be consistent across viewpoints and presentation conditions: original footage displays and animacy displays. The results showed that, as expected, people with ASD were able to judge the intentions in the displays, but at levels reduced from the neurotypicals, and that this difference varied depending on the viewpoint and presentation condition that was being used to display the intention.

When judging intention in the animacy displays, neurotypicals showed a similar pattern of results as to what has been seen throughout this thesis. Neurotypical participants showed that ability to recognise intentions is greater for overhead displays than it is for side view displays, though at a rate higher than previously reported. In previous experiments we have found neurotypicals to score a hit rate of around 50% in the overhead displays and 30% in the side view

displays, however hit rates of 63% and 48% were achieved in this experiment, though this difference may be a fortuitous effect due to the random selection of participants.

The ASD group scored hit rates of 25% for the side view and 50% for the overhead displays, in the animacy presentation condition. As the hit rates were above chance, i.e. 16.67%, it would appear that the high functioning autistics that took part in the experiment could successfully perform the intention recognition task. Furthermore, it was shown that the ASD population displayed the same pattern of results as seen in the Neurotypical group, i.e. ability at judging intention is better for overhead displays.

The confusion matrices of animacy displays for both groups show similar results with the ASD group's matrices being almost an exaggerated version of the Neurotypicals, in terms of confusion. These similarities in overall rates and in the confusion matrices are interesting as they point to both groups attempting to perform the task in the same manner using the same motion cues, however this ability is slightly less accurate in the ASD group. The results also show that overall, this particular ASD group do have the ability to interpret the global motion of shapes as intentional motion of humans, at two viewpoints, one being a less common viewpoint.

It was expected that for both experimental groups, ability to judge intentions would improve when viewing the original footage displays, given the increase of available visual information. This was found in the Neurotypical group, scoring high overall hit rates of around 80% for both viewpoints. However, the ASD group only showed improvement in the side view display, with an overall hit rate of 65%, conversely showing a decrease in ability in the

overhead displays, dropping from 50% in the animacy displays to 43% in the original footage displays.

The confusion matrices in the original footage presentations for the neurotypicals showed only slight confusions between playing and fighting and between chasing and following, for both viewpoints. That there is still some confusion in the original footage displays may help explain confusions in the animacy displays of these intentions, not only in this experiment but also in experiments in the previous chapter where these animacy displays were used. It would be clear that if the original footage displays are not correctly categorised every time then it is unlikely that the animacy displays would be categorised to a better degree. This would suggest there was ambiguity in some of the original footage displays as regards the intention that they portrayed; an aspect that should be addressed in future experiments.

The confusion matrices in the original footage presentations of the ASD group are quite contrasting between viewpoints. The confusion matrix for the side view shows reasonable clarity in all displays except guarding and flirting. On the other hand, the confusion matrix for the overhead shows large confusion for all displays. This difference and subsequent decrease in ability is surprising given the increase of information. However, certain aspects of underlying theories of autism may help explain the results.

The Empathizing-Systemizing (E-S) theory (Baron-Cohen, 2002) would suggest that people with ASD would be poor at the task of the current experiment due to their reduced ability to attribute appropriate intentions to others. Furthermore, a suggestion of the E-S theory is that people with ASD have poor imaginative ability and cognitive thinking, both of which would be required in

the current task; not only to perceive the circles as animate and moving as humans, but to be able to imagine oneself or others from an overhead perspective. It is unclear as to how the Systemizing aspect would explain the results of this experiment as there is no real capacity for participants with ASD to learn rules that govern the motion, and then use these rules to categorize the intentions. Furthermore, the only improvement shown by the ASD population is in categorising displays in the original footage from the side view. It is possible that people with ASD are using rules previously learnt about human motion to categorize intentions in the side view displays of the original footage, and that this prior learning is not useful or transferable to motion shown from overhead; however it is more likely that increased visual information, as discussed in Chapter 5, is the cause of this improvement. In addition, it may be that people with ASD are using systematic rules to make their judgements of intentions, but that these rules are not beneficial for judging human motion. Applying a feedback mechanism to the experiment, where participants are informed if their decision is correct or incorrect, would have enabled both populations to make more accurate systematic judgements. However, as it was desired to see how efficient both groups were at judging intentions in these basic, minimal animacy displays, using only the cues available and any prior knowledge that may be helpful, feedback was not given. If the E-S theory (Baron-Cohen, 2002) holds true, we would have expected both populations to have shown increased ability, with perhaps greater accuracy seen in the ASD population, due to their proposed superior systemising skills.

Weak Central Coherence (WCC) theory would suggest that people with ASD would be poor at this experiment due to an inability to bind local details of

a scene into a global percept (Frith, 1989). If we examine the displays in terms of global and local aspects we would have the animacy displays that show only global motion and the original footage displays that show global and local motion, if we accept trajectories and speed, etc to be global cues, and the movement of limbs to be local cues. In terms of the WCC theory, people with ASD would show impairment in the animacy displays due to poor use of global cues, which may explain a similar but reduced pattern of performance compared to neurotypicals. In the original footage display, a reduction in performance due to an increase in local cues would also be consistent with Weak Central Coherence theory. We suggest that a probable cause, for reduced performance in the overhead, original footage displays, may be due to the saliency of the local motion cues, e.g. the arms, in the overhead displays of the original footage, compared to the side view displays. In the side view displays, the whole body of the actor is seen and the arms and legs are in relative proportion to the body, with global motion cues possibly more apparent, resulting in high hit rates. In the overhead displays, only the tops of the heads of the actors, the shoulders and the arms are seen and thus, the arms make up a larger proportion of the visible actor than they do in the side view. It is therefore possible that the local motion of the arms is now increased. If ASD participants did focus on the local motion of the arms in the overhead displays, rather than focusing on the global motion of the agents, then this would indeed result in greater confusion and lower hit rates.

In terms of a neurological theory, Baron-Cohen and Belmonte (2005) indicated that sensory hyperactivity and abnormal attentional selectivity might result in increased noise and a reduction in capacity for information that would be of use for judging the actions of others. Their theory may help explain the

results of this experiment if it can be shown that people with ASD are focusing on the irrelevant local details of the displays to make their judgements, though neuroimaging studies would be required to confirm the validity of this statement. Furthermore, it would also be of interest to compare neuroimaging results of an experiment using the current set of stimuli with that used by Castelli, Frith, Happé and Frith (2002) to examine whether similar levels of activation are seen in brain areas associated with social understanding (Castelli et al, 2000) when using animacy displays derived from human motion, as seen when using cartoon representations of human actions.

One must be careful when drawing conclusions from the work presented in this current experiment. Though the results point to clear deficits in the ability to attribute correct intentions to animacy displays by people with ASD's, other factors or detriments may be interwoven in the results that the experiment does not test. For example, Bertone et al (2003; 2005) and Bertone and Faubert (2006) have previously shown people with ASD to have reduced performance, compared to a control population, in motion direction judging tasks, using first-order (luminance defined) and second-order (texture-defined) displays. This would suggest that people with ASD's may have a general motion perception impairment, and without testing for this, it is unclear to how great an extent the results of this study are an inability to perceive intention or an inability to perceive motion in general.

In conclusion, it has been shown that people with Autistic Spectrum Disorders can successfully distinguish intentions in video displays of human motion and in the animacy displays derived from them. Furthermore it has been shown that people with ASD are successful at distinguishing intentions in

displays shown from both a side view and an overhead view. However the ability to do so was reduced in both viewpoints in the animacy displays and in the overhead viewpoint in the video displays, compared to an age-matched neurotypical control group. It is proposed that impaired cognitive reasoning, in conjunction with impaired motion perception, is the underlying cause of these differences.

Chapter 7 – General Discussion

Perceiving Animacy

For many of us the ability to understand and interpret the actions and intentions of those around us would appear to be a straightforward procedure. Nevertheless, the simplicity with which we achieve this understanding would appear to be in no way indicative of the complex processing of visual information that appears necessary to perform this task. It has been shown that in order to study the underlying mechanisms of action understanding we can make use of simplified scenarios referred to as animacy displays. The perception of animacy is the attribution of human-like qualities, such as beliefs, thoughts, intentions and desires, to the motions of moving geometric shapes, first reported by Heider and Simmel (1944). This thesis looked to further research into the attribution of intentions by exploring a new method for the creation of animacy displays, using human motion as the basis for the displays, and by looking at the perceptual cues within these animacy displays that drive the cognitive interpretation of the motion. By doing so, we look to advance previous theories regarding the attribution of intention by typically developing people and people with Autistic Spectrum Disorders.

The Creation of Animacy Displays

Fundamental to the research of this thesis is an examination of the methods by which animacy displays are created. Heider and Simmel (1944) created their original display by using a stopgap, or key-frame, technique

involving a frame being filmed, moving the shapes, filming the next frame and so on. Subsequent research has used various other techniques for the creation of animacy displays including distortions and variations of the Heider and Simmel movie (Abell, Happé and Frith, 2000; Berry, Misovich, Kean and Baron, 1992; Bloom and Veres, 1999; Csibra, Gergely, Biro, Koos and Brockbank, 1999; Castelli, Happé, Frith and Frith, 2000; Dittrich and Lea, 1994; Gergely, Nadasdy, Csibra and Biro, 1995; Heberlein, Adolphs, Pennebaker and Tranel, 2003; Kuhlmeier, Wynn and Bloom, 2003); mathematical algorithms (Zacks, 2004); complex computer games (Barrett, Todd, Miller and Blythe, 2005; Blythe, Todd and Miller, 1999); or simple displays involving one or two shapes moving on simple trajectories (Bassili, 1976; Blakemore, Boyer, Pachot-Clouard, Meltzoff, Segebath and Decety, 2003; Gelman, Durgin and Kaufman, 1995; Stewart, 1982; Tremoulet and Feldman, 2000; Tremoulet and Feldman, 2006). Results from behavioural and neuroimaging studies using these methods have been discussed in terms of how humans attribute intention and the global cues that we use to do so. However to the best of our knowledge, prior to this thesis there has be no reported research that has used actual human motion as the basis of animacy displays.

The method of animacy display production presented in this thesis uses a subtractive approach, via a combination of the Eyesweb open platform (Camurri, Trocca and Volpe, 2002), Matlab (the Mathworks, Natick), Showtime (Watson and Hu, 1999) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) to extract positional co-ordinates of actors from video recordings, and uses these positional co-ordinates to create the animacy displays. It is suggested that the resultant animacy displays are the first incidence of displays that are truly

representative of human motion, allowing for more accurate conclusions about the attribution of human intention. In addition, the technique has the added advantage of allowing multiple viewpoint recordings of the same action/interaction, depending on the availability of video cameras, therefore creating the possibility of comparing judgements of intention from motion, from differing viewpoints. Furthermore, this subtractive technique for the creation of animacy displays facilitates a comparison between the ability to judge intentions in the original video recordings and in the resultant animacy displays. Problems were raised with this technique, especially when extracting the co-ordinates of the actors, and possible amendments to the technique may be considered. These amendments may include using single markers on the bodies of the actors to clearly extract the co-ordinates of the centre of the actor, or filming the actors in different colours, such as red and green, and performing colour extractions to get more accurate recordings of the co-ordinates.

Through a series of explorative experiments in Chapter 3, using a wide range of interactions and scenarios, it was shown that this new method for the production of animacy displays was a valid technique for creating displays involving moving geometric shapes that would be viewed as alive by participants. Results from post-ratings of self-propulsion tasks in Experiment 3.3, using a 9 point Likert scale, indicated that the agents/shapes in many of the displays created using this technique, were viewed as being in control of their own movement. This was seen for simple movements such as walking and jogging, as well as dyadic interactions such as flirting and chasing, and finally using motions taken from modern dance routines involving two dancers.

Spontaneous uses of animate terms to describe animacy displays

Of interest was the infrequent use of animate terms used to describe the displays when using free response tasks. Of all the displays created via this method in the first experimental chapter, few displays were readily and consistently described in animate terms by participants. In Experiment 3.3, only the displays referred to as "SocialInteraction" and "Attraction" were often described in terms that would suggest that the participants viewed the shapes as alive, and in the second experimental chapter, Experiment 4.5, we found only displays involving chasing and following intentions to be regularly described in animate terms. All other displays, except one, were less frequently described in animate terms: the monadic modern dance animacy display from Experiments 3.2 and 3.3 was never described in animate terms.

The displays often described in animate terms appear to have either very simple and obvious motion patterns, such as following and chasing, or have an obvious storyline attached to the motion, as in the SocialInteraction and Attraction. SocialInteraction was the display that was most often described in animate terms and displayed a crouching/injured actor/square being approached by a second actor/square and resulting in the crouched actor chasing the second actor off-screen. In this display there is a clear story that the first agent is pretending to be hurt to draw the second one closer in, in order to strike, and the second actor is curious as to the what the first actor is doing or is in some way trying to help the first actor. There are of course other interpretations of this display, including that we could be witnessing the end of a fight that has being going on prior to the display we see. This display is similar to the original Heider and Simmel movie in that both displays have agents with clear roles and

both displays have a fluid underlying story, where actions and interactions evolve and change in the one display. We propose that it is the evolving of intentions, e.g. a chase becoming a fight in the Heider and Simmel display, as a scenario develops, which facilitates free responses that clearly indicate participants view the shapes as having purpose and intentions. Conversely, we put forward that when there is no obvious evolving of intentions or an underlying story, the motion trajectories displayed must be simple and obvious, such as chasing or following, in order to obtain free responses indicating the perception of animacy. Even with such displays of simple motion trajectories, the free responses that are obtained will be simple and will give no more than the basic action, for example a response of 'one ball is following another' may be given. It may seem obvious, but it would appear that simple displays of one motion of an action/interaction, such as a chase, in a sparse environment will be met with simple responses, and displays involving rich content with clear roles of agents, each with their own intentions, desires, beliefs, purposes, will receive rich responses. The research carried out in this thesis may not yet fully support this claim, and an analysis looking at the quantity of words and type of words used would be required to do so. However, the basis is there and would make interesting work in researching the content of the motion and its context within the environment, similar to previous work of Tremoulet and Feldman (2006).

Free Response Tasks Versus Forced Choices or Ratings Tasks

Differences between results of free response tasks and results of post-ratings of self-propulsion tasks raise the question of what is the most appropriate task for these experiments. As we have seen in this thesis, even though some

displays are not described in animate terms, participants may still give these displays high ratings of self-propulsion or will be able to accurately judge the intention that the agents are portraying. It is possible that the free response task is not sensitive enough for these experiments, with the wording of the task question being of great importance. However, I believe that this difference is again due to the level of content of action and story in the display. In experiments where the free response task has proven to be a suitable task, displays have been rich in content (Abell, Frith and Happé, 2000; Bloom and Veres, 1999; Castelli, Happé, Frith and Frith, 2000; Castelli, Frith, Happé and Frith, 2002; Heberlein, Adolphs, Pennebaker and Tranel, 2003; Heider and Simmel 1944) as opposed to when rating tasks or multiple choice tasks have been used, where the displays have been sparse (Barrett, Todd, Miller and Blythe, 2005; Blythe, Todd and Miller, 1999; Gelman, Durgin and Kaufman, 1995; Stewart, 1982; Tremoulet and Feldman, 2006; Tremoulet and Feldman, 2000). It is therefore proposed that the task used in animacy displays must be appropriate for the displays being used. However, this use of an appropriate task may still not ensure that human motion can be degraded into animacy displays and be perceived as animate for the monadic dance animacy display, from Chapter 3, was never described in animate terms, and received low ratings of self-propulsion indicating that the movement appeared to be controlled by an external force.

Monadic Dance display & the Newtonian-Violation Hypothesis

One of the main theories of animacy is referred to as the Newtonian-violation hypothesis (Stewart, 1982) and suggests that an agent that is perceived to undergo unexplained changes in direction and speed, i.e. defy Newton's laws of physics (1687), will be perceived as animate whether it portrays intentional motion or not. Tremoulet and Feldman (2000) showed that this was true of a single shape that underwent a change in direction or speed in a sparse environment, however later work (Tremoulet and Feldman, 2006) suggested that the role of the environment and the context of the motion within that environment are more important than first thought. Tremoulet and Feldman (2006) show support for a second theory on the perception of animacy, the Intentionality hypothesis. This hypothesis suggests that animacy is perceived when intentionality is perceived, even if the intention is non-visible, i.e. when the intention of the motion is directed towards an entity or object that is not seen on the screen.

The monadic dance display, depicting a solo dancer freely moving around an empty stage, found in Experiments 3.2 and 3.3, would appear to raise an interesting point about the perception of animate agents as regards to which of these two theories stated above is more likely. Results from the two experiments incorporating this display showed that the solo dancer animacy display was never described in animate terms, in Experiment 3.2 and Experiment 3.3, and was also given low ratings of self-propulsion, in Experiment 3.3, suggesting that the motion of the shape appeared to not be animate but to display more random action. This is curious as the motion of the animacy display derived from the footage of the solo dancer was continually undergoing changes in direction and speed, for the original footage showed a dancer moving across an entire stage at

various speeds with numerous stops and starts. The intention in the display is hard to perceive, as the display is a modern dance theme. If the Newtonian-violation hypothesis is accurate then the solo dancer display should have received higher ratings of self-propulsion and be described in animate terms of some form, as the motion pattern fits with this theory. However as the display is not rated or perceived as such and as the intention in the motion is hard to perceive or is unclear, this is indicative of the Intentionality hypothesis being accurate, which would have suggested that if the intention of the agent had been clearer then the motion of the agent would have been perceived as animate. Though this understanding of the results is based on the comparing of data from one display across two experiments, nevertheless, further analysis using displays of freely moving agents with intentions that are unclear or hidden, may vindicate this interpretation.

Another possibility is that the movement of the monadic dance display, when taken out of context and shown as an animate block, shows motion that is greatly stochastic and people do not perceive it to be that of a self-propelled being. This reasoning would suggest that there might be a spectrum of sorts where at one end there is Newtonian motion and the shapes are viewed as having non-animate causal motion (Michotte, 1946), and at the other end there is Brownian motion (Brown, 1828), relating to the random movement of particles being bombarded by other forces, and the shapes are viewed as having non-animate random motion. In between these two extremities, would be what people would view as self-propelled animate motion: neither simple linear motion nor stochastic motion. This is merely conjecture and would require analysis into the point of subjective equality both for when simple linear motion

becomes animate and for when animate motion becomes stochastic non-animate motion.

The role of Viewpoint in the interpretation of animacy displays

A main advantage of this new technique for the production of animacy displays is being able to record and examine the same action/interaction simultaneously from different viewpoints. This has led to the study in this thesis of the importance of viewpoint in animacy displays: an area that has received no attention prior to this. The majority of animacy displays have shown the motion of agents from an overhead perspective, i.e. a bird's eye view (Abell et al, 2000; Barrett et al, 2005; Bassili, 1976; Blakemore, Boyer, Pachot-Clouard, Meltzoff, Segebath & Decety, 2003; Bloom & Veres, 1999; Blythe et al, 1999; Castelli et al, 2000; Castelli et al, 2002; Gelman et al, 1995; Heider & Simmel, 1944; Tremoulet & Feldman, 2006, Tremoulet and Feldman, 2000) with only a small number of experiments using displays showing the motion of the agents from the side view, our natural viewpoint (Gergely, Nadasdy, Csibra and Biro, 1995; Csibra, Gergely, Biro, Koos and Brockbank, 1999; Kuhlmeier, Wynn and Bloom, 2003). The data from experiments throughout this thesis has shown a consistent advantage for perceiving animacy displays from the overhead perspective.

Experiment 3.4 showed that agents in displays shown from an overhead perspective regularly appeared more self-propelled than the equivalent agents shown from the side view, when using a self-propulsion rating task. Experiments throughout Chapter 4 continually showed that there was an advantage for judging intentions in motion shown from the overhead viewpoint over the side view

displays. Furthermore, there was also found to be more chance of people describing the events in animate terms in a free response task, when shown from the overhead, in Experiment 4.5. It is proposed that the key to this difference, in the judgement of intention across viewpoints, is the cues between the agents that are visible in the overhead displays but that are not visible in the side view displays. Blythe et al (1999) and Zacks (2004) examined cues that were useful for determining intentions and event parsing, respectively. These cues ranged from positional co-ordinates of each agent to relative cues between agents, such as relative angles, speed and distance between agents, and finally absolute cues between an agent and the background, such as absolute speed. Indeed, our own analysis of the motion properties, in Chapter 5, showed various differences between the motion patterns and the kinematics of displays, across viewpoints. Furthermore, the stepwise regression analysis attempted, confirmed previous research highlighting the importance of speed and acceleration cues for predicting intentions in the Overhead displays, and the importance of positional cues in the Side View displays..

Comparing the viewpoints, the overhead displays portray the motion in two dimensions, whereas in the side view, the motion is generally one-dimensional. As a result of this difference many of the above mentioned cues between agents, in particular relative cues, are lost in the side view; therefore it is hard to get a true understanding of the position of one agent compared to another. Where in the overhead view agents will appear to be a distance from each other going in a circle, on the side view they will appear as passing right next to each other going on linear paths of left to right. Attempts to give a clearer understanding of the motion, by using occlusion cues, giving ordinal depth and

identity, made only a small significant difference suggesting that the information required to make accurate judgements goes beyond ordinal depth or identity. It is possible that the inclusion of a perspective cue, where a distal agent would look smaller, and vice versa, or other cues to depth such as stereo depth or motion parallax, might clarify the displays and produce higher ratings.

The fact that the side view displays tended to receive lower ratings of self-propulsion would be consistent with the motion being described as random motion in free responses tasks and of poor results in judging intentions of displays. If a participant views the display as being random motion then they will describe it as so, and it will be harder to judge the correct intention when they do not see any intention in the motion. Nevertheless, it is worth noting that despite this lack of information, participants were still able to judge intentions at levels above chance, indicating that some cues still exist for this purpose.

It would appear that the removal of certain motion cues is the most likely reason for a difference in perceptions between the two viewpoints used in this thesis. However, we are still left with the interesting scenario where participants are particularly good at perceiving animacy and intention in motion shown from the overhead viewpoint, which is an uncommon viewpoint. There are very few occasions where people will actively observe the motion of others from directly overhead, perhaps only when looking down from the windows of tall buildings. Of course, it is not clear as to whether participants actually view the shapes as being human or as bugs or ants, for which there are many possible incidences of people viewing them from an overhead viewpoint. Nevertheless, the ability to perceive intention and animacy from an overhead perspective would appear to be strong. One of the few times in life that we do perceive action from an overhead

position is as children playing with our toys, or as adults playing with toys but that is less common. It is a plausible suggestion that we learn to perceive intention from above as children when playing, thus giving a prior-knowledge solution to judging intention from above. Blythe et al's (1999) suggestion that we learn various intentions via play, both by ourselves and with others, would support this claim. An alternative explanation for the ability to perceive intention from above is that perhaps as humans we are highly adept at imaging scenes from different viewpoints, this would be akin to mental rotation theories with shapes (Shephard and Metzler, 1971). It would be hard to clarify which of these reasons is the most plausible. An interesting control would be to use a clinical population where childhood play and imagination skills are reduced. People with Autistic Spectrum Disorders (ASD), where both these traits have been shown to be reduced to differing degrees, compared to typically developed people (Baron-Cohen, 2002; Wing and Gould, 1979), may provide such a population. Previous studies with such populations have shown them to have reduced ability for judging the intentions in animacy displays, generally preferring to give more mechanical descriptions of the motion rather than descriptions suggesting that the agents had beliefs or desires (Abell et al, 2000; Castelli et al, 2002). It is clear therefore, that some form of higher cognitive thinking is required to perceive intention from an overhead perspective, be it learned or innate.

The attribution of intentions by people with Autistic Spectrum Disorders

Animacy displays are regularly used in studies of Autistic Spectrum Disorders (ASD) and in the research of the attribution of intention in people with ASD. As mentioned above, people with ASD generally refer to such displays in mechanical terms, preferring to describe them in terms of their geometric qualities, rather than perceiving the shapes as moving in human-like ways. Furthermore, people with ASD have been shown to ascribe increased wrong or inappropriate intentions to animate motions, compared to typically developed subjects (Abell et al, 2000). The summation of this thesis was a study into the attributions of intentions in people with Autistic Spectrum Disorders.

As with neurotypical participants, using our new method for creating animacy displays derived from human motion, allowed a comparison of the judgement of intentions in animacy displays by people with ASD, across the two viewpoints of overhead and side. Furthermore, we compared the people with ASD at judging intentions in the full video displays and the resultant animacy displays.

Results from animacy displays showed a reduced ability of people with ASD to correctly assign the appropriate intention to the display, similar to results shown by previous authors (Abell et al, 2000; Castelli et al, 2002). Furthermore, it was found that people with ASD followed a similar pattern of results with respect to viewpoint, in both confusion matrices and ability to judge intentions, with accuracy in the side view animacy displays being weaker than the accuracy in the overhead animacy displays. In the original video displays, we expected the Autistic population to increase in ability to assign the appropriate attention to the displays, as well as the Neurotypical population, given the increase in

available visual information. Results from the side view displays showed an increase in both populations, and similar confusions patterns, with the people with ASD being slightly weaker than the Neurotypicals. However the results from the overhead displays produced the surprising result of the ASD population actually decreasing in their ability to judge intentions, compared to when viewing animacy displays. The ASD population showed much greater confusion than the Neurotypical population, which would be consistent with an overall decreased ability for assigning the appropriate intention to a display.

The reduced ability by the ASD population for judging correct intention in animacy displays was expected, however the same pattern across viewpoints as neurotypicals, was not expected. It was thought that the people with ASD would not be good at judging the intention in the overhead animacy displays given the cognitive load and level of imagination required to understand this viewpoint, thus we would have expected low hit rates in both viewpoints of the animacy displays. However, in people with Autistic Spectrum Disorders, the level of variance between people, for different tasks, can be substantial and it is possible that the 5 subjects used in the study were not that impaired in aspects required for this intention recognition task. It would have been beneficial to incorporate animacy displays which previous research has shown people with ASD to have reduced ability to understand, as a standard or base-line measurement. Displays such as the Heider and Simmel display (Heider and Simmel, 1944) or the displays used by Abell et al (1999) and Castelli et al (2000), would have been suitable.

The fact that the same pattern of results, i.e. ability on overhead is better than the ability on side view, was found in both participant groups is interesting

as it suggests that both groups use similar motion cues to judge intentions, yet the Neurotypical group would appear to be either more efficient at using these motion cues or are more aware of other cues available. However, it is also possible that the ASD population used a systematic method for judging intentions rather than cognitive reasoning. This would be consistent with the Empathizing-Systemizing theory of Autism (Baron-Cohen, 2002), which suggests that people with ASD are impaired in mentalizing about the actions of others, but have superiority in establishing rules of systems. The results from this experiment would benefit from analysing the cues used by the ASD population for judging intentions, and comparing this to the cues used by the typically developed population, shown in Chapter 5.

In regards to the original video displays it is uncertain as to why the ASD participants should be worse at judging intentions when shown the full video display from the overhead position. This combination of viewpoint and presentation condition should have been ideal for judging intentions as you are given all the possible visual information for that viewpoint, and that this viewpoint has previously been shown to be best for judging intentions. It is unlikely that this result is due to problems with imagination or cognitive reasoning, as this could not explain that people with ASD are better at judging intentions for animacy displays from the overhead viewpoint than in the original video recordings. An alternative possibility would be an effect due to the video displays showing real people, with ASD people known to have poor social understanding skills and to dislike interacting with people (Robins, Dautenhahn and Dubowski, 2004; Robins Dautenhahn, te Boekhorst and Billard, 2005). However, an argument could be made that would suggest that if this explanation

was the case then there should have been a decrease in ability at judging intention in the side view video displays as these perhaps involve more social understanding and interaction given that the faces of the actors can be seen. It is possible that a combination of these elements, i.e. poor social understanding and motion perception, are causing this reduction and that it is more pronounced in the overhead viewpoint due to the unusual angle from which perspective has to be taken.

A final probable cause for the differences in judgement of intentions by people with ASD compared to neurotypicals may lie with the Weak Central Coherence theory (Frith, 1989), which suggests that people with ASD are poor at grasping global aspects of a situation due to an inability to ignore irrelevant local motion. A subjective comparison of the video displays across viewpoints, looking at the ratio aspect of the area of moving body (global motion) to area of the arms and legs (local motion) in the video displays, may suggest that local motion cues are more prominent in the overhead displays than in the side view. This would require further analysis of the displays used, examining the kinematics and the Quantity of Motion (Camurri, De Poli, Leman and Volpe, 2001) of the actors in the displays, at both a global level (whole-body) and at a local level (limbs, torso, head, etc). If there is a difference in levels of local and global motion, across viewpoints, then it is possible that the increased local motion has become the focus for the people with ASD and is thus preventing the ASD population from correctly perceiving and interpreting the global motion.

Limitations of Thesis and future steps

Results from the confusion matrices of the Neurotypical group showed that even in the full video presentation condition of both viewpoints, for a couple of intentions, there was confusion as to what intention was being portrayed; namely chasing and playing. Both of these displays, though well above chance, were not as accurately recognised as would have been expected, given the abundance of visual information. Both of these intentions have consistently been confused with other intentions in various experiments throughout this thesis, chasing with following, and playing with fighting. The results from the full video overhead displays indicate that these confusions will partly have risen from aspects of the original displays. As suggested from the kinematic analysis, Chase was consistently confused as follow mostly likely due to the speed of the chase intention not being fast enough, resulting in people perceiving it as a follow intention, but at jogging pace. The play display was often mistaken for a fight and this was most likely due to the physical contact between the actors in the display. As mentioned in Chapters 4 & 5, the fight display was more of a heated argument as there were no physical blows between the actors, though it was still perceived as a fight in the full video display. The play display, on the other hand, did feature physical contact between the actors, as it portrayed a game of "Tag". It is possible that the physical contact between the two actors and the speed and acceleration that they approached each other with, may have allowed participants to sometimes perceive it as a fight. It is unfortunate that these discrepancies were not picked up on, in early viewings of the displays made during their creation. It would be wrong to argue that the faults in these displays invalidate the conclusions reached in the experiments throughout this thesis, as

these intentions were still always recognised above chance in all forced choice experiments. Nevertheless, it would be fair to concede that perhaps some of the confusions noticed may have become reduced had the original displays been better representations of the intentions that they were meant to show. It is suggested any future work that develops from this thesis would benefit from the use of various recordings of various actors, with less choreographed scripts to create the displays. These displays could then be piloted in their original format in order to achieve a set of appropriate displays, thus eradicating any effects from nuances in certain displays.

The work of this thesis has introduced new avenues in the research of the attribution of intentions, including viewpoint and method for the production of the displays. However, it does not provide extensive conclusions about what specific cues inform these intentions, due to the concise stimuli set, and only broad conclusions have been drawn.. Previous research by other authors (Bassili, 1976; Blythe et al, 1999; Csibra, Gergely, Biro, Koos and Brockbank, 1999; Gergely, Nadasdy, Csibra and Biro, 1995; Heider and Simmel, 1944) has shown various cues ranging from specific global motion cues to reciprocal movement, surface changes etc, to show that these have an effect on the perception of animacy. Though some cues have been addressed in detail in this thesis, namely cues such as self-propulsion, changes in speed, changes in direction, cues to occlusion, identity and boundaries, perhaps more focus on the global motion cues involved would have provided more insight into the precise mechanisms underlying the perception of animacy. Though Chapter 5 does attempt to explore the global cues of the motion, and how they relate to percept of intention by

participants, if more stimuli had been available for this analysis to be performed on, stronger conclusions could have been drawn.

An interesting cue to analyse in future research would be the Quantity of Motion (QoM) cue, used briefly in the first three experiments of Chapter 3. The Quantity of Motion is a first, rough approximation of the physical momentum of an actor (Camurri, De Poli, Leman and Volpe, 2001), and was used as a motion cue in the first three experiments of this thesis, with the effect that the square which represented the actor increased and decreased in size, in accordance with increases and decreases in QoM. It was found that the occurrences of animate terms and ratings of self-propulsions were slightly higher when this cue was added. Changes in surface shape have been shown to help the perception of animacy but, similar to self-propulsion, are not essential for the percept. In Chapter 3 the use of this cue was discontinued in favour of exploration of viewpoint and also because it was unclear if participants were actually making use of the Quantity of Motion cue or if the changing in shape and size was giving a cue to depth. This would be an interesting aspect to do further research on to see if people are susceptible to such a cue for judging intention, as they have already shown to be for judging emotions in dance (Camurri, Lagerlöf and Volpe, 2003; Timmers, Camurri and Volpe, 2003). Furthermore it would be interesting to use the QoM cue as well as more common global cues to research the perception of emotion in animacy displays, another area that, though emotions are regularly reported in animacy displays, has received little attention.

The intention of this thesis was to explore certain cues that lead to the attribution of social intentions to animacy displays, however certain manipulations of the displays were not incorporated, which previous research has

shown to alter the perception of human motion. Such manipulations would include playing the displays in reverse, showing the displays inverted along the horizontal axis, or manipulating the frame rate at which the displays are shown. Previous authors have shown, in animacy displays, that playing the display in reverse will produce greater variation in free response reports of participants (Heider and Simmel, 1944), which would be attributable to the cognitive understanding of the display, given that the low-level kinematics and dynamics of the display are maintained. Berry, Miscovich, Kean and Baron (1992), and to an extent Bassili (1976), showed that altering the temporal relationship between shapes in animacy displays greatly affects the likelihood of seeing the shapes as self-propelled, animate agents. Berry et al (1992) showed that if frames were dropped from the Heider and Simmel display, thus disrupting the dynamic relationship of agents, participants were less likely to report the actions using terms relating to human movement. Both of the these cues, manipulating playback rate and playing the displays in reverse, could have been explored in this thesis to explore how altering them would have affected animacy displays created from human motion. Though we can only hypothesise their effect, there would be no reason to suggest that manipulating the current displays in the mentioned ways would produce outcomes any different from the findings of the previous research (Berry et al, 1992; Heider and Simmel, 1944). Altering the frame rate, or dropping frames from the display will affect the temporal relationship between agents and we would expect reduced ability to correctly attribute the appropriate intention. Playing the displays in reverse would be expected to again reduce ability to perceive the correct intention but perhaps not to as great an extent as manipulating frame rate. When Heider and Simmel

(1944) showed their display in reverse, the underlying story of the display was affected, i.e. why was it all happening. However, from the reports, it can be seen that the understanding of the movements within the story were only altered slightly. For example, people still recognised that the intention of a chase or a fight was present at certain parts of display, what became altered was who was the instigator of the chase and fight: in the normal display, the big triangle is the instigator, and in the reversed display, the smaller shapes become the instigators. With this reasoning, we could imagine that in the displays of this thesis, showing them in reverse would have only a small impact on the ability to correctly perceive the intention, for we are only asking to perceive the actual intent, and not to give cognitive understanding of the display as a whole. The effect of showing the movies in reverse is perhaps reliant on the task involved: understanding the global cohesion of the display using a free response task such as in Heider and Simmel (1944); or understanding the intention portrayed in short animacy displays using a forced choice task, as in this thesis.

Another possible manipulation, often used in the perception of human motion and in the perception of faces, is the use of inverted displays. In biological motion perception, using point light displays where only the joints of a human body are visible marked by white dots, the use of inverted displays has been shown to greatly reduce the ability to judge various traits that are accessible in upright displays (Berenthal and Pinto, 1994; Clarke, Bradshaw, Field, Hampson and Rose, 2005; Dittrich, Troscianko, Lea and Morgan, 1996; Grossman and Blake, 2001; Johansson, 1973; Kozlowski and Cutting, 1977; Pavlova and Sokolov, 2000; Shipley, 2003; Troje, 2003). Similarly, in face perception literature, evidence shows that it is more difficult to recognise faces

that are inverted, compared to upright faces (Farah, Wilson, Drain and Tanaka, 1995; Lewis & Edmonds, 2003; Thompson, 1980; Yin, 1969). This would suggest that the human visual system is weaker at perceiving inverted displays. It would therefore have been an appropriate line of research to test the effect of inverting animacy displays on the ability to correctly to perceive the intention portrayed by the agents in the displays. However, it is unclear to what extent this manipulation would have on the perception of animacy. Looking at the overhead perspective, the regular vantage point for animacy displays, it is probable that inverting the display along either the horizontal or the vertical axis would have little or no effect. This is suggested as there is no canonical overhead position to view motion from, and viewing the motion going from left to right, or from top to bottom, would subjectively, appear no different from viewing it in the flipped. In side view animacy displays, I would hypothesise that inverting the displays would only have an effect in certain situations. In life, watching a chase or a fight moving from right to left is the same in the flipped display, and we would not expect it to be different for when perceiving animacy displays: so this manipulation would not work. There is perhaps an argument that inverting the displays along the horizontal would affect ability to perceive the correct intention as there is a canonical way to perceive people from the side: feet to the bottom of our visual field and the head to the top. It is possible that inverting this viewpoint may affect the perception of animacy, but this would require research into whether or not people are privy to the orientation of an agent in an animacy display, in regards to real world anatomical orientation. Another situation would be where context is given to the display, for example the boundary cue used in Experiment 4.3. As these displays gave anchor points, i.e. the corners and edges

of the boundaries, that people could use to determine the orientation of the display, we would suggest that altering these displays horizontally may affect peoples' perception of animacy as their means of orientation has been removed.

The above discusses only some of the possible manipulations that could be performed to see the effect on the perception of animacy. All would have been valid inclusions to the research of this thesis, however were not incorporated due to the focus of this thesis on other cues such as occlusion and identity.

General Conclusions

In conclusion, this thesis has provided valuable research into the attribution of social intentions via animacy displays. Throughout this research a new technique for the production of such displays has been conceived and developed that creates animacy displays that are now truly representative of human actions. Using video recordings of actors as the starting blocks of a subtractive method, we create displays where geometric shapes give direct representation of the movement of human actors, allowing for more accurate postulations about how people judge social intent. From this technique research has shown that the perception of intentions is more salient from an elevated viewpoint, due to global cues that can be perceived from this position, however it is unclear whether the ability to perceive intention from above is innate, learned, or is solved via cognitive thinking and imagination. It has also been shown, that the global motion cues of speed and acceleration appear to be the motion properties that most influence the perception of animacy and intention. Furthermore, this thesis has also advanced research into the perception of intentions in people with ASD, showing an interesting deficit in judging

intentions in overhead video displays, speculated to rise from theoretical explanations based on weak central coherence and inappropriate focus of attention. This work has relevance in various psychological fields including cognition, perception, theory of mind and neuroscience, where there is interest in the understanding of the attribution of social intention; in clinical fields, dealing with populations that are poor in understanding intention from motion; and finally, in public sector fields including people interested in surveillance and intention from human motion, via manual or automated systems using video cameras.

References

Abell, F., Happé, F. and Frith, U. (2000). Do Triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development, ***Cognitive Development****, 15, 1-16.*

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., Happé, F., Frith, C. and Frith, U. (1999). The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans, ***Neuroreport,*** 10*, 1647-1651*.

Akaike, H. (1974). A new look at the statistical model identification. ***IEEE Transactions on Automatic Control,*** *19 (6), 716–723.*

APA. (1994). ***DSM-IV diagnostic and statistical manual of mental disorders***, 4th Ed. Washington DC: American Psychiatric Association.

Asperger, H. (1944). "Autistic psychopathy" in childhood. In Autism and Asperger Syndrome, U. Frith, editor and translator 1991 (Cambridge: Cambridge University Press), pp 37-92.

Bachevalier, J. (1991). An animal model for childhood autism: memory loss and socioemotional disturbances following neonatal damage to the limbic system in monkeys. *In Tamminga, C.A., Schulz, S.C. (Eds.), Advances in Neuropsychiatry and Psychopharmacology (Vol. 1). Raven Press, NY, pp 129 – 140.*

Baron-Cohen, S. (1995). ***Mindblindness: an essay on autism and theory of mind.*** Cambridge, MA: MIT Press.

Baron-Cohen, S. (1998). The autistic child's theory of mind: a case of specific developmental delay, ***Journal of Child Psychology and Psychiatry****, 30, 285-298.*

Baron-Cohen, S. (2000). ***Autism and 'theory of mind'***, in *Hartley, J. and Braithwaite, A., The Applied Psychologist, Open University Press.*

Baron-Cohen, S. (2002). The extreme male brain theory of autism, ***Trends in Cognitive Science****, 6, 248–54.*

Baron-Cohen, S. and Belmonte, M.K. (2005). Autism: A Window Onto the Development of the Social and the Analytic Brain, ***Annual Review of Neuroscience****, 28, 109-26.*

Baron-Cohen, S. and Bolton, P. (1993). ***Autism: the facts***: Oxford University Press.

Baron-Cohen, S., Leslie, A.M. and Frith, U. (1985). Does the autistic child have a "theory of mind", ***Cognition****, 21, 37-46.*

Baron-Cohen, S., Leslie, A.M. and Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children, ***British Journal of Developmental Psychology****, 4, 113-125.*

Baron-Cohen, S., Wheelwright, S., Griffin, R., Lawson, J. and Hill, J. (2002). The Exact Mind: Empathising and Systemising in autism spectrum conditions. *In U. Goswami (Ed.) Handbook of Development, Blackwells.*

Baron-Cohen, S., Wheelwright, S., Lawson, J., Griffin, R., Ashwin, C., Billington, J. and Chakrabarti, B. (2005). Empathising and Systemising in autism spectrum conditions. *In F. Volkmar, A. Klin and R. Paul (Eds.) Handbook of Autism and Pervasive Developmental Disorders. Wiley.*

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. and Clubley, E. (2001). The Autistic Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High Functioning Autism, Males and Females, Scientists and Mathematics, ***Journal of Autism and Developmental Disorders,*** *35, 5-17.*

Barrett, H.C., Todd, P.M., Miller, G.F. and Blythe, P.W. (2005). Accurate judgements of intention from motion cues alone: A cross-cultural study, ***Evolution and Human Behaviour****, 26, 313-331.*

Bassili, J.N. (1976). Temporal and Spatial Contingencies in the Perception of Social Events, ***Journal of Personality and Social Psychology***, *33 (6), 680-685.*

Bauman, M.L., and Kemper, T.L. (1994). Neuroanatomic observations of the brain in autism. *In ML Bauman and TL Kemper (Eds.), The Neurobiology of Autism, pp. 119–45. Baltimore, MD: John Hopkins Univ. Press.*

Bertenthal, B. I. & Pinto, J. (1994). Global processing of biological motions. ***Psychological Science,*** *5, 221-225.*

Berry, D.S., Misovich, S.J., Kean, K.J., and Baron, R.M. (1992). Effects of Disruption of structure and motion on perception of social causality, ***Personality and Social Psychology Bulletin***, *18(2), 237-8.*

Bertone, A. and Faubert, J. (2006). Demonstrations of Decreased Sensitivity to Complex Motion Information Not Enough to Propose an Autism-Specific Neural Etiology, ***Journal of Autism and Developmental Disorders***, *36, 1, 55-64.*

Bertone, A., Mottron L., Jelenic, P. and Faubert, J. (2003). Motion Perception in Autism: A "Complex" Issue, ***Journal of Cognitive Neuroscience***, *15(2), 218-225.*

Bertone, A., Mottron, L., Jelenic, P. and Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity, ***Brain***, *128, 2340-2441.*

Blakemore, S.J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C. and Decety, J. (2003). The Detection of Contingency and Animacy from Simple Animations in the Human Brain, ***Cerebral Cortex***, *13, 837 – 844.*

Bloom, P. and Veres, C. (1999). The Perceived Intentionality of Groups, ***Cognition***, *71, B1-B9.*

Blythe, P.W., Todd, P.M and Miller, G.F. (1999). How Motion Reveals Intention: Categorizing Social Interactions, in Simple Heuristics that make us smart. In *Simple Heuristics That Make Us Smart* (Gigerenzer, G. *et al.*, eds), pp. 257–285, Oxford University Press.

Brainard, D.H. (1997). The Psychophysics Toolbox, ***Spatial Vision***, *10, 433-436.*

Brothers, L. (1990). The social brain: a project for integrating primate behaviour and neurophysiology in a new domain, Concepts ***in Neuroscience,*** *1, 27-51.*

Brown, R. (1828). A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies, ***Phil. Mag.*** *4, 161-173.*

Camurri, A., De Poli, G., Leman, M. and Volpe, G. (2001). A Multi-layered Conceptual Framework for Expressive Gesture Applications, *in* ***Proc. Intl MOSART Workshop****, Barcelona, November.*

Camurri, A., Krumhansl, C., Mazzarino, B. and Volpe, G. (2004). An exploratory Study of Anticipating Human Movement in Dance, *in* ***Proc. 2nd International Conference on Measurement, Analysis and Modelling of Human Functions,*** *Genova, Italy, June.*

Camurri, A., Lagerlöf, I. and Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques, ***International journal of Human-Computer Studies,*** *59(1), 213-225.*

Camurri, A., Mazzarino, B. and Volpe, G. (2003). Analysis of expressive gestures in human movement: the Eyesweb expressive gesture processing library, *in* ***Proc. XIV Colloquium on Musical Informatics,*** *Firenze, Italy, May.*

Camurri, A., Mazzarino, B. and Volpe, G. (2004). Expressive gestural control of sound and visual output in multimodal interactive systems, *in* ***Proc. International Conference Sound and Music Computing****, Paris, France, October.*

Camurri, A., Mazzarino, B., Volpe, G., Morasso, P., Priano, F. and Re, C. (2003). Application of multimedia techniques in the physical rehabilitation of Parkinson's patients, ***Journal of Visualization and Computer Animation****, 14(5), 269-278.*

Camurri, A., Trocca, R. and Volpe, G. (2002). Interactive Systems Design: A KANSEI-based Approach, *in* ***Proc. NIME2002,*** *Dublin, Ireland, May.*

Castelli, F., Frith, C., Happé, F. and Frith, U. (2002). Autism, Asperger syndrome and brain mechanism for the attribution of mental states to animated shapes, ***Brain****, 125, 1839-1849.*

Castelli, F., Happé, F., Frith, U. and Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns, ***NeuroImage****, 12, 314-325.*

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E. & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue, ***Perception,*** *34(10), 1171 – 1180.*

Constantino, J.N. (2002). ***The social responsiveness scale***. *Los Angeles, California: Western Psychological Services.*

Constantino, J.N. and Todd, R.D. (2003). Autistic traits in the general population: A twin study, ***Archives of general Psychiatry,*** *60, 524-530.*

Csibra, G., Gergely, G., Biro, S., Koos, O. and Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy, ***Cognition***, *72, 237 – 267.*

Dasser, V., Ulbaek, I. and Premack, D. (1989). The Perception of Intention, ***Science***, *243, 365-367.*

Dittrich, W.H and Lea, S.E.G. (1994). Visual Perception of Intentional Motion, ***Perception***, *23, 253-268.*

Dittrich, W. H., Troscianko, T., Lea, S. E. G. & Morgan, D., (1996). Perception of emotion from dynamic point-light displays represented in dance, ***Perception,*** *25, 727- 738.*

Farah, M.J., Wilson, K.D., Drain, H.M., & Tanaka, J.N. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. ***Vision Research***, *35, 2089-2093.*

Frith, U. (1989). ***Autism: Explaining the Enigma***. Oxford: Blackwell.

Frith, U. (2001) Mind Blindness and the Brain in Autism, ***Neuron***, *32, 969-979.*

Frith, U. and Frith, C.D. (2003). Developmental and neurophysiology of mentalizing, ***Philosophical Transactions of the Royal Society of London Series B - Biological Sciences,*** *358(1431), 459-473.*

Frith, U. and Frith, C.D. (2001). The biological basis of social interaction, ***Current Directions in Psychological Science,*** *10, 151-155.*

Gelman, R., Durgin, F. and Kaufman, L. (1995). Distinguishing between animates and inanimates: not by motion alone. *In Sperber, D., Premack, D. and Premack, A.J., Causal Cognition: A multidisciplinary debate, Oxford, Clarendon Press.*

Gergely, G., Nadasdy, Z., Csibra, G. and Biro, S. (1995). Taking the Intentional stance at 12 months of age, ***Cognition***, *56, 165-193.*

Grossman, E., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. ***Vision Research***, *41, 1475–1482.*

Happé, F. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. ***Journal of Child Psychology and Psychiatry***, 37, 873–877.

Hashimoto, H. (1966). A phenomenal analysis of social perception, ***Journal of Child Development***, *2, 3-26.*

Heberlein, A.S., Adolphs, R., Pennebaker, J.W. and Tranel, D., (2003). Effects of Damage to Right-Hemisphere Brain Structures on Spontaneous Emotional and Social Judgements, ***Political Psychology***, *24(4), 705-726.*

Heider, F. and Simmel, M. (1944). An experimental Study of Apparent Behaviour, ***American Journal of Psychology***, *57(2).*

Hill, E. and Frith, U. (2003). Understanding autism: insights from mind and brain, ***Philosophical Transactions of the Royal Society of London B,*** *358, 281-289.*

Hooper, H.E. (1983). ***The Hooper Visual Organization Test Manual***. Los Angeles: Western Psychological Services.

Howell, D.C. (1997) Statistical Methods for Psychologists. Wardsworth Publishers, 5th Edition.

Iacoboni, M., Lieberman, M.D., Knowlton, B.J., Molnar-Szackas, I., Moritz, M., Throop, C.J. and Fiske, A.P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline, ***Neuroimage**, 21 (3), 1167-1173.*

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. ***Perception & Psychophysics**, 14, 201-211.*

Jolliffe, T. and Baron-Cohen, S. (1997). Are people with autism or Asperger's Syndrome faster than normal on the Embedded Figures Task? ***Journal of Child Psychology and Psychiatry**, 38, 527-534.*

Jolliffe, T. and Baron-Cohen, S. (2001). A test of central coherence theory: can adults with high functioning autism or Asperger syndrome integrate objects in context? ***Visual Cognition**, 8, 67-101.*

Kanner, L. (1943). Autistic disturbance of affective contact, ***Nervous Child**, 2, 217-250.*

Krumhansl, C.L. and Schenk, D.L. (1997). Can dance reflect the structural and expressive qualities of music? ***Musicae Scientiae**, 1, 63-83.*

Kuhlmeier, V., Wynn, K. and Bloom, P. (2003). Attribution of Dispositional States by 12 Month Olds, ***Psychological Science***, *14(5).*

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. ***Perception & Psychophysics***, *21, 575-580.*

Le Couter, A., Rutter, M., Lord, C., Rios, P., Robertson, P., Holdgrafer, M. and McLennan, J. (1989). Autism Diagnostic Interview: A standard investigator-based instrument, ***Journal of Autism and Developmental Disorders***, *19, 363-387.*

Lewis, M.B., & Edmonds, E.J. (2003). Face detection: Mapping Human performance. ***Perception,*** *32, 903-920.*

McKay, L., Mackie, J., Piggott, J., Simmons, D.R. and Pollick, F.E. (2006). Biological motion processing in autistic spectrum conditions: Perceptual and social factors, ***Journal of Vision,*** *6(6) 1036a,* http://journalofvision.org/6/6/1036/,doi:10.1167/6.6.1036.

Michotte, A. (1946/ English translation. 1963). ***The Perception of Causality***, Basic Books.

Nevarez, H.G. and Scholl, B.J. (2000). An adaptation of Heider and Simmel. http://research.yale.edu/perception/animacy/HS-Blocks-QT.mov

Newton, I. (1687). ***Philosophiae Naturalis Principia Mathematica***, *1st edn (London, Streater).*

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion Perception, ***Perception and Psychophysics,*** *62, 889-899.*

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, **Spatial Vision,** *10, 437-442.*

Premack, D. and Woodruff, G. (1978). Does the Chimpanzee have a 'theory of mind'? ***Behavioral and Brain Sciences***, *4, 515-526.*

Robins, B., Dautenhahn, K. and Dubowski, J. (2004). Investigating Autistic Children's Attitudes Towards Strangers with the Theatrical Robot – A New Experimental Paradigm in Human-Robot Interaction Studies, *in* ***Proc. IEEE Ro-man 2004, 13th IEEE international workshop on Robot and Human Interactive Communication,*** *Kurashiki, Okayama, Japan, September.*

Robins, B. Dautenhahn, K., te Boekhorst, R. and Billard, A. (2005). Robotic Assistants in Therapy and Education of Children with Autism: Can a Small Humanoid Robot Help Encourage Social Interaction Skills? *Special Issue "Design for a more inclusive world" of the international journal,* ***Universal Access in Information Society (UAIS)***, *4(2).*

Schlottmann, A., Allen, D., Linderoth, C. and Hesketh, S. (2002). Perceptual Causality in Children, ***Child Development****, 73, 6, 1656-1677.*

Scholl, B. and Tremoulet, F. (2000), Perceptual causality and animacy, ***Trends in Cognitive Sciences****, 4(8).*

Schultz, J., Imamizu, H., Kawato, M. and Frith, C.D., (2004). Activation of the Human Superior Temporal Gyrus during Observation of Goal Attribution by Intentional Objects, ***Journal of Cognitive Neuroscience****, 16(10) 1695-1705.*

Shephard, R.N. and Metzler, J. (1971). Mental rotation for 3 dimensional objects, ***Science****, 171, 701-703.*

Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. ***Psychological Science****, 14, 377-380.*

Shor, R. (1957). Effect of preinformation upon human characteristics attributed to animated geometric figures. ***Journal of Abnormal and Social Psychology****, 54, 124–126.*

Stewart. J.A. (1982). The Perception of Animacy, PhD thesis, University of Pennsylvania.

Tagiuri, R. (1960). Movement as a cue in person perception. *In H.P. David and J.C. Brenglemann (Eds.) Perspectives in personality research.* New York: Springer.

Thayer, S. and Schiff, W. (1969). Stimulus factors in observer judgment of social interaction: facial expression and motion pattern, ***American Journal of Psychology**, 82, 73–85.*

Thompson, P. (1980). Margaret Thatcher: A new illusion. ***Perception**, 9(4), 483-484.*

Timmers, R., Camurri, A. and Volpe, G., (2003). Expressive gestures and their relation to emotional engagement in three performances of a Skriabin Etude, *in **Proc. International Conference Music and Gesture**, Norwich, UK, August.*

Tremoulet, P. and Feldman, J. (2000). Perception of animacy from the motion of a single object, ***Perception**, 29, 943-951.*

Tremoulet, P. and Feldman, J. (2006). The Influence of Spatial Context and the Role of Intentionality the Interpretation of Animacy from Motion, ***Cognition**, 99, 131-165.*

Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. ***Perception**, 32, 201-210.*

Watson, A.B. and Hu, J. (1999). ShowTime: A QuickTime-based infrastructure for vision research displays, ***Perception 28(Supplement)**, 45.*

Wing, L., and Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification, ***Journal of Autism and Developmental Disorders**, 9, 11-29.*

Wing, L., Gould, J., Yeates, S.R. and Brierley, L.M. (1977). Symbolic play in severely mentally retarded and in autistic children, ***Journal of Child Psychology and Psychiatry***, *18, 167-178.*

Witkin, H., Oltman, P., Raskin, E. and Karp, S. (1971). ***A manual for the embedded figures test***. Palo Alto: Consulting Psychological Press.

World Health Organisation. (1994). *International Classification of Diseases*. Geneva, Switzerland: World Health Organ. 10th Ed.

Yin, R.K. (1969). Looking at upside-down faces. ***Journal of Experimental Psychology**, 81, 141-145.*

Zacks, J. (2004). Using movement and intentions to understand simple events, ***Cognitive Science**, 28, 979-1008.*

Zacks, J.M. and Tversky, B. (2001). Event Structure in Perception and Conception, ***Psychological Bulletin***, *127, 1-21.*

Zacks, J.M., Tversky, B. and Iyer, G. (2001). Perceiving, Remembering, and Communicating Structure in Events, ***Journal of Experimental Psychology:*** **General**, 130 (1), 29-58.

Appendix A – Relevant Publications of the Author

EXPERIMENTS 3.1 AND 3.2

McAleer P., Mazzarino B., Volpe G., Camurri A., Smith K., Paterson S. & Pollick F.E. (2004). Perceiving Animacy And Arousal In Transformed Displays Of Human Interaction. *in* ***Proceedings for the 2nd International Symposium on Measurement, Analysis and Modelling of Human Functions***.

EXPERIMENT 3.3

McAleer P., Mazzarino B., Volpe G., Camurri A., Smith K., Paterson H.M. & Pollick F.E., (2004). Perceiving Animacy and Arousal in Transformed Displays of Human Interaction, ***Journal of Vision***, 4 (8) *pp 230*. – Abstract.

EXPERIMENT 3.4

McAleer P., Paterson H.M., Mazzarino B. & Pollick F.E., (2005). Towards canonical views of animacy from scenes of human action. ***Journal of Vision,*** 5 (8) *pp 936a*. – Abstract.

EXPERIMENT 4.1

McAleer P., Mazzarino B., Volpe G., Camurri A., Smith K., Paterson S. & Pollick F.E. (2005). Creating animacy displays from scenes of human action, ***Journal of Cognitive Neuroscience*** *pp 145*. – Abstract.

EXPERIMENT 4.3 AND 4.4

McAleer P. & Pollick, F.E., **Understanding Intention from Minimal Displays of Human Activity** (under review)

Pollick F.E. & McAleer P., (2005). Understanding Intention from Minimal Displays of Human Activity**,** *in* ***Proceedings of the 46th Annual Meeting of the Psychonomics Society****, November 10th - 13th, Toronto, Canada* – Abstract.

EXPERIMENT 4.5

McAleer P & Pollick F.E., (2006). Extracting cues to intention utilising animacy displays derived from human activity, ***Perception****, 35(3), pp425. – Abstract.*

EXPERIMENT 6.1

McAleer P., McKay L., Piggott J., Simmons D.R. & Pollick F.E., (2006). Intention Recognition in Autistic Spectrum Condition (ASC) using Video Recordings and their corresponding Animacy Displays, ***Journal of Vision*** 6 (6) *pp 1035a*. – Abstract.