University of Edinburgh
School of Informatics

Investigating the Attribution of Intentions and Goals to Simple Animations by Children with Autistic Spectrum Conditions (ASC): An Interactive Approach

4th Year Project Report
Artificial Intelligence and Computer Science

Pierre Pauly

April 04, 2012

Abstract: In the context of this honours project an interactive method for conducting the theory of mind animation description task was developed. This theoretical method was implemented in a computer based application designed to be used by 6 to 9 year old children with autism spectrum conditions and typically developing children. Subsequently a pilot study was designed to assess the usefulness of the new interactive approach and the quality of the software. The study was run with a small group of participants and provided sufficient data to illustrate the usefulness of the interactive approach.
The development process of the software and of the pilot study are described in detail in this report.

Acknowledgements

I would like to thank Helen Pain and Alyssa Alcorn for their continued support in the throughout this honours project. I would also like to thank Sue Fletcher-Watson and Thusha Rajendran for their help in evaluating Act and the experimental method. Many thanks to all the parents and children who agreed participate in this study. Finally, I would like to thank my family and my friends for their invaluable support and encouragements throughout the last six months.

Contents

1 Context and Purpose

Perception: (derived from the Latin perceptio (n-)) is defined as the process of attaining awareness or understanding of the environment by organizing and interpreting sensory information." (Pomerantz, 2006; Oxford Dictionaries, 2010)

Our everyday decisions are based on our knowledge and perception of the environment we live in. All sentient beings react to their senses, enabling them to recognize changes in the environment, in other living beings, and most importantly in themselves. The understanding that pain is a sensation caused my damaging effects to the own body is just as important as understanding social interaction cues from a member of the hive.
Influential theories in the psychological domain of object perception and perceptual development say that knowledge about a human person can be gleaned directly from their physical appearance, movement, and vocal quality (E. J. Gibson, 1969; J. J. Gibson, 1966, 1976-1986). Baron-Cohen in his work underlines the importance of perception and goes further by claiming that not only are we able to observe the behaviour of others but also *read* other people's *minds* (his *mind reading* term was coined in his paper in 1995). Furthermore perception is not only about observing an individual and inferring knowledge from their behaviour, but also entails a pervasive tendency to explain one's own and others' actions in terms of beliefs, desires and goals - attributing the ability to have thoughts, feelings and intentions to other animate beings. This aspect has been referred to as 'theory of mind' (ToM) or 'mentalizing' in the field of psychology (Abell, Happé, & Frith, 2000; Castelli, Frith, Happé, & Frith, 2002).

Autistic spectrum conditions (ASC, aka. autism spectrum disorders, ASD) are defined as a group of pervasive developmental disorders of which autism is the second most common with 1-2 people on 1,000 (Newschaffer et al., 2007). Currently, the only method of diagnosis is through the observation of the person's external behaviour. Research has put forward multiple reoccurring behavioural patterns which match the widely distinct characteristics of the individuals with ASC. These patterns are sometimes referred to as the *red flags* of autism and can be recognised in the early two years of a human baby's life (A. M. Wetherby et al., 2004) . The two most striking behavioural patterns are strong difficulties in social communication, and repetitive body movements. It has been proven by several research teams that individuals with ASC perceive their environment very differently than typically developed people. Several experiments have been conducted to analyse how these individuals attribute ToM constructs to other beings, which revealed that most people had problems with this kind of task (Abell et al., 2000; Baron-Cohen, Leslie, & Frith, 1985; Kane, Winchester, Scarpa, & Smith-jackson, 2011; and others.). This large body of research lead to various theories within the field of psychology as to how ASC can be understood and potentially treated. One of the most influential theories being that people with autism would lack the ability to read minds thus being *mindblind* (Simon Baron-Cohen, 1994, 1995).

This honours project focusses on how children with Autism Spectrum Conditions (ASC) assign intentions, goals and more complex Theory of Mind (ToM) attributes to the actors of a collection of animations. Previous research developed, among other testing methods, an experiment based on so called *ToM animations*. These are animations in which simple geometric shapes act out

common social behavioural patterns only through movements and interactions with the environment. Early investigations have shown that people generally attribute desires, emotional states and intentions to those simple geometric shapes, thus attributing a mind to the shapes (Heider & Simmel, 1944). Later research showed that not everyone followed this trend, but that people with ASC consistently showed to have difficulties with this kind of task (Abell et al., 2000, and others.). Both the false belief tasks and the ToM animation experiments are to this day two very influential methods which enabled great insight into how the mentalising process works.
The success of the previous research motivated the search for other experimental methods as well as extensions of existing ones in the hope to find better approximations of the cognitive model and finally the fundamental cause of ASC.
In the given context this honours project proposes a variation of the classic ToM animation experiment in which an interactive element is introduced to the experimental setup which is hoped to enable further engagement with the actors of the animations as well as at the same time introducing elements of exploration, replayability and possible learning opportunities for the participant. All these could potentially provide further insight into the interpretation and reasoning process through an iterative assessment method. Thus following the participant every step of the way; from getting to know the software and its controls, to the introduction of the individual actors as well as their reactions to and interactions with the other actors.
Throughout this report the fundamental ToM animation experiment is referred to using the *traditional observational method* while the variation is referred to as the *interactive assessment method,* both of which are explained in detail in section 3 of this report.

2 Outline:

This report is structured as follows. The first section explained the general context of the project as well as some general terms as preparation for the literature review which is found in the second section. The literature review covers many different topics within the field of perceptual and developmental psychology and is sub-divided accordingly.
What follows is a more detailed description of the project itself. First the context is picked up again in more detail, defining paper specific terms and moving on to stepping through the software development phases from design to implementation and iterative testing.
Section 5 describes the experimental study in detail. First Act's ToM animations are described in the materials section, followed by the description of the experimental method and the scoring process. The expected outcomes of the study are spelled out before the actual results are presented. The evaluation naturally takes place in two steps. Step one is the evaluation of the experimental data in term of the attribution of ToM. Step two briefly looks at the data from a software engineering point of view and how it informs about the quality of the software. A conclusion is drawn from the evaluations and followed by a discussion in which a critical light is shone upon the project as a whole and possible improvements are outlined in the last section.

3 Literature review

In the years 1944 and 1946 the domain of the psychology of perception was expanded by two important studies which to this day continue to inspire modern research. These studies were “An Experimental Study of Apparent Behaviour” (1944) by Fritz Heider and Marianne Simmel, and “La Perception de la Causalité” (1946, The Perception of Causality) by Albert Michotte. Their work is outlined and described in this review, before shifting to theory of mind experiments and their connection to autism spectrum conditions (ASC) and developmental disorders (DD) is discussed. Finally this literature review will conclude with recent projects in the fields of social perception and ASC intervention techniques.

In 1944 the 57th volume of the American journal of psychology included an article from the University of Illinois in which Fritz Heider and Marianne Simmel shared the recent results of their experimental study on apparent behaviour. The article still remains relevant today with a total of 1227 citations from it (JSTOR,April 2nd, 2012) , of which four references are from papers published in 2012 (Web of Knowledge, April 2nd, 2012).
Their research investigated the descriptions of a silent short film made by three different audiences. The first group of participants (34 subjects) was asked to describe the film. The second group (36 subjects) was asked to interpret the movements as actions of persons and to answer a number of questions related to them. The third group (44 subjects) was treated like the second group except that the film was shown in reverse and they had to answer fewer questions.
The film itself showed three geometric shapes: a big triangle, a small triangle and a small circle (or disc), which followed a list of movement patterns in order to illustrate human social behaviour. Besides the three geometric shapes, one bigger structure was situated on the side of the screen. The structure was aimed to represent a house as it had four walls and a 'door' which could be opened and closed by the actors of the film. Illustration 1 shows one frame from the original film,as shown in the original paper (Heider & Simmel, 1944).

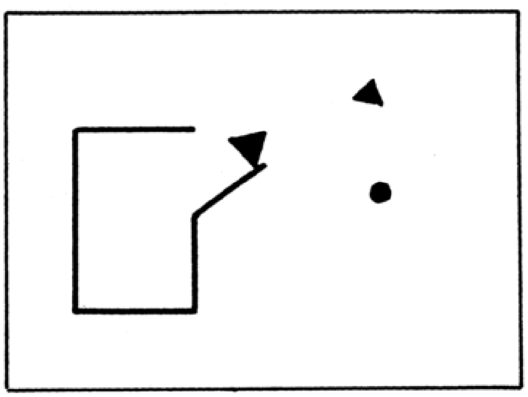


Illustration 1: one frame of the H&S animation. Showing Large triangle, small triangle, disc and house.

The descriptions gathered from the first and third group showed a general tendency to formulate the behaviour of the geometric shapes in anthropomorphic terms, thus describing the shapes as if they were alive and capable of having intentions, goals and emotions, and furthermore generally matching the descriptions of the second group.
A surprising degree of unity was seen in the descriptions. For instance the question “What kind of person is the big triangle” 97% of the participant group used the following terms: “Aggressive, warlike, belligerent, pugnacious, quarrelsome, troublesome, mean, angry, bad-tempered, temperamental, irritable, quick to take offence, bully, villain, taking advantage of his size, picking on smaller people, dominating, power-loving, possessive.“

But there was one exception to this general trend. In the first group one of the participants only used geometrical terms. Since this is an important aspect of the results her full description of the H&S animation is included in this review and goes as follows:

> "A large solid triangle is shown entering a rectangle. It enters and comes out of this rectangle, and each time the corner and one-half of one of the sides of the rectangle form an opening. Then another, smaller triangle and a circle appear on the scene. The circle enters the rectangle while the larger triangle is within. The two move about in circular motion and then the circle goes out of the opening and joins the smaller triangle which has been moving around outside the rectangle. Then the smaller triangle and the circle move about together and when the larger triangle comes out of the rectangle and approaches them, they move rapidly in a circle around the rectangle and disappear. The larger triangle, now alone, moves about the opening of the rectangle and finally goes through the opening to the inside. He (sic!) moves rapidly within, and, finding no opening, breaks through the sides and disappears."

Furthermore, the researchers explain that the interpretations of the actors' movements are strongly connected to the origin of the perceived causal relation. The collision between the small triangle and the big triangle appears to transfer kinetic energy from one character to the other, thus the characters take on roles according to perceived causality effect, one becomes the effector the other the effected, or in their words the characters become causally connected. By watching the film in reverse the origin of the events is changed and so the apparent energetic movement is reversed as well, changing the effected to the effector and vis versa. Putting their theory in pictures, Heider and Simmel use the following metaphor:

> "If one sees two animals running in file through high grass, one will interpret these movements in accordance with other data. If the one in front is a rabbit and the one behind a dog, he will perceive a dog chasing a rabbit. If the first one is a big rabbit and the second a small one, he will not see 'chasing' but 'leading' and 'following'."

The effect of causal connection is observed in the third experiment. In this condition the observers still interpreted the movements of the shapes as human actions although with more variation between the descriptions. Moreover the agency in the descriptions was effectively reversed, for instance the big triangle is attacked and teased by the small triangle and the circle. In this variation the small triangle and the circle are described as being 'evil' whereas in the other two groups it was the big triangle which was described as 'evil'.
This implies that the interpretation of personality-traits of the actors is intimately connected to the interpretation of their movement patterns. The causal origin of the action varies according to the ordering of events and movements. Moreover the perception of the movements provoked the participants to assign motives and character traits of animated beings to the geometric shapes (Heider & Simmel, 1944).

In 1946, Albert Michotte's book "La Perception de la Causalité" was published which was only in 1963 translated into English under the title "The Perception of Causality".
In his work Michotte builds upon Heider and Simmel's work in that he further investigated what

kind of movement patterns would cause the perception of causality. Among many different movement patterns, he investigated one specific kind which he named the *launching effect*. It constituted of one object (A) moving toward another stationary object (B) until they are adjacent, at which point A stops and B starts moving along the same path. Michotte's fundamental observation about this event is that it is in fact perceived as an event: beyond its objective kinematics (i.e., as two objects moving at certain times and to certain locations), a collision appears to happen, wherein A causes B's motion (Wagemans, van Lier, & Scholl, 2006). Or put differently, a single motion is perceived, which is transferred to (or, in Michotte's terms, "phenomenally duplicated" in) the second object.
This effect is caused alone by two simple shapes moving on a projection. Michotte himself put his work into context and wrote:

> "In ordinary life, the specifying factors – gestures, facial expressions, speech – are innumerable and can be differentiated by an infinity of nuances. But they are all additional refinements compared with the key factors, which are the simple kinetic structures"(Michotte, 1950)

Thus the role of perception of causality is not a small one. At the time of the publication most researchers and philosophers thought of causality as a high-level cognitive concept and had a tendency to think of perception only in terms of lower-level properties such as colour, texture, and motion. But Michotte on the other hand makes a point in suggesting that causality might be processed in the visual system rather than in the higher order functions of the brain (Wagemans et al., 2006).

A large body of research followed both of these studies. Fritz Heider continued to work on his theories and later in 1958 published his book "The Psychology of Interpersonal Relations". Which had an large impact in the field of social psychology and perception with 12848 citations (Google Scholar, April 2nd, 2012). In his book Heider proposes the *attribution theory* which analyses how people explain the cause of events and behavioral responses through the attribution of behaviour to a disposition (e.g. attitudes, motives, personality traits) or to situations (e.g. external pressures, social norms, accidents of the environment, act of God, random choice, etc.). He made the argument that people tend to overweight internal, dispositional, causes over external causes, which was later known as the *fundamental attribution error* (Ross, 1977) or *correspondence bias (Gilbert & Malone, 1995;* Jones, 1979, 1990).

Many researchers supported Michotte's contention that causal perception has an innate basis but also suggested that mechanisms of causal perception may – like many other aspects of vision – be a specialized part of our genetic endowment(Scholl & Tremoulet, 2000; and many more) . At the same time, other investigators have stressed the possible role of individual differences and learning in causal perception (Oakes, 1994; and many more).
Specifically Scholl and Tremoulet further investigated how certain simple visual displays consisting of moving 2-D geometric shapes can give rise to percepts with high-level properties such as causality and animacy. They suggested that causal perception meets most of the criteria of modularity (Scholl & Tremoulet, 2000). But neuropsychology still has to conduct research in order to explore the possibility of specific impairments in causal perception with brain damage (Wagemans et al., 2006).

The measurement of causal perception remained problematic since it was so far based on direct reports and ratings of causal perception. When used carefully such probes can yield useful and rigorous data, quantifying the probability of causal percepts over many trials and observers (Schlottmann et al., 2006). Nevertheless Choi and Scholl (2006a) claim that such methods have several problems – especially when they fail to adequately distinguish percepts and higher-level cognitive inferences. As a result, for many years researchers have striven to develop other more implicit measures of causal perception, e.g., based on priming (Kruschke & Fragassi, 1996) , representational momentum (Hubbard, Blessum, & Ruppel, 2001), spatial illusions (B. J. Scholl & Nakayama, 2004), or neural signatures (e.g. Fugelsang et al.,2005). Nevertheless, Choi and Scholl (2006a) argue that such attempts have not yet been successful, and that we are still in great need of new dependent measures which are able to track causal perception implicitly with both rigor and quantitative precision (Wagemans et al., 2006).

3.1 The Characteristics of ASC

Autism spectrum conditions (ASC, aka. autism spectrum disorders, ASD) are defined as a group of pervasive developmental disorders which includes the following five distinguished forms: Autism, Asperger's syndrome, childhood disintegrative disorder, Rett syndrome and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) (American Psychiatric Association, 2000). The prevalence for ASC is estimated to 6 per 1,000. The highest prevalence within this category of conditions is held by PDD-NOS and estimated at 3.7 per 1,000 followed by autism on second place with 1-2 on 1,000. Asperger syndrome was estimated roughly 0. (1er 1,000, and childhood disintegrative disorder at 0.02 per 1,000(Fombonne, 2009; Newschaffer et al., 2007).

In the last 20 years the diagnostic practices and overall public awareness have changed to great degrees which are reflected in the numbers of reported cases of occurrences of ASC (Fombonne, 2009; Wing & Potter, 2002). Although an overall increase in the occurrence of this kind of conditions should not be discarded (Rutter, 2005). Yet the diagnosis of these conditions currently solely rely on the analysis of the external behaviour of the individuals (American Psychiatric Association, 1980; Kanner, 1943; Ritvo & Freeman, 1978; Rutter, 1978). This makes it firstly very hard to monitor with high certainty and is time a time consuming process.

Thus a lot of effort is put into recognising the behavioural characteristics, or sometimes called the red flags of ASC, and in developing more systematic approaches in the diagnosis and intervention for people with ASC. For instance, Wetherby et al. (2004) conducted a prospective longitudal study in order to identify the red flags of ASD. They cross-compared three groups consisting of people with ASC, developmental delays (DD), and typical development (TD) respectively. Two comparisons were published, the first was a comparison between the people with ASD and a group of people with DD and TD (18 people all together) and the following red flags were noted:

(1) lack appropriate gaze,
(2) lack of warm, joyful expressions with gaze,
(3) lack of sharing enjoyment or interest,
(4) lack of response to name,
(5) lack of coordination of gaze, facial expression, gesture and sound,
(6) lack of showing,

(7) unusual prosody,
(8) repetitive movements or posturing of body, arms, hands, or figures,
(9) repetitive movements with objects.

The second experiment investigated the significant differences between ASD and TD groups (but not ASD and DD):
(1) lack of response to contextual cues,
(2) lack of pointing,
(3) lack of vocalization with consonants,
(4) lack of playing with a variety of toys conventionally.

The most commonly reported impairment of childhood autism is a profound disorder in understanding and coping with the social environment, regardless of IQ (Simon Baron-Cohen et al., 1985). Interacting with the social environment requires a large array of prerequisite skills which a child would normally learn during the first years of life. These skills are commonly referred to social communication skills. Social communication is an umbrella term which groups together different theoretical constructs, such as "shared or coordinated attention","intentionality" and "reciprocity", all of which are not directly observable and thus hard to test for in diagnostic environments (Charman & Stone, 2006). The study of social communication is relatively new and the field is still trying to determine what the latent variables are and what good items to measure the latent variables would be. (Charman & Stone, 2006)
In children with ASC the social communication skills are commonly not developed to an functional degree, thus having a negative effect on acquiring related abilities such as understanding shared meaning and overall social awareness. Furthermore it was documented that people with ASC are characterized by a distinct constellation of strengths and weaknesses in parameters of social communication (Charman & Stone, 2006; A. Wetherby, Prizant, & Hutchinson, 1998). The magnitude of this difference in social cognitive problem solving ability is in recent years been documented using modern techniques such as eye tracking, which allows the researchers to see and measure how individuals with autism search for meaning when presented with naturalistic social scenes. The data gathered through these modern approaches is analysed to find reasons about why these difficulties in social communication prevail and how to social communication skills could be tought more efficiently in current intervention methods. (Klin et al., 2003)
The difference between high level cognitive reasoning and the social communication skills is illustrated in this paragraph from the paper of Klin et al. (2003):

> "When the viewer with autism was later questioned, in an explicit fashion, about whether he knew what the pointing gesture meant, he had no difficulty defining the meaning of the gesture. And yet, he failed to apply this knowledge spontaneously when viewing the scene from the film."

Klin and his collegues in the same paper explain that even though adolescents and adults with autism do not display normative reactions to the social communication stimuli as typical young children do, this does not mean that their ability to function in the world is at the very early stage of development. But rather does it raise the possibility that these individuals learn about the social world in a different manner. How this alternative developmental path would look like would be of strong significance, both in clinical and research domains.

3.2 The Theory of Mind and Mindblindness

When the behaviour of another being is described by an observer as being triggered by its thoughts, desires, and intentions, then the observer bases his description on the theory that the being has a mind. The observer assumes that it is able to have thoughts, desires and intentions just like the observer himself. The observer can then be said to have a *theory of mind* (ToM) (Premack & Woodruff, 1978) which is also known as *mentalising (Abell et al., 2000; Castelli et al., 2002)*.
The ability to make inferences about what other people believe to be the case in a given situation allows one to predict what they will do. This is clearly a crucial component of social skills (Simon Baron-Cohen et al., 1985).

A common method to test the ToM skills of a participant are *false belief tasks*. This method is based upon the idea that a person who does not have a ToM would be unable to relate to a specific situation through a third person's perspective. Thus the person would be unable to attribute a belief different than its own to the third person. This case provides the strongest evidence for the capacity to conceive of mental states (Dennett, 1978). Wimmer and Perner (1983) originally constructed a false belief task which assesses a first order false belief ('she thinks x'). Baron-Cohen et al. (1985) reproduced their results with an adapted version of the test which from then onwards was known as the Sally-Anne test.
Other false belief tests like the Ice-Cream story (Perner & Wimmer, 1985) and Birthday Puppy test (Sullivan et al., 1994) examine the ability to attribute a second-order false belief (`she thinks that he thinks x') (Abell et al., 2000).

The Sally-Anne test provided the conclusive data upon which Baron-Cohen based his very influential theory that autism is in fact the result of impaired mentalising, which is manifested in a lack of social insight and impaired communication. Furthermore individuals with autism would lack the ability to *read minds*, a complex group of skills which is used in mentalising, and thus are *mindblind* (Simon Baron-Cohen, 1994, 1995; Simon Baron-Cohen et al., 1985).
Through the comparison of various tasks that differ only in the mentalising component, it was ruled out that mentalising difficulty is due to greater task complexity or lower general ability (e.g. Leslie & Thaiss, 1992; Perner et al., 1989; Sodian & Frith, 1992). Furthermore it was found that people with high-functioning autism can perform perfectly on standard laboratory tests of false belief attribution, but they experience long developmental delays when acquiring the skill and are prone to errors on more advanced tests of ToM (Baron-Cohen et al., 2001; Happé, 1994; Klin, 2000; Roeyers et al., 2001). Thus suggesting that even able individuals with high-functioning autism read minds differently.

An alternative ToM experiment was used besides the false belief task, which was based upon the animation developed in Heider and Simmel's (H&S) study of apparent behaviour (1944).
Firstly it was shown by many researchers that young children would indeed have the skills required to produce anthropomorphic descriptions of the simple ToM animations. Montgomery

and Montgomery (1999) showed that even 3 year old children can detect the intended goal of an animated shape, on the basis of a simple pattern of motion. Furthermore there is growing evidence for the ability to attribute mental states to others, and its development from the second year of life onwards (Bretherton, McNew, & Beeghly-Smith, 1981; MacNamara, Baker, & Olsen, 1976; Shantz, 1983; Shultz & Cloghesy, 1981; Shultz, Wells, & Sarda, 1980).
Thus the H&S animation could be shown to children of 3 years and up with a potential to gain further insight into when the mental state attributions are made by typically developing children and adults. And that is exactly what Berry and Springer (1993) did. In their study they showed the H&S animation and three digitally altered versions of the animation to children of ages 3, 4, and 5 in order to compare them with the results of their previous study with adults (D. S. Berry, Misovich, Kean, & Baron, 1992). In the first altered version of the animation the structure of the animation's characters was disrupted. In the second, the motion paths of the characters were disrupted. In the third alteration, both structure and motion were disrupted. Again the descriptions of the participants were scanned for anthropomorphic terms and compared across the different age groups. It was found that the children (between 3 and 5 years old) were less likely to provide anthropomorphised descriptions in comparison to the adult group overall. Nevertheless significantly less children provided anthropomorphic descriptions when the animation's movement patterns were disturbed. Moreover, the disruption of the structure did not yield a significant decrease in the number of children and adults who used anthropomorphic terms in their descriptions. Altogether the data indicated that children followed the same trend as adults in describing the animations in to same degree of anthropomorphism. Moreover Berry and Springer again illustrate how the motion path has a direct effect on the interpretation of the animation. In succession to the experiment, Springer, Meier, and Berry (1996) showed the H&S animation to 3-, 4- and 5-year-old children, and found that character attributions to the figures in the animation were more differentiated and more similar to adult attributions in the 5-year-olds than in the younger children.

Later research analysed the different degrees of ToM attribution and its connection to developmental delays and ASC was investigated by Abell, Happé and Frith (2000). Their work consisted of a series of ToM animations which illustrated different levels of ToM complexity: Random animations (e.g. bouncing), Goal-directed (G-D) sequences (fighting), and Theory of Mind (ToM) sequences (tricking). Their results indicate that high-functioning children with autism used mentalising descriptions less often than normally developing 8-year-olds, but just as often as children with general intellectual impairment did. However, with the difference that the autism group frequently referred to mental states that were inappropriate to the animation. Even those children with autism who passed standard false belief tasks showed inappropriate descriptions of ToM animations, revealing continuing impairments in mentalising (Abell et al., 2000). A very large body of research was conducted using their ToM animation, which is commonly referred to the Frith-Happé stimuli. For instance, Klein et. al (2008) used the Frith-Happé stimuli in combination with eye-tracking technology to find that indeed the degree of mentalising is reflected in the time the eye concentrates on the stimuli. Thus, the random sequence entailed shorter fixation times where as the goal-directed sequence provoked intermediate fixation times. Which in turn means that the mentalising ability can be measured using non-verbal methods like eye-tracking. (Their paper also includes a large list of references to other behavioural and neuroimaging studies conducted using the same stimuli and a variety of participants.)

3.3 Related Research Projects:

In the following section two recent research projects which are mentioned throughout this honours project are briefly introduced.

ECHOES (2008-2011) is a technology-enhanced learning environment (TEL) which was designed to support the development of social communication and interaction skills. The target group are children with ASC and typically developing children between 5 to 7 years old. (Porayska-Pomsta et al., 2011)
The system consists of a *sensory garden* – a multi-modal 3D environment in which various interactive objects are located – and an intelligent, semi-autonomous virtual character (embodied agent) which inhabits environment. The children are encouraged to interact with the objects in the environment while the embodied agent guides them through a collection of learning activities. These activities are designed around specific learning goals that relate to different forms of joint attention and free exploration of the environment.
The system uses a large (42") multi-touch LCD display as an interactive interface between sensory garden and the children. Advanced computer vision is used to detect where the child is looking at any given point in time (Porayska-Pomsta et al., 2011).
The project as a whole provides a platform for researchers, teachers, parents, practitioners to investigate the individual children's strengths and difficulties in social communication. Furthermore it investigates how technological intervention can be used in this context. (ECHOES website, 2011)

The CLICK-EAST project is a research project of the University of Edinburgh and its name stands for “Computer-based Learning in Children: The Edinburgh Autism Social-attention Trial”. The project build an learning tool for children with autism which aims to bring basic social and communicational skills closer to the children in a fun and engaging way. The application is based on a tablet pc and designed to be as family friendly as possible. (CLICK-EAST Website, 2012)

For more information on these projects, please refer to the individual project websites in the references.

4 The Project

The main objective of this honours project was to develop and evaluate an alternative method for conducting Theory of Mind experiments, specifically for the ToM animation description task. The common approach to the ToM animation description task was studied through an extensive literature review and will be referred to as the *traditional observatory method* throughout this report. The *interactive assessment method* is proposed, which is based on this previous research and effectively extends the traditional method.
Furthermore, a computer based application was developed to implement this new interactive method. Subsequently an experimental study was designed in order to put the new approach into context and test its software implementation.
An interactive variation of the commonly used ToM experiment could potentially benefit the field of psychology, specifically developmental psychology in respect to ASC research and social perception. A large body of work used ToM animations to gain insight into how children and adults attribute mental states to objects (Abell et al., 2000; Diane S Berry & Springer, 1993; Klin, 2000; and many more). Thus far an interactive approach which uses a stimuli similar to the Heider and Simmel or the Frith-Happé animations has not yet been investigated, as far as the review of the previous research could tell.

Both the interactive method and the software were designed specifically for young children with autism spectrum conditions (ASC) as well as typically developing (TD) children in the age range of six to nine years. ToM experiments targeted at children would benefit the most from the interactive approach as it aims to keep the participants engaged and encourages a dialogue between the participant and the experimenter.
This user group also put specific user requirements on the software and the layout of the experimental study.

The development of the interactive assessment method and “Act” (the software implementation) is explained in the following sections of this report. Beginning with the definition of the *traditional observatory method* and *interactive assessment method*.

4.1 The Traditional Observatory Method

The methodology in the research of Heider and Simmel (1944), and Abell, Happé, and Frith (2000) were taken as representative cases for the *traditional observatory method* within this paper. This method can be defined as follows; taking Heider and Simmels method as a starting point. In their research a 2 ½ minutes long ToM animation was shown to three groups of participants. The interactions between the actors were silently observed by the participants who subsequently were asked to fill in a questionnaire. Three variations of the questionnaire were prepared for the individual groups, all of which in essence asked for the personal interpretation of the film's contents. The first group was asked to describe the events of the film in strictly anthropomorphic terms, the second group should simply describe the events (no specific language/interpretation requirements), and the questionnaire of the third group was an adjusted

version of the second, since they had watched the films in reverse compared to the first two groups.
The methodology of Abell et al., differed in the way that they used multiple animations which varied in ToM complexity. Their work covered a broader range of factors and normalised the results in a systematic way, yet the essence of the experimental method remains equivalent to Heider and Simmel's approach.
In both experiments the participants followed a general “observe, describe” action sequence (OD), which is the defining characteristic of the traditional observatory method. There is no interaction between the experimenter and the participants or between the participants and the animations. Furthermore the descriptions are recorded in a single step after the animations were watched through a questionnaire.

4.2 The Interactive Assessment Method

In contrast with the observatory method, the *interactive assessment method* attempts to gain insight into a person's mentalising routine through the analysis of simple interactions between the participant and the animated object. The method implements the participant action sequence “observe, describe, act” (ODA) and repeats the sequence multiple times throughout an experiment session. Thus the general experimental procedure takes the following form.

In the beginning the participant is introduced to the ToM animation and the environment in which the characters of the animation interact (animation environment). Through observation the participant starts to familiarise with the general rules of the environment and make their first inferences about the underlying rules. These initial inferences are then extracted by requesting a description of the introductory animation - insofar the method is identical to the traditional observatory method.
In the next step the participant is requested to interact with the characters of the ToM animation environment. This request is also called the *interactive challenge,* and represents the one of the core features of the interactive method. The formulation of the interaction challenge could have a direct impact on the degree of mentalising that is done by the participant.
For instance setting the challenge as “How can we make Alice happy?” with options “Give her a cake” or “Steal her bike”. This challenge primes the participant in believing that one of the characters has a name and is called Alice (which might not be obvious from the stimuli) and literally tells the participant that the character is able have emotions such as being happy. Furthermore it is a rather complex task as the participant needs to infer multiple things. Firstly that Alice is not happy at the time the question is asked, and that giving Alice a cake will make her happy, whereas stealing her bike will make her sad.
In contrast the interactive challenge could also be formulated in a more objective way such as “Pick one” with options “Cake” or “Bike”. In this case it is not at all clear what either option will entail. Hence the challenge is reduced to a relatively simple choice of preference.
Thus the interactive challenge can be used to set the frame for the interpretation of the character's reaction, effectively promoting or demoting mentalising through priming.
Moving on through the action sequence of the interactive method, the option chosen in the interactive challenge than has its effect on the animation environment and its characters. This provokes the creation of new theories and inferences about the reactions to the interaction and

further expands the participant's understanding of the environment.
Then a new description phase extracts the new inferences and ToM models constructed in the interactive step and consequent observation step.

The interactive assessment method can be implemented in many different ways, ranging from a completely immersive environment in which the participants are embodied and can freely move around and interact with other characters, to a simpler turn-based environment in which the participants take on the role of directors by observing the behaviour of the characters before taking decisions for one or more characters of the ToM animation.
Although it should be considered that the method's efficiency is directly related to the complexity of the interaction (degree of realism) and the target group's mentalising ability. Naturally complex social interactions in real life are often confusing for a target group with difficulties in social communication and mentalising. In this case, an immersive environment would only be able to highlight that there is a mentalising problem by means of confrontation but it would be difficult to extract as to why this is the case or where the mentalising path deviates from the typical path.

The instantiation of the interactive assessment method described in this paper was developed in order to fit the requirements of 6 to 9 year old children with ASC and typically developing children.
Recent research has shown that children with ASC describe the classic ToM animations (Heider and Simmel / Frith-Happé animations) in similar terms as typically developing children as long as the animation only shows goal directed behaviour. The main difference although was found as soon as mentalising was necessary to describe the behaviour accurately. The children with ASC either didn't use mentalising terms at all or they didn't match the events in the animation. (Abell et al., 2000; Berry & Springer, 1993; Klin, 2000). This suggests that a more abstract and controlled environment would be more suitable for this target group, since other distracting features of the stimuli are removed and the participant can concentrate at the social behaviour which on its own is already hard enough.
Furthermore the interactive assessment method could potentially provide a more detailed insight into how the mentalising process evolves over time. The interactive challenges could be carefully adopted to slowly prime or train the participant in using a ToM in their interpretation of the ToM animation. The change of language used in the descriptions would then indicate a change in the interpretation of the ToM animation.
Besides the interactive assessment method is attractive for experiments with children as it is gives the participants control over the story of the ToM animation and thus has a small gaming element to it. Additionally through the iterative description process the experimenter can adopt the questions to fit the child's language skill level.

4.3 User Characteristics

The user group or target group, as mentioned before, were 6 to 9 year old children with ASC and typically developing children. Through the review of the literature on autism spectrum conditions a collection of user characteristics were found which needed to be considered when designing the the software and the experimental study.

Young children are full of energy and sometimes very hard convince to sit still for even a few minutes. Attention is sometimes on everything but the task at hand as often task are found boring relatively quickly and interest is lost (personal experience).
Specifically young children between 5-9 years of age demonstrate significant differences in their emotional regulation, emotional recognition as well as categorisation abilities (Tartaro & Cassell, 2006), and also thematic interests, communication and literacy skills (Porayska-Pomsta et al., 2011).
Children with ASC have very strong thematic interests and preferences of colour up to the degree that a task is refused if it falls outside of their domain of interest (Dr. H. Pain, personal communication, November, 2011). One example was given from the Click-East project which ran on a tablet pc. Even though the software was well designed the single button on the tablet caught the attention of many children who inevitably pressed it and minimized the running application (Dr. H. Pain, personal communication, October 13, 2011).

Moreover fine motoric skills are often impeded through ASC. Repetitive movements are often used as coping mechanism to regulate emotional stress (Levy, Mandell, & Schultz, 2009; A. M. Wetherby et al., 2004) .

For the whole list of ASC characteristics please refer to section 3.1 “The Characteristics of ASC” in the literature review.

4.4 User Requirements

Developing software for children in general puts multiple requirements on the way the user interface of the software is designed. These requirements also needed to be considered while designing the exploratory study since the experimenter is interacting with the user while the software is being used. Following the user characteristics from the previous section the system needs to be:

- Easy to use; Big buttons which are easy to see and click on. The action of the button should be easily predicted, through a clear label and intuitive positioning. The hardware user interface should be easy to use and avoid distractions (shiny LEDs on the keyboard, Post-its on the screen etc.)
- Visually attractive; A balanced set of colours should be used in order to avoid a refusal to participate due to strong colour preference conflict. The colours green and red should be used sparingly due to the possibility of colour blind participants.

- Easy to understand; Long and complicated text instructions should be avoided. Short and direct directions of what to do are preferred. Text should be easy to see (not moving around or on top of complex images).
- Robust; The user should not be able to leave the application by pressing random buttons or clicking. The software needs to be prepared to be used in totally unpredictable ways. Trigger happy users should not be allowed to rush through the software too easily.
- Fun to use; Penalties and failure messages should be avoided at all cost. Positive and encouraging feedback needs to be supplied to the user in order to keep motivation up and emotions regulated. “Try again!” instead of “No. Your are wrong!”.
- Provide two different interfaces: one for the participant and one for the experimenter; The participant should not be able to modify the experiment settings. Experimenter needs specialised interface to set up the experiment before the participant comes.
- Portable; Location is uncertain which includes internet access.

4.5 The Designing Process:

The project went through several development phases before Act came to be. The red string that runs through the project's development was the idea to add an interactive aspect to ToM animation experiments.

An initial draft of the interactive assessment method was made and with it an ambitious computer based application that would implement the method. After a theoretical evaluation of the first design, feasibility was strongly criticised and the second design was developed upon a revised interactive assessment method. This second design evolved into Act and underwent several testing stages before it was tied into a experimental setup for the final field test.
These design stages are explained in detail in the following sections.

4.5.1 The First Draft

The review of the literature motivated a more complex ToM experiment, more complex than only observing and describing a ToM animation. Especially Wetherby and Wood's work in 2004 on identifying the red flags of autism as well as Heider and Simmel's work in 1944 about ToM animations were strong influences for the design decisions in the initial design.
With a strong user centered design approach the first draft aimed to firstly make the software fun and engaging and secondly train the user in recognising patterns of social behaviour in abstract figures.

Looking at the work of Heider and Simmel (1944) – one single ToM animation was created and shown to the participants of the study. The actors of the animation, as previously explained, were simple geometric shapes which interacted with each other. The initial draft aimed to investigate how a participant would behave when put into the situation together with the geometric shapes.

Perhaps the reactions of the participant would give some insight into how and to what degree the social situation is understood.
It should be mentioned that no interviewing or description process was planned for this implementation. The participant's understanding would be gleaned of their behaviour in the application and a rough eye tracking system would be implemented through video footage of the participant in order to gauge their attention level towards the computer screen.
Interactive social challenges would play a central role in this implementation. A social situation would be introduced and the participant would be invited to engage with the situation. The interactive challenges could take on two forms: direct or indirect.
In the direct challenge the user is directly challenged to take action, this could take the form of a group of characters rapidly approaching the user in a menacing way, or a prisoner sitting in their cage. The social activity is directed towards the user alone and their reaction has a direct impact to the situation, ie. the menacing group surrounds the inactive user or the released prisoner is thankful for the user's action.
In contrast, the social activity in indirect challenges would be directed to one of the characters or an object in the environment. This kind of challenge is judged more complex as the participant first has to recognise the social situation, then find how they could potentially help, and find a way to do so. Such challenges would take the form of helping to dissolve a conflict (e.g. helping a character escape) or help a group of characters achieve a common goal (e.g. pushing a ball into a hole).
This aspect was mainly inspired by Michotte's work on the *launching effect* and the research that followed. As it would heavily rely on the recognition of social patterns and attributing a quality to movements of the geometric shapes.

In order to provide more control to the user and to not just throw them from one social challenge to the next, the environment was planned to take the form of a maze with multiple chambers that waited to be explored. Through this layout the user would explore the maze and solve the challenges in their own pace and it would add a more game-like feeling to the application potentially making it more fun to replay. Another reason for implementing the maze was in order to monitor repetitive behaviour in the participants, which is one of the red flags of ASC as put forward in Wetherby and Wood's work in 2004. The maze would remain unchanged between different runs, but the chambers would randomly change their location in the maze. The recorded user movements could be used to create heat-maps in order to identify heavily used routes and movement patterns. Naturally this aspect would only be observable in a very big maze and a continuous study would have to be done to compare the moving patterns in all the replays. Additionally a cross user study would have to be done to compare the movement data in order to rule out simpler paths (paths which are simpler to navigate through or more attractive and are therefore more popular).

Upon completion of an interactive challenge the participant would be rewarded and then free to explore the maze further.

This first draft of the system would take a rather different role than the second draft. It would be a training software that would be used multiple times and a study would have to be done over a long period of time.

Through the evaluation of the feasibility of the first draft it became evident that there were multiple hard problems that would have to overcome.
Firstly the assessment of the participant's actions would be a very hard task. Distinguishing

between intentional movements and random uninterested movements of the participant would be a challenge, even with the rough gaze tracking through the video footage. But reading from those movements whether the participant fully understood the social situation would be impossible. The only way to assess understanding would be through specific user actions, which leads to another problem. How can it made obvious to the user that these actions exist without priming them and telling them the answer to the social challenge? Assuming the available actions were made obvious and the user would know about them, how could the user be prevented to 'game the system'? (just pressing the buttons in order to progress but without the understanding of what it entails, a common problem in adaptive learning environments, for instance in: (McQuiggan, Rowe, Lee, & Lester, 2008))
Additionally time was judged to short in order to implement the application in time for a formative evaluation and experimental study. Thus the design was trimmed down and the interactive assessment method was re-evaluated.

4.5.2 The Second Draft

Through the evaluation of the first draft it became clear that a more robust assessment method was required and with it a better defined user interface. Hence the interactive assessment method was revisited, thought over and extended, and thereupon a new software implementation was designed.

Firstly the traditional observational method was analysed and defined as such. This laid out the foundation for the second draft of the interactive assessment method. In contrast to the first draft it primarily concentrated on the ToM assessment methods used in previous research. The description phases in the traditional experiments had proven to return useful results (Abell et al., 2000; Heider & Simmel, 1944) therefore it was kept in the interactive method as well.
In essence the traditional ToM animation experiment was taken and split into smaller animations each followed by a description phase. One experiment would run through a collection of these smaller ToM experiments. Consequently a criteria for choosing the next animation was needed, which took the form of *decision phases* – here the participant would interact with the actors of the ToM animations which in turn would affect which animation is played thereafter. Although it is a relatively weak type of interaction, it facilitates a relatively easy way to dose the level of priming through the design of the decision phase. For instance one could use the method to build a training environment in which the decision phase primes the user as much as possible into thinking in anthropomorphic terms in order to familiarise the user with the practise of mentalising (Dr S. Fletcher-Watson, personal communication, February 6, 2012).

The experiment's action sequence can be defined as many consecutive iterations of the "observe, describe, act" (ODA) sequence and effectively extends the traditional observatory method which takes the form of a single execution of the "observe, describe" (OD) sequence.

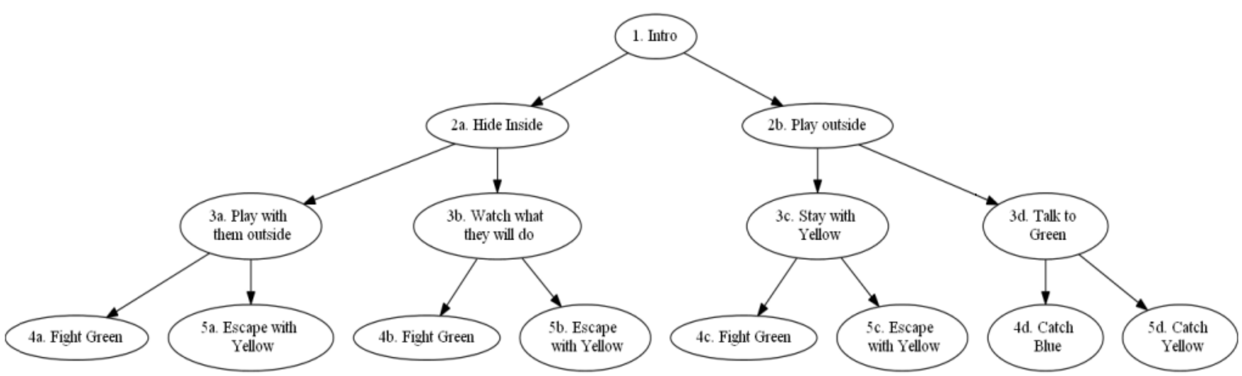
The updated notion of the interactive assessment method was directly implemented into a computer based application which a bit later was named "Act". The ODA principle was put into practice in the following ways:

Observe – Eight stories were created each of which consisted of four short ToM animations. Together the eight stories formed a story tree (see Graph 1 below) as all of them shared at least one common node: the root node which represented the introductory ToM animation.

Throughout the story the complexity of the ToM interactions increased. Thus the introduction animation is the simplest of all and the ending animations (the fourth) were the most complex in general. Additionally, when played the application would wait two seconds before starting the playback of the animation and it would freeze the picture for two seconds in the end. This was done in order to prepare the participant for the transitions to and from the animation screen.

Describe - After watching one animation the participant would be asked to describe it before progressing further down the story tree. In the software this took the form of a simple prompt screen as it was assumed the software would only be used in an experimental environment. This directly addressed the issues found in the first draft. The user interface is well designed – two actions are presented in each *decision phase* (see below for definition) whose decision can be reasoned about. The participants are interviewed throughout the experiment in order to gauge their mentalising levels and understanding of the consequences of their actions. The experimenter was expected to ask the appropriate questions to the participant in order to extract the description of the animation but also minimise anthropomorphic priming (the experimental study is thoroughly explained in section 5 of this report).

Act – After the description phase the participant progressed to the decision phase. Here the participant would be asked to decide what one of the actors of the ToM animation should do and was introduced to three options in form of three clearly labelled buttons. Two of the buttons were labelled in relatively mild anthropomorphic action based terms, like “Play” or “Hide”, and would play the appropriate clips when pressed. The third button always carried the label “I don't know” and acted as an escape button. It was hoped that those participants who did not understand the anthropomorphic language terms of the other two buttons would press this button and through that would effectively demonstrate their difficulty in understanding. The “I don't know” button led to a simpler decision problem in which the participant only had to pick one of two buttons which each had the picture of an actor of the ToM animation on them. The two buttons on this alternative decision phase led to the same animations as the two anthropomorphic buttons.



Graph 1: The first version of the Act story tree; Nodes represent decision phases and their corresponding ToM animation. The user starts at the root node (1.Intro) and moves down to the leaf nodes. The application closes after the description of the last clip was given.

In general one run through the software would show one whole story, which consists of four ToM animations and passing through four description phases and three decision phases. In order to see all the different endings the application would have to be run through at least eight times.

This provided some motivation to play through the application again to see all the variations. The application can be scaled up relatively easily by adding options and more content. The results of the experimental study although are not as strictly correlated with the size as it was in the first design. The workload in terms of preparation should still not be underestimated however – with two options on each level of the story tree the number of ToM animations grows in the order of 2^n. The more options, the faster the tree grows per level. In general, it grows in an exponential rate of k^n were k is the number of options and the n is the level (depth) of the tree.

4.5.3 The First Prototype – Act

The implementation of the user requirements followed the basic human-computer interaction (HCI) guidelines and favoured simplicity and robustness.
The user requirements influenced many decisions in the design of Act's ToM animations. The style of the animations closely matched the Heider and Simmel (H&S) animations, in that simple geometric shapes interacted with each other, without sound, and in a very simple environment.

Some notable differences although were firstly that the background colour was chosen to be black in this project. This decision followed the practises in early studies in the field of causal perception and perception of animacy (Bassili, 1976). Secondly the actors of the ToM animations were equally sized white squares which were surrounded with a coloured aura (blue, green or yellow) in order to make them easily distinguishable from one another (see illustration 1). The black background was found to give a better pop-out effect than a white background such that it was easier to concentrate on the actors.

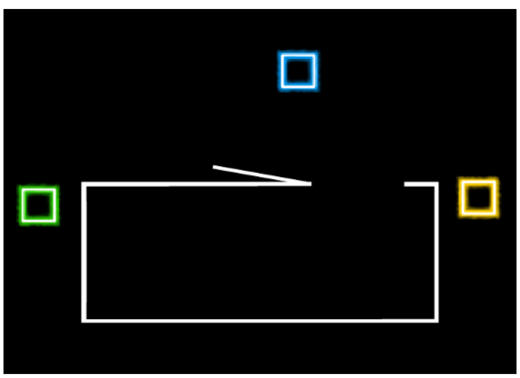


Illustration 2: Screenshot of the "Talk to Green" animation.

In the H&S animations the triangle actors were often described as aggressive males and the small square as frightful girl. Following their work, a large number of studies further investigated the link between the perception of animacy and different features of the stimuli, and consistently found that the movement patterns of the actors have the strongest effect on perception. Nevertheless shape and colour of the actors had significant effects on the test cases which only used single frames of an animation, thus eliminating the movement percept. (Berry & Springer, 1993; Hubbard, 2004; Rimé, Boulanger, Laubin, Richir, & Stroobants, 1985; Scholl & Tremoulet, 2000; to name a few). Following this research it was decided to prevent this priming through the appearance of the actors and to equalize the shape and size. The colours were chosen to be relatively neutral and believed least likely to provoke a premature classification of the actor's role.
Additionally a house, or open and closable box, similar to the one used in the H&S animation was included in the animations with the difference that it was located on the bottom of the scene with the door on the top side. This was done in order to have a relatively large field on top of the house where interactions could take place, and to be able to fade the actors in and out on the sides on the screen as is common practise in animation movies.

In order to maintain a consistent design in the whole application the user interface was designed in a black-board and chalk drawing theme (see illustration 2). This way the black background could be kept throughout the whole application. The contour of the buttons were drawn in chalk and changed colour depending on their state. The black-board theme was introduced as it was believed that it would seam familiar to most children and as such worked well with the watch, point-and-click interface of the software. Consistency was important as it would reduce the emotional stress originating out of the task of learning how to use a new software. If the style would change drastically between the different phases it would most likely confuse and frustrate the user.

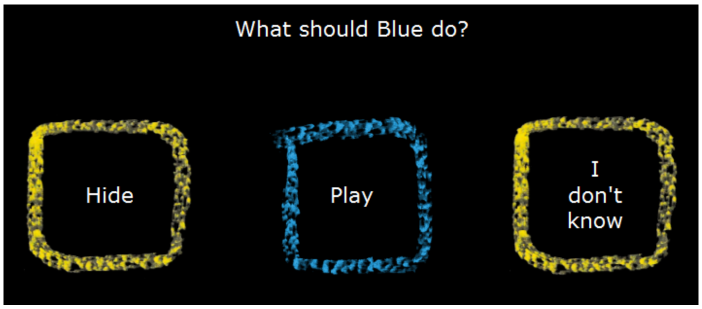


Illustration 3: Screenshot of the decision phase, showing the black-board theme. The "Play" button is currently hovered over.

The buttons of the interface were big and easy to see and select (size: 175x170 px), and selection was made clear by a change of colour when hovered over and an increase of thickness when pressed. Again the colours blue and yellow were chosen as they are opposites and so provide an optimal distinction factor, and it matched the colours used in the animations. The position of the buttons varied such that the general surface area of the buttons did not overlap from one screen to the other. Thus preventing rushing through the description and decision phases by repeatedly clicking on the same location.
Another feature that address the emotional regulation of the user was the “I don't know” button, as it made failure as such impossible. Even when the task is not understood, the user can still continue using the software by choosing this option and on the alternative decision phase choose one of the options.

On the technical side, Act was implemented using the Flex application framework (v. 4.5) and the animations were created using Flash (Actionscript 3.0). This made the application very easy to use on every platform. All the application requires to run is the newest version of Adobe's flash player (v.11 upwards) which is usually installed with any up-to-date internet browser. The application was tested and worked with Internet Explorer, Mozilla Firefox, Safari and Google Chrome; although the full-screen feature of the Google Chrome browser was judged best for this purpose, and Internet Explorer did not support one of Act's features (disabling right clicks).

Robustness of the system was assured in several ways. Firstly the software is always run in full-screen mode, removing all other distractions and also all other possible ways to close the application. Additionally, it was made impossible to skip an animation by clicking on the screen as the mouse pointer was hidden during the playback of the animation.

Right clicks were deactivated throughout the application, since a flash menu popped up on every right click through which the settings of the player could be modified. As this was highly undesirable, the wrapper class which included the Flex/Flash file into the browser had to be modified. This was done with the help of a javascript module specifically developed for creating custom menus in Flex/Flash applications, see reference for more information (Florio & Uza, 2007) .
In order to assure robustness on a hardware level it was decided to run the application on an external display and only provide a computer mouse as interface to the user. By removing the keyboard out of the users field of view the whole setup appears clearer and offers less opportunities for distractions. In terms of robustness, removing the keyboard also removes the possibility to press break-out key combinations, like Alt+F4 or Ctrl+Alt+Del, or other unpredictable keystrokes and their effects.
Since Act was developed in Flex the application could potentially be ported to a tablet pc, which would make the interface a bit more intuitive but potentially less robust.

Additionally Act also recorded the experiment ID as well as all the decisions of the participant. Naturally it was recorded which option was chosen in the decision phase, with the specific motivation to capture how many times the “I don't know” option was chosen. Furthermore the number of replays of each animation was saved as well.
Act conveniently featured a setup screen for the experimenter where the experiment ID could be entered and the experiment could be started. It also included a text area in which the gathered data could be inspected, but it mainly served for debugging purposes.
After the experiment is done, i.e. when the participant presses the “have a break” button, a save dialogue will open on the experimenter screen and the collected data is compiled into a text file and saved on the specified location.

4.6 Software Testing and Changes:

The design and implementation stages of Act's development were followed by three testing stages, which are explained in the following section.

The first general evaluation of the computer system was done with the help of a focus group in an informal meeting where the software was briefly presented in its unpolished state. The group comprised members of the Learning and Adaptive Environments Research (LAER) lab which was part of the School of Informatics' Institute for Language, Cognition, and Computation at the University of Edinburgh. The members were all academics and research students and were friendly enough to take a look at the general quality and some high-level design aspects of the system.
It should be noted that the version of Act that was discussed was missing some of the robustness features, the third option in the decision phase was labelled “?”, and it was still lacking the final black-board theme, both of which were added at a later point in time.
After the software and the idea behind it were explained to the focus group, Act was demonstrated and critically assessed. For this purpose a questionnaire was put together (Appendix A), which covered basic questions about the appropriateness of the animation titles (decision labels), the level of Anthropomorphism, the expressiveness of the events in the animation and the speed or fluency of the animation. Additionally the group was asked to write

down what happened in the animations that were shown and add any comments they had about that specific animation.
As time was very limited in this session only two ToM animations were shown, there were the introduction and the “Play outside” animation (right choice of the first decision).

Even though the testing session was relatively brief it provided a lot of constructive ideas and made some bugs of the software clear.
The bug was found in the animation selection algorithm due to a mistake in the tree arithmetic, but only became obvious through the demonstration of the software.
The group provided general feedback concerning the representation, indicating that the colours of the characters' aura were not strong enough, and that the delays when starting and finishing the animation should be extended. Moreover it was reported that the door behaved illogically in the clips as it did stood open to far after the introductory animation. Through a long discussion about the decision labels of the animations it was agreed that some of the labels simply didn't fit the animation, as in the “Play outside” animation as the characters ended up playing inside the house. And it was highlighted that marking a button with “?” will most likely trigger some curiosity clicks instead of acting as an help or escape button.

All the feedback was gathered and all the ToM animations were revised for illogically moving doors, the colours of the characters were made more intense, the delays extended, the movements in general were smoothened out, and the decision labels were simplified. Naturally the animation selection bug was fixed as well. Furthermore the black-board theme was implemented in the whole application and the “?” decision label was changed to a more self explanatory “I don't know”.

4.6.1 The Academic Testing Session

The second testing session was done together with Dr. Helen Pain and Alyssa Alcorn (PhD student and research assistant). The questionnaire was the same as in first testing session and again used order to collect ratings, more detailed comments and discuss the new decision labels. In this meeting 10 of the 15 animations were watched, described and criticised. Furthermore the overall user interface and the questionnaire itself were assessed.
The main points of criticism were the lack of a replay function and a door inconsistency was across several animations was highlighted. General minor comments were given about the fluency of the animations and the delays needed to be extended by half a second as well.
The replay function was implemented in two ways. The default number of replays, meaning the number of times the animation is played without any action from the user, can be adjusted by changing the value of the DEFAULT_PLAYS variable in the code. Additionally a replay button was added to the description screen. The reasoning behind implementing the replay function was that the participants should not be tested on their working memory, so on how much they remember of the ToM animation, but on their interpretation of the animation. Thus being able to play the animation again is only advantageous.
The door inconsistency depended on how the animation was interpreted, specifically how the house was seen. If it was assumed the house was pictured from the top, or from the bird's eye perspective, then the door would behave as expected. But if the box was not seen as a house but rather as a cellar with a trapdoor, looked upon from the side, then the door would be perceived to behave erroneous because it did not obey the gravitational pull. Then the door would hang in mid

air sometimes and not fall shut as one would expect from a normal trap door.
In order to remove this effect however the box (or house) would have to be repositioned to the left side of the screen in order for the door to not break the law of gravity, which in turn would mean that all the animations would have to be remade. Firstly since the time was lacking to modify all of the animations in such a profound way and secondly that such a change would have obscure some of the events in the animations, the animations were left unchanged in that respect. Besides it would be interesting to see how many participants would perceive the environment in this way, which would provide some insight into their causal perception and mentalising process. Putting it into context, if even in such an abstract world of self propelling geometric shapes the law of gravity is still assigned to objects it must be proof that at least a minimal amount of mentalising takes place.

On a side note it was also commented that black-boards are not actually that common in Scottish primary schools and treatment centres, but that mostly whiteboards or smart-boards were used nowadays. The blackboard theme was kept nevertheless as at least some children will have seen a blackboard or have drawn with chalk at some point in their lives.
Also some of the formulation in the formal software evaluation questionnaires were discussed and changed in later versions of the questionnaire.

4.6.2 The Expert Testing Session

The third testing iteration was done through two discussion sessions with professionals from related fields in order to gather their professional opinion about the project. A small questionnaire was prepared for this occasion (see appendix B). The questionnaires were updated according to the criticism from the previous session and also since the meetings were limited to half an hour each, only two sample stories (or runs) were chosen to be analysed by the professionals. These two stories were judged by the animator to have the highest and the lowest degree of anthropomorphism in the interactions. The expert would use and explore the software and then fill in the questionnaire. This meant ranking the appropriateness of the decision titles, the level of anthropomorphism in the decision titles, the appropriateness of the content for children, the level of anthropomorphism in the animations, the quality of the animations, the quality of the artwork, and the robustness of the system. Additionally some design decisions were discussed. The experts were only told why the two runs were chosen after they had rated them.

The first meeting was with Dr. S. Fletcher-Watson (Nuffield Research Fellow, Moray House School of Education), project leader and chief investigator of the clinical trial of the Click East project. Her practical experience in developing technology for children with autism was invaluable in the evaluation of the application.
In general the current version of the application was found appropriate to be used by children. It was highly robust and the content of the animations were also found appropriate for the age level. The later was a concern as some amount of violence as well as affection were shown in the animations. The updated decision titles were good in anthropomorphic dosage, since they were brief and to the point. The priming effect of the decision title itself should be seen as a drawback from using the interactive assessment method. Furthermore it was confirmed that, with the added replay functionality, the description task was in reasonable bounds in terms of memorization.

Moreover it was decided upon playing the animation only once and then moving on to the description phase, as it would provide a clear split between the replays. While the user is in the description phase the replay option can always be chosen if desired.
Changes were recommended in the phrasing of the prompt message in the description phase. “What happened in the animation” was found to technical for children between 6 and 9, instead either “film” or “clip” should be considered. In the end “film” was used.

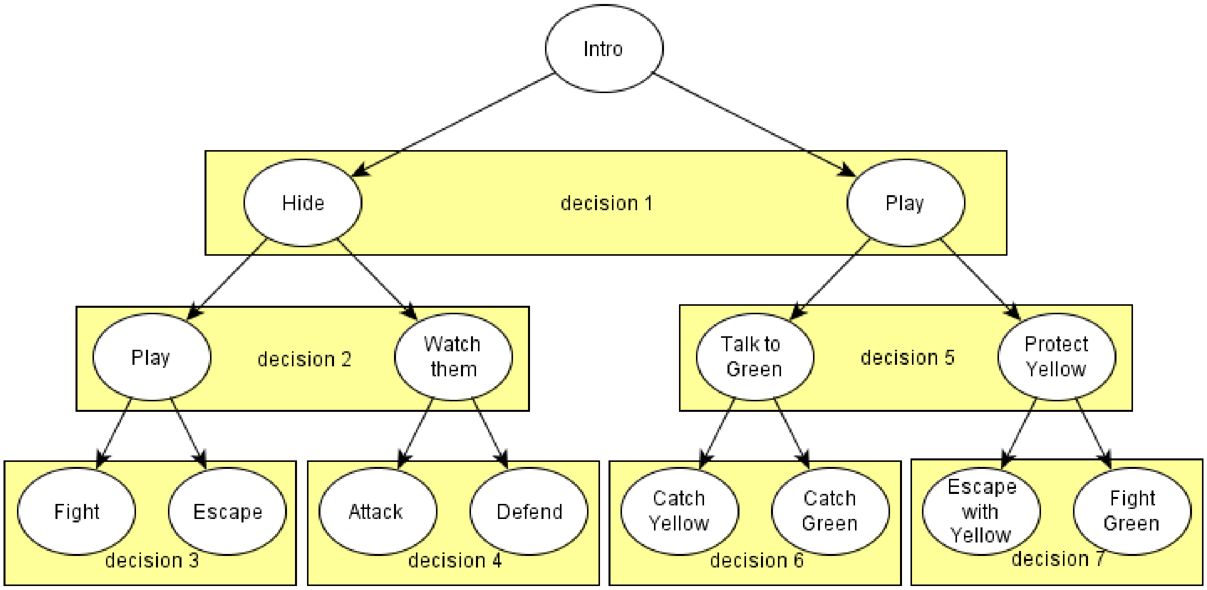
Furthermore, the addition of a “Play again” screen was agreed upon, which provided a fluent transition from the last animation to the end of the story and made it obvious to the user that there was a choice of seeing another story or “have a break”.
And finally the blackboard theme was commended.

The second meeting was with Dr G. Rajendran (Senior Lecturer, School of Psychological Sciences and Health, University of Strathclyde), principal investigator of the ECHOES II project. His insight in developmental psychology and practical experience were of great help in putting the project into context and how to proceed with the experimental study.

The general ratings of the two stories classified the animations as highly appropriate in every way and some aspects of the animations were discussed. The social interactions were found to be good in terms of using moving in circles to illustrate dancing and happiness, the hitting animation correctly conveyed aggression, and the exist animations were found to provide a smooth transition and made the ending clear.
Additionally the experimental study was analysed, which is further discussed in section 5 of this report.

4.6.3 Summary

Summarising, the three testing phases challenged the design decisions of the project but the open discussions helped improve the software. Some mistakes, especially those related to the animations themselves, were only recognised through these testing phases. Act was extended by two features, the animation replay function and the “Play again” screen, and the ToM animations were greatly improved through smoother movement paths and better door interactions as well as decision titles. The updated story tree is illustrated in graph 2 below.



Graph 2: Act story tree – final version; The nodes represent the decision labels in each decision phases and their corresponding ToM animation. The user can play each animation as many times as they desire and also when the last description was given, they have the option to start over.

5 The Experimental Study

To recapitulate, the purpose of this honours project was to define and implement an interactive approach to the ToM animation-description experiment. The interactive assessment method was defined and finally implemented through a purpose build computer based application – Act. Through several testing phases the software optimised and found suitable to be used in a experimental study which would determine the suitability of the interactive assessment method. In the following section the materials of the experimental study is further explained, these are Act's ToM animations.
Subsequently the methodology is presented, expected results are briefly described and the evaluation of the gathered data is put forward.

5.1 Materials

The experimental study is mainly based upon Act – the software implementation developed in this project. The software is used by one participant at a time, who is accompanied by the experimenter. Act is essentially a collection of ToM animations bound together through a story tree and presented in a child friendly interface.
Detailed descriptions of Act's ToM animations (which are used as scoring guidelines) can be found in Appendix X.

During the design of the ToM animations it was aimed to model an increase of the complexity throughout the story tree, i.e. the further down the story tree the higher the number of events eliciting ToM reasoning or the more demanding the ToM task.

The average duration of a ToM animation was 16.57 seconds (st. deviation of 4.00s). The root node of the tree represented the introductory animation and all other nodes represented decision phases leading to other animations. Every node in the first three levels of the story tree had two children. On a single run through the application the user would only see four of the fifteen animations, thus eight play-throughs would be necessary to see every story variation. A full story is composed of four ToM animations and on average has a duration of 66.83 seconds (st. dev. = 4.49s).
Table 2 shows the number of events over time relative to the branch of the story tree. These result illustrate that the animations' event density is highest in the second level of the tree, i.e. in the clips following the second decision phase. Since the complexity of the ToM interactions is highest in the third level of the tree, the number of events goes down. For example, the “Play” animation of the decision 2 (see graph 2 in section 4 for reference) shows how two square are playing outside the house and how the blue square square first approaches the group, then communicates with one of the other squares, then plays and then talks to another square who in response hits the blue square. The subsequent clip “Escape” in decision 3, shows how the green and the blue square get into a fight and the yellow square coaxes the green square who is then locked in the house.
Thus the transition between the second and the third level of the tree changes from complex interactions to mentalising events.

Tree level	Average #events / time per level	St. Dev.		left tree	right tree
0	0.14	0.00			
1	0.20	0.10		0.28	0.13
2	0.29	0.05		0.25	0.33
3	0.21	0.04		0.22	0.21
			Average	0.25	0.23

Table 1: ToM animations analysis: number of main events over time. Level 0 represents the root node.

5.2 Methodology

5.2.1 The Experiment

All the experimenters were assured to have the legal requirements to work with children through the Disclosure Scotland initiative. The ethical approval of the ECHOES project covered all foundation studies conducted by students, under which terms this experiment was covered. Additionally an information sheet and a consent form were given to the parents, and an assent form was read aloud to each child before each experiment (see Appendices C,D and E respectively). Upon agreement to participate the forms were signed and the child was once more explicitly asked for permission to audio record the experiment session.

After this formal procedure, the experiment proceeded in the following way.
Act was started on a laptop which was connected to an external monitor and a computer mouse. The experimenter entered the participant ID (or experiment ID) and started the experiment by clicking on the “start experiment” button. Then the program was moved to the external monitor so that the participant could see it, and the mouse was handed to the participant. An effort was made not to let the participant see the experimenter's setup screen, in order to avoid some of the examination stress connected to knowing that it is an experiment one is participating in.
To begin, the participant was asked for permission to audio record the session and upon agreement the experimenter explained what the computer application is about and how it works. It was also mentioned throughout the experiment that the participant could stop at any time if they desire. Finally the participant was encouraged to click on the screen which would start the first animation.

Throughout the experiment, the experimenter followed an interviewing script which was compiled in order to maintain a consistent level of quality and similar wordings of the questions across the experiments. (The interviewing script can be found in appendix F.) In general all the text was read out aloud when it appeared and when unclear answers were given a question was formulated from the words of the participant, in order to avoid priming. In each description phase the participant was asked to describe why they chose their option, whether that was what

they had expected, and what happened in the animation.
The experiment was run as long as the participant wanted and in the end the data was saved on the laptop through the save dialogue that appeared on the experimenter's laptop screen.

5.2.2 Data Categorisation and Scoring

All the data is collected as described in the experimental method. Before it was possible to score the descriptions, the audio recordings of the experiment sessions had to be transcribed and categorised according to who was describing which ToM animation and whether it was an actual description or a decision dialogue, or expression of confusion or frustration.
In order to be able to effectively rate the descriptions of the ToM animations a scoring criteria and a 'gold standard' was needed. The designer of the ToM animations compiled a list of main events as well as highly anthropomorphised descriptions for every ToM animation in the experiment into a document which was used as the 'gold standard' in the scoring process (see appendix H).
The paper of Abell et al. from 2000, was taken as a starting point to design the scoring criteria for the descriptions of the ToM animations (see appendix G). The concepts of description accuracy and type were kept although the type was expanded by one category: priming. The description accuracy of one ToM animation the average accuracy of all the descriptions of the animation's main events. The type of the description is given by the highest level of mentalising recognised in the event descriptions. The priming type was used in event descriptions which used the terms in the decision label of the animation, for instance in the ToM animation with the decision label “Hide” every occurrence of the term “hide” would be considered of type “priming”. Every description of the “priming” type was ignored as it did not reflect the mentalising process of the participant, the child is being primed by the software.
Additional to Abell et al.'s scoring criteria specific rules for scoring the decision descriptions (transitions) were introduced. Three scores were associated to a description of this kind: usefulness, strength, and type. “Usefulness” took a boolean value (true/false) and indicated whether the description carried any sort of useful information. This was mainly used to distinguish between simple decision exclamations (saying aloud which button was pressed) and more elaborate descriptions. “Type” indicated the category of the description. This could either be any of the ToM description categories or “stress”,”engaged”, or ”questions”.
“Strength” was similar to accuracy in the normal descriptions, with the difference that it captured how strong the expression was. For instance a decision description could be rated “True, 2, Stress” indicating that the description reflects a high stress level in the participant.

5.3 Expected Outcomes

On the psychology side, the expected outcome of this study was to find that the ASC group would describe the animations using less anthropomorphic terms or make more mistakes in the attribution of mental states than the TD group. Moreover the ASC group is expected to express more signs of emotional stress during the decision making phase and description phase, since the task is assumed to be considered harder in this group.
In general the results from the traditional ToM animation tasks are hoped to be replicated.
It is also expected to observe a learning effect in the data, representing the combination of the learning curve associated with the understanding of the software and the priming effect that is inevitably present in the decision phase. As the story tree is explored (or as the software is played again) more of the decision labels are known to the participant. Since the terms are mildly anthropomorphic they will have an effect in what terms the descriptions are formulated. As this is known to happen the evaluation procedure can be adjusted to take this effect into account.

The expected outcomes in terms of the comparison between the interactive assessment method and the traditional observatory method were to find the interactive method to provide some additional data that would provide some further insight into the participant's mentalising process. Since the interactive assessment method was composed of ODA sequences, a comparison with the traditional method can be made by only looking at the data obtained in the OD section of the application and analysing the data gained through the Act (decision phase) section.
If significant difference in useful data is found, the interactive method can be put into context and found better or worse than the traditional approach.
Therefore the interactive method will only be a success if the data obtained through the participant's interaction with the system and their reasoning behind their actions carries additional relevant information describing their mentalizing process.
One strong example that would confirm this would be a consistent difference in the participant's behaviour. For instance, the participant would be able to explain what happened in the ToM animation correctly but does not understand the action requested by the system in the decision phase. This could be possible if the participant could explain the events on a high level of cognition yet on a spontaneous level of interaction would fail to understand the meaning of the interactive challenge and the anthropomorphised terms in that context.
This behaviour would be analogous to the reaction towards the pointing gesture - although people with ASC can explain the gesture upon request they do not react to it on a spontaneous occurrence (Klin, 2000).
Thus if this kind of behaviour would be recorded through Act, it would mean that the interactive assessment method would be able to capture the spontaneous mentalising reaction of the participant which would make it much more useful than the traditional observatory method.

5.4 Results:

5.4.1 The Participants

The participants for this experimental study were recruited through individual contact with the parents who were all professional academics.
In total five children took part in the experiment (see table 1). Of these five, one child (P3), male and age 9, was diagnosed with ASC (Asperger's syndrome) but also achieved above average academic results. The other four were typically developing children with age range of 6 to 8. All of the children had taken part in formative evaluation studies before, either as part of the Click-East or the ECHOES projects.
Originally it was planned to make a larger study in collaboration with a local school but due to practical reasons this was rendered impossible.

Participant ID	Age	Gender	TD/ASC
P1	8	Male	TD
P2	6	Female	TD
P3	9	Male	ASC (Asperger's syndrome)
P4	8	Male	TD
P5	7	Female	TD

Table 2: Overview of the five participants of the study

5.4.2 Description Data

The descriptions of the ToM animations were gathered from the transcripts of the audio recordings of the experiment and scored as explained in section 5.2.2 of this report.
The full set of annotated data can be found in appendix J.
A summary of ToM animation relative description scores shown in Table 3 below.

In general the average accuracy of the descriptions was 49.05% with standard deviation of 0.203. This would indicate that on average less than half of the main events of all the animations were recognised and described. Since two of the animations were not seen at all in the course of the whole experiment some description are missing (PwB-escape and Swy-Escape). Both animations were situated in the lowest level of the tree. Furthermore three animations (Hplay-StayWYellow, SwY-Fight, and TtG-ChaseYellow) were only seen once, thus only described once with relatively low accuracy scores but medium to high description types.

The data in table 3 provides a first hint that the ToM animations were indeed perceived in an increasing order of ToM complexity. Decisions 3, 4, 6 and 7 were all situated on the third level of the story tree, hence the most complex animations. It can be seen that all the participants used interaction terms in describing these animations with an average accuracy of 46.3% (st. dev. 0.247). Whereas on the second level (decision 3 and 5) the lowest description type varies between all the three types, and lowest type (action) was observed most often in the introduction as well as in decision 1.

Decision ID	Branch	Animation file name	#events / time	Accuracy		Type		
				Avg	St. dev.	lowest	highest	majority
0	Root	House-Intro	0.14	0.50	0.19	action	mentalising	action
1	Left	House-Stay	0.28	0.44	0.17	action	mentalising	draw a/i/m
	Right	House-Play	0.13	0.71	0.29	action	mentalising	interaction
2	Left	Hstay-PlayWBoth	0.25	0.69	0.27	interaction	mentalising	draw i/m
	Right	Hstay-YellowVSGreen	0.26	0.48	0.36	action	mentalising	draw a/m
3	Left	PwB-fightGreen	0.25	0.69	0.27	interaction	interaction	interaction
	Right	PwB-escape	0.18	-	-	-	-	-
4	Left	YvsG-HuntGreen	0.21	0.58	0.11	interaction	mentalising	draw i/m
	Right	YvsG-Escape	0.23	0.38	0.13	interaction	mentalising	interaction
5	Left	Hplay-StayWYellow	0.37	0.10	0.00	mentalising	mentalising	mentalising
	Right	Hplay-TalkToGreen	0.30	0.58	0.44	action	mentalising	action
6	Left	SwY-Fight	0.17	0.30	0.00	interaction	interaction	interaction
	Right	SwY-Escape	0.20	-	-	-	-	-
7	Left	TtG-ChaseGreen	0.19	0.75	0.12	interaction	mentalising	draw i/m
	Right	TtG-ChaseYellow	0.28	0.20	0.00	interaction	interaction	interaction

Table 3: Description scores for each animation.

Participant ID	# Play-throughs	Accuracy		type		
		Avg	st. dev	action	interaction	mentalising
P1	3	0.48	0.26	27.3%	36.4%	36.4%
P2	3	0.38	0.24	54.5%	27.3%	18.2%
P3	4	0.63	0.23	26.7%	53.3%	20.0%
P4	2	0.59	0.28	12.5%	62.5%	25.0%
P5	1	0.44	0.34	25.0%	25.0%	50.0%

Table 4: Participant relative results; illustrating average description accuracy and relative description type occurrences.

Participant ID	Tree depth	type		
		action	interaction	mentalising
P1	0	1	0	0
	1	0.33	0.67	0
	2	0	0	1
	3	0	0.67	0.33
P2	0	0.67	0	0.33
	1	1	0	0
	2	0.67	0	0.33
	3	0	1	0
P3	0	0.75	0.25	0
	1	0	0.75	0.25
	2	0.33	0.67	0
	3	0	0.5	0.5
P4	0	0	1	0
	1	0	0.5	0.5
	2	0.5	0	0.5
	3	0	1	0
P5	0	0	0	1
	1	0	0	1
	2	1	0	0
	3	0	1	0

Table 5: Decision matrix relative to each participant

The evaluation of the participants' performance is summarised in table 4.
Table 5 shows the participant's description type usage relative to the depth of the tree. Hence effectively capturing the distribution of description types. Since all this data is based on a statistical analysis normal distributions (gaussian distribution) can be assumed.

The expected results would take the form of a normal distribution on every level of the tree.
On level 0 , the distribution would have its mean on the action description type.
On level 1 and 2, the mean would be closer to the interaction description type.
And on level 3, the mean would be close to the mentalising description type.

Although it should kept in mind that the study was only run five times, hence there is relatively little data to analyse. P5's results for instance cannot taken as representative since she only played through the application once. Which explains the very clear cut distribution of types in her case.

5.4.3 Decision Data

The Act's log file of each experiment can be found in appendix I.

The amount of data collected during the decision phases of the application was substantially less than the description data. Yet 67.62% of the 43 dialogue snippets were rated to carry informative data, of which 6 carried mentalising descriptions and one an action description.

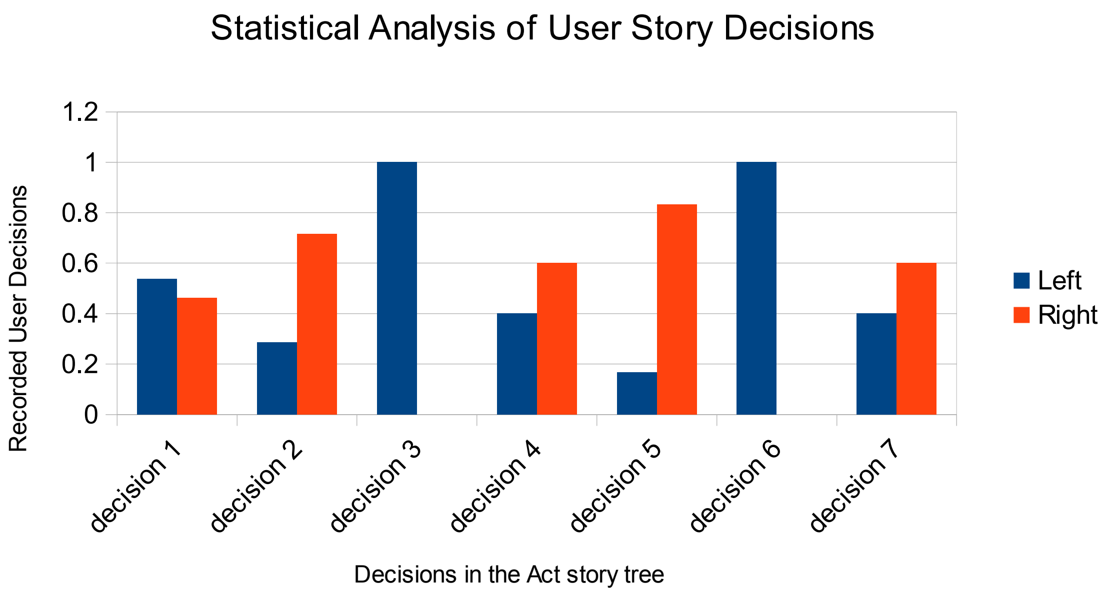
An overview of the collected data is given in Table 6 below.
The description type varied heavily between the participants. For instance the data recorded from the experiment session with P5 shows that all of the transition dialogue that was recorded showed strong signs of stress and confusion.

Participant ID	# recorded transitions	Informative (in %)	type											
			priming		stress		question		engaged		mentalising		other	
			Avg Strength	split	Avg Strength	split	Avg Strength	split	Avg Strength	split	Avg Strength	split	Avg Strength	split
P1	5	80.0%	2	25%	0	0%	2	25%	0	0%	2	25%	1	25%
P2	13	69.2%	2	11%	1.7	67%	0	0%	1	22%	0	0%	-	0
P3	9	88.9%	0	0%	0	0%	2	13%	1	25%	2	63%	-	0
P4	11	0.0%	-	-	-	-	-	-	-	-	-	-	-	0
P5	5	100.0%	0	0%	1.8	100%	0	0%	0	0%	0	0%	-	0

Table 6: Description data gathered during the decision phase relative to each participant

5.4.4 Statistical Analysis and Context

The five participants explored on average 50.67% (st.dev: 0.137) of the story tree. In total 86.67% of the story tree were explored in the experimental study, which means all the animations were seen except 2.
Graph 3 provides an overview of the decision preference of the participants. It was not surprising to find that the boys when presented with the choice to either “fight” or “escape” in decision 3 and 6, both boys chose to fight, which left the 2 alternative animations unexplored.



Graph 3: Statistical Analysis of User Story Decisions; The graph illustrates which animations where chosen the most, relative to the two choices in each decision phase. The decisions are defined in the Act story tree diagram (graph 2, section 4)

5.5 Evaluation

The experimental data is evaluated in two individual parts. In the first part the data is assessed with emphasis on the psychological meaning of the data. In the second part the data is used to evaluate Act.
The suitability of the interactive assessment method is determined through both of these evaluations and is described in the conclusion.
In the discussion section the results are put into context and the future improvements section highlights the most valuable changes that could be done in a follow up study.

5.5.1 The Experimental Study

As a reminder, the description type is ranked accordingly: action < interaction < mentalising. (further details on the scoring method can be found in the appendix G).

The data obtained through the descriptions of the participants in general followed the expected trend of ToM complexity that was implemented in Act's story tree.

The ToM animation relative summary of the scores (table 3) shows that the lowest description type used to describe the root node was "action". The same is the case for the two animations on level 1. The lowest description type then varies between "action" and "interaction" on level 2. Until finally on the level 3 all of the descriptions have at least been of the "interaction" type. The general increase of the minimum description type would suggest that an increase in complexity was recognised. The data in table 5 also shows the shift of the type distribution from "action" in the top left corner of the matrix, to "mentalising" in the bottom right corner.
Furthermore it is seen that the highest description type is always "mentalising", except for two animations which were only described once with a lower type. One possible explanation for this trend would be that some of the participants always used ToM terms to describe the animated events. The participant data in table 5 confirms that at least one participant on each level of the tree used a mentalising description.
Specifically, participants P2 and P5 used the mentalising type in their descriptions of the root node, which is the least complex of them all. The description data shows that P2's mentalising description of the root node was of very low accuracy (0.16) which means that only one of the three main events was described with intermediate detail. Also her description was given on the second time the root introduction animation was seen. Looking at her wording itself: "He really was hiding." makes reference to a ToM animation that was played on the first run, namely the animation related to the "Hide" decision label (with animation file name: House-Stay.swf). Thus her description most likely originates from previous priming through knowledge of other animations. P5's mentalising description has a slightly higher accuracy of 0.33, indicating that two events were described with intermediate accuracy and she only played through Act once. She described the house in the animation upon first sight as "trap door thing" and refered to the yellow square as "guy". This would suggest that the participant in general uses ToM more often than the other participants and thus by default uses anthropomorphic language to describe unfamiliar behaviour and events.

By studying the data obtained through the descriptions gathered during the decision phase, it can be seen that both P2 and P5 showed the highest number of descriptions rated to demonstrate stress. This might be an indication that the task was harder for the two participants than for the others. Another indication is that P5 played through Act only once and used the “Rephrase” button twice, thus opting for the easier interactive challenge in two out of three decision phases. A strict conclusion cannot be made due to the very small number of participants in the study. One possible explanation would be that girls are more likely to express their emotions explicitly, whereas boys do not. Hence no substantial conclusion can be taken on this data.

Having a closer look at the description type distribution over the story tree depth (Table 5), it can be seen that three of the five participants come close to the expected type distribution: P1, P3 and P4. Moreover the same three participants are also the top three in terms of accuracy.
Looking at the descriptions of the participants again it reveals that all three are the oldest participants and male. The higher age most likely entails better language skills which would be one explanation for their performance.
P2, the youngest of the participants, showed the lowest accuracy and also the highest number of descriptions of the “action” type. Although she shifted to describing the last animations in interaction terms in both of her play-throughs.

P3's performance in the experiment was interesting in multiple ways.
Firstly he went against the trend given in previous ASC research which predicted that children with ASC are less likely to use anthropomorphic terms when describing the events of the ToM animation (Dittrich & Lea, 1994; Klin, 2000; Rutherford, Pennington, & Rogers, 2006). This could be related to his high academic results, or to the fact that interviewing process was followed through better in this iteration of the experiment.
Either way the data obtained through the interactive stage made it explicitly clear that P3 was using much more anthropomorphic and mentalising terms than all the others. In fact most of the the dialogue recorded reasoning about the interactions with the characters in the animations in mentalising terms. Moreover the data did not report any signs of stress but instead that some questions were asked, thus hinting to constructive confusion, and signs of engagement with the software. This conclusion could not be made only considering the animation description data. This data set would only indicate that P3 followed the expected complexity trend of the story tree, but his very strong mentalising skills are not reflected in those descriptions. In fact only 20% of P3's animation descriptions were of the mentalising type, which is the second lowest split among all the participants.

Another general trend that could be observed in the data was the way the house structure in the ToM animations was described by the participants. As mentioned in the testing phase, the animation environment could either be interpreted from the bird's eye or top-down view, or from a side view, from which the house is more of a cellar or box which is accessed through a trap door and which then in some cases appeared to not follow the laws of gravity.
The data shows that in fact all the participants who described the house to be a “box” with a “lid” or “trap door” used mentalising terms in their descriptions as well. This effect was surprisingly consistent across all the descriptions.

In general, the priming effect through the interactive challenges, or a learning effect based on the fact that the participant is familiar with the procedure of the experiment could not be securely determined in this experimental study due to the small number of participants.

5.5.2 The Software

The description data obtained through the decision phase provided, besides further ToM descriptions, information about the emotional state of the participant while using the software. Confusion and frustration were clearly audible in the audio recordings through sighs, nervous laughs, or repetitive questions.
Although the experiment was explained to the participant before it was started, many showed initial confusion with the task at hand. This is most clearly visible in Act's log file, which captured all the button presses of the user. The initial use of the “rephrase” button in two experiments points to possible problems in understanding the purpose of the system and their interaction. In one case (P5) the initial “rephrase” button press was followed by another a bit later in the experiment, this was most likely caused through stress.

Some of the experiments were followed by a brief rating of the software by the participants. The children said they found the task was easy but a bit boring, and that they would generally not change anything on the software.

5.5.3 Conclusion

The evaluation of the description data set gave clear insight into what type of terms were used in the participants' descriptions and as such makes up the main body of results.
The decision data provided additional information about the emotional state of the participant as well as further ToM descriptions.

Although the experimental study was relatively small, it can be concluded that the interactive stage of the method provided further insight into the mental states of the participants, and as such both the interactive assessment method and the software implementation fulfilled their purpose.

5.5.4 Discussion

Originally it was planned to make a larger study in collaboration with a local school but due to practical reasons this was rendered impossible. This left the experimental study with a relatively small number of participants of which only one participant was diagnosed with ASC. Hence the results of previous research could not be completely replicated.
The optimal evaluation of both the interactive assessment method and the software would encompass a large quantitative study, with at least two different test groups and a randomized control condition. The study would have to be taken over a relatively long period of time in order to capture the individual differences in learning abilities and personal development.
Ideally the control condition would take the form of a random animation, meaning an animation in which the actors move around randomly. The participants' descriptions of these random movements could then be used to establish a baseline measure for all future measurements. Thus the data could be normalized by removing the 'average anthropomorphic awareness' (or likelihood to use anthropomorphic language), which is measured through the control condition. As the animations within the Act computer application increase in anthropomorphic content the further down the story tree the participant progresses, a comparison between the objective increase of social events in the animation and the recognition of these anthropomorphic events by the participant could be done. Most ideally the results of a typically developing child would match a linearly increasing function (with the number of anthropomorphic events on the x axis and the number of correctly matching descriptions of these events on the y axis.), whereas the results of a child with ASD would be expected to be also linear but with a slower ascent rate (smaller positive gradient). (Dr G. Rajendran, personal communication, February 15, 2012)

As time and resources were very restricted for this project, the ideal evaluation described above could not be completed. Nevertheless it was attempted to get a a variety of participants in order to get more insight into the suitability of the chosen approach of investigation.

Besides the lack of participants, the experiment also struggled to fully implement the interviewing process in a consistent manner. Unfortunately in some experiments the interaction of the participant was not questioned subsequently which resulted in substantially less data as originally expected. Only in a few cases the structure of the interview met the actual design. But in those cases the information the interactive approach provided was significant (as in the case of the experimental session with participant P3).
Another aspect of the interviewing process which should be put forward is the intimidation effect it has on the child participants. Describing a green square as being jealous of the blue square might be left out, if the participant fears that it might sound silly and could be judged badly for it. This effect is of course reinforced through the fact that it is an experiment and the participants answers are recorded. A questionnaire based approach might be able to avoid the effect as the participant's answers wont be evaluated straight away.

5.5.5 Future improvements

By design the interactive assessment method provides insight into the mental state of the participant. The the data gathered in the decision and the description phases combined can be used to highlight frustrating sections in the experiment and thus it could be used to design better experiments. “Better” experiments could either use the emotional data to reduce the stress put on the participant or the data could be used to elicit a specific emotional state in the participant.

The decision data made it apparent that some participants had problems understanding the task itself, or found it difficult to express their interpretation of the animation. Especially the younger participants showed more signs of stress through sighing and repetitive questioning, which might be an indication that the task was not quite suitable for children under 7 years.
But the origin of the confusion might also be caused by the user interface of the software. The 'I don't know' button in the decision phase only proved to be useful when the participant was very confused about the application. In the end the anthropomorphic interaction challenge defaulted to a simple decision challenge by itself, since the labelling of the options is understood, hence a preference can be established. Thus another way needs to be found to enable the participant to express their difficulty in understanding the events in the ToM animations.

Future implementations of the interactive assessment method could investigate the attribution of mental states as a function over time. This could potentially be done through a systematic increase in priming through the interactive challenge.
In a similar way the method could also be used to build mentalising learning environments.

6 Bibliography

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1-16. doi:10.1016/S0885-2014(00)00014-9

American Psychiatric Association. (2000). Diagnostic and Statistical Manual of Mental Disorders.

Baron-Cohen, S, Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry*, *42*, 241-251.

Baron-Cohen, Simon. (1994). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *Cahiers de Psychologie Cognitive*, *13*, 513-552.

Baron-Cohen, Simon. (1995). *Mindblindness, An Essay on Autism and Theory of Mind*. MIT Press.

Baron-Cohen, Simon, Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37-46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9775957

Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology*, *33*, 680-685.

Berry, Diane S, & Springer, K. (1993). Structure, Motion, and Preschoolers' Perception of Social Causality. *Ecological Psychology*, *5*(4), 273-283.

Berry, D. S., Misovich, S. J., Kean, K. J., & Baron, R. M. (1992). Effects of disruption of structure and motion on perceptions of social causality. *Personality and Social Psychology Bulletin*, *18*, 237-244.

Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication When do infants acquire a "theory of mind." In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition*. Hillsdale, NJ: Erlbaum.

CLICK-EAST Website. (2012). CLICK-EAST Website. Retrieved from http://www.clickeast.co.uk/

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain : a journal of neurology*, *125*, 1839-1849. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12135974

Charman, W., & Stone, T. (2006). *Social & Communication Development in Autism Spectrum Disorders: Early Identification, Diagnosis, & Intervention* (p. 348). New York: Guildford Press.

Choi, H., & Scholl, B. J. (n.d.). Measuring causal perception: Connections to representational momentum? *Acta Psychologica*, *123*, 91–111.

Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *4*, 568-570.

Dittrich, W. H., & Lea, S. E. G. (1994). Visual perception of intentional motion. *Perception*, *23*, 253-268. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7971105

ECHOES website. (2011). ECHOES website. Retrieved from http://echoes2.org/?q=node/2

Florio, D., & Uza, P. (2007). custom-context-menu: Capture right-click events in Flash / Flex (AS 3). Retrieved from http://code.google.com/p/custom-context-menu/

Fombonne, E. (2009). Epidemiology of Pervasive Developmental Disorders. *Pediatric Research*, *65*(6), 591-598.

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., & Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive Brain Research*, *24*, 41–47.

Gibson, E. J. (1969). *Perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (Original work published 1979).

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Happé, F. G. (1994). An advanced test of theory of mindL understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J Autism Dev Disord*, *24*, 129-54.

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behaviour. *The American Journal of Psychology*, *57*(2), 243-259.

Hubbard, T. L., Blessum, J. A., & Ruppel, S. E. (2001). Representational momentum and Michotte's "launching eVect" paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 294–301.

Hubbard, Timothy L. (2004). The Perception of Causality: Insights from Michotte's Launching Effect, Naïve Impetus Theory, and Representational Momentum. In A. M. Oliveira, M. P. Teixeira, G. F. Borges, & M. J. Ferro (Eds.), *International Society for Psychophysics* (pp. 116-121). Coimbra, Portugal: Fechner Day.

Kane, J. L., Winchester, W. W., Scarpa, A., & Smith-jackson, T. L. (2011). *Exploration of Computer Game Interventions in Improving Gaze Following Behavior in Children with Autism Spectrum Disorders*. Virginia Polytechnic Institute and State University.

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

Klein, A. M., Zwickel, J., Prinz, W., & Frith, U. (2008). Animated triangles: an eye tracking investigation. *Quarterly journal of experimental psychology*, *62*(6), 1189-1197. doi:10.1080/17470210802384214

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The Social Attribution Task. *Journal of child psychology and psychiatry, and allied disciplines*, *41*(7), 831-846. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11079426

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1430), 345-360. doi:10.1098/rstb.2002.1202

Kruschke, J. K., & Fragassi, M. M. (1996). The perception of causality: Feature binding in interacting objects. *In Proceedings of the eighteenth annual conference of the cognitive science society* (pp. 441–446). Hillsdale, NJ: Erlbaum.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition*, *43*, 225-251.

Levy, S. E., Mandell, D. S., & Schultz, R. T. (2009, November 7). Autism. *Lancet*. doi:10.1016/S0140-6736(09)61376-3

MacNamara, J., Baker, E., & Olsen, C. (1976). Four year olds' understanding of pretend, forget, and know evidence for propositional operations. *Child Development*, *47*, 62-70.

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008). Story-based Learning : The Impact of Narrative on Learning Experiences and Outcomes. *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems* (pp. 530-539). Montreal, Canada.

Michotte, A. (1950). The emotions regarded as functional connections. *In Feelings and Emotions: The Mooseheart Symposium (Reymert, M., ed.), pp. 114–125, McGraw-Hill. Reprinted 1991: In Michotte's Experimental Phenomenology of Perception (Thinès, G. et al., eds), pp. 103–116, Erlbaum*.

Montgomery, D. E., & Montgomery, D. A. (1999). The influence of movement and outcome on young children's attribution of intention. *Br J Dev Psychol*, *17*, 245-261.

Newschaffer, C. J., Croen, L. a, Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., Mandell, D. S., et al. (2007). The Epidemiology of Autism Spectrum. *Annual Review of Public Health*, *28*, 1-24. doi:10.1146/annurev.publhealth.28.021406.144007

Oakes, L. (1994). Development of infants' use of continuity cues in their perception of causality. *Dev. Psychol.*, *30*, 869–879.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: knowledge, belief, and communication. *Child Dev.*, *60*, 688-700.

Pomerantz, J. R. (2006). Perception: Overview. *Encyclopedia of Cognitive Science*.

Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., Foster, M. E., et al. (2011). Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing*, *16*(2), 117-127. doi:10.1007/s00779-011-0384-2

Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515-526.

Rimé, B., Boulanger, B., Laubin, P., Richir, M., & Stroobants, K. (1985). The perception of interpersonal emotions originated by patterns of movement. *Motivation and Emotion*, *9*(3), 241-260. doi:10.1007/BF00991830

Ritvo, E. R., & Freeman, B. J. (1978). National Society for Autistic Children definition of autism. *Journal of Autism and Developmental Disorders*, *8*, 162–167.

Roeyers, H., Buysse, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: empathic accuracy in adults with a pervasive developmental disorder. *J Child Psychol Psychiatry*, *42*, 271-278.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 173–220). New York: Academic Press.

Rutherford, M. D., Pennington, B. F., & Rogers, S. J. (2006). The perception of animacy in young children with autism. *Journal of autism and developmental disorders*, *36*(8), 983-992. doi:10.1007/s10803-006-0136-8

Rutter, M. (1978). Diagnosis and definition of childhood autism. *.Journal of Autism and Developmental Disorders*, *8*, 139–161.

Rutter, M. (2005). Incidence of autism spectrum disorders: changes over time and their meaning. *Acta Paediatrica*, *94*(1), 2.

Schlottmann, A., Ray, E. D., Mitchell, A., & Demetriou, N. (2006). Perceived physical and social causality in animated motions: Spontaneous reports and ratings. *Acta Psychologica*, *123*, 112–143.

Scholl, B. J., & Nakayama, K. (2004). Illusory causal crescents: Misperceived spatial relations due to perceived causality. *Perception*, *33*, 455–469.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, *4*(8), 299-309. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10904254

Shantz, C. U. (1983). Social cognition. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of Child Psychology* (Vol. III.). New York: Wiley.

Shultz, T. R., & Cloghesy, K. (1981). Development of recursive awareness of intention. *Developmental Psychology*, *17*, 456-471.

Shultz, T. R., Wells, D., & Sarda, M. (1980). The development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social & Clinical Psychology*, *19*, 301-310.

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistics, retarded and normal children. *J Child Psychol Psychiatry*, *33*, 591-605.

Springer, K., Meier, J. A., & Berry, D. S. (1996). Nonverbal bases of Social Perception: Developmental Change in Sensitivity to Patterns of Motion that Reveal Interpersonal Events. *Journal of Nonverbal Behavior*, *20*(4), 199-211.

Tartaro, A., & Cassell, J. (2006). Authorable virtual peers for autism spectrum disorders. Riva del Garda, Italy: Paper presented at the combined workshop on language-enabled educational technology and development and evaluation of robust spoken dialogue systems at the 17th European conference on artificial intelligence (ECAI06).

Wagemans, J., van Lier, R., & Scholl, B. J. (2006). Introduction to Michotte's heritage in perception and cognition research. *Acta psychologica*, *123*(1-2), 1-19. doi:10.1016/j.actpsy.2006.06.003

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of autism and developmental disorders*, *34*(5), 473-93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15628603

Wetherby, A., Prizant, B., & Hutchinson, T. (1998). Communicative, social/affective, and symbolic profiles of young children with autism and pervasive developmental disorders. *American Journal of Speech-Language Pathology*, *7*, 79-91.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103-128.

Wing, L., & Potter, D. (2002). The epidemiology of autistic spectrum disorders: is the prevalence rising? *Mental Retardation And Developmental Disabilities Research Reviews*, *8*(3), 151.

perception. (2010).*Oxford dictionaries*. Oxford University Press. Retrieved from http://oxforddictionaries.com/definition/perception

7 Appendices

A. Formal Evaluation Form 1

B. Formal Evaluation Form 2 – Expert Session

C. Parent Information Sheet

D. Blank Consent Form

E. Blank Assent Form

F. Experiment Interviewing Script

- Description of the project:
1a. "You will see a few small films and you can decide how the story should develop.
As you go along, please describe what is happening in the film."

preferred: 1b. "You will see a short story which I want you to describe to me. In turn you can choose how the next chapter of the story should be."

- Questions to ask:
-- "Can you describe what you saw?"
"What happened in this animation?"

if "chapter" was used (intro 1b.):
"What happened in this chapter?"
"What did the Blue/Yellow/Green square do?"
"What happened to the Blue/Yellow/Green square?"
"Why?"
"Can you describe Blue/Yellow/Green to me?"
"Why did you want for Blue to <action>/do that?"
"Why should Blue do that?"
"Why did you choose that action?"

& repeating phrases the participant used, as questions.

- Instructions:
"Click when you are ready to start the film."
"Do a mouse click / click with your mouse to play the first film."
"Watch carefully and describe what you see"
“If you are unsure, choose 'I don't know'".
"You have to choose what Blue should do next."
"You are deciding how the story continues"
"You can choose either one of them" (when the 'i don't know' button was pressed)

Prompts:
"Let's watch the film again. (Pause)
Please describe what you saw."
"Would you like to watch the film again?"
[Yes] "then choose 'Play again'."
[No] "then choose 'continue'."

"You can stop at any time"
(make a note where they stopped in the log
later, technically you'll have to play the game until the end and press 'have a break')

"Do you want to play again or have a break?"
[]"alright, choose ..."

G. Scoring Criteria

Scoring criteria
Adaptation of Abell et al., 2000

For each ToM animation the main events were noted by the designer. (see "Act – Animation clip full descriptions " document).
The accuracy of description of the whole ToM animation is the average accuracy score of the individual main event description phrases.
A score is given by the accuracy rating and the type of description. The lowest (default) score is "0, action".

1. Accuracy of descriptions: general rules:

Each description phrase is scored on a scale from 0 to 2, depending on how accurately it reflects the event.

2 spot-on description of the story or the actions represented; All the main events are described; Can be concise just capturing gist, or can be discursive
1 partial description of the sequence; description is related to the sequence, but imprecise or incomplete
0 bizarre descriptions, plainly wrong descriptions, and responses that focus solely on a minor unimportant aspect of the sequence

2. Type of description: general rules

Each event description (regardless of accuracy score given) is rated as showing *random action, interaction action* or *mental state attribution*.
These rules apply to all descriptions for any sequence. In each case the *highest* level of descriptive language is scored (i.e. mentalising trumps interaction trumps action).
Priming of the description through the decision labels is assumed, thus primed phrases will be ignored.

Random action: Simple action, no mention of goal, no reference to interaction of the characters, e.g. floating.

Interaction: Specific reference to purposeful movement, without reference to mental states, e.g. following; fighting; copying; having a race. More than one action may be described, e.g. leading and following. May involve use of direct speech without mental state verb; may include qualification of verb by 'trying to', e.g. boy trying to ask her something, but teacher kept walking away.

NOT: purposeless action.
NOT: implied mental state attribution

Mentalising attribution: Use of mental state verbs to describe reciprocal interactions, e.g. wanting; hiding; tricking; pretending; being naughty;
Also use of
NOT: complex interaction, e.g. chasing each other round the house; x pushing y out of the way.
NOT: solely direct speech.
NOT: solely 'trying to'.

Priming: descriptions or exclamations which demonstrate the priming effects the software, specifically the decision labels, has on the participants. e.g. Only the event is only described with the words the participant just saw in the decision phase.
The phrase is ignored and its accuracy is not counted towards the total accuracy of the whole description. If there is only little priming, meaning the participant described in different ways, that alternative description should be considered.

For transitions only:

Stress: Signs of confusion, strong emotional regulation exclamations and disorientation while operating the software.

Engaged: description / exclamation shows willingness to spend more time with the system and a good mood = opposite of stress; Participant is engaged and wants to further explore the system.

Questions: General questions raised during the experiments, marked for later analysis and recognition of potential improvement opportunities.

3. Transition descriptions / exclamations

Transitions are rated on usefulness (True/False). Such a dialogue is deemed useful if it gives insight either into a technical problem with the presentation software or more importantly into the mentalising process of the participant. If the later is the case, a score constituting of a strength measure (0-2 similar to accuracy) and type are assigned to it just as above.

H. Scoring Guidelines / Gold Standard

Act – Animation clip full descriptions
Evaluation guidelines / Gold standard descriptions

To be used interchangeably:
- the blue square = Blue
- the green square = Green
- the yellow square = the orange square = Yellow
- box = house (the structure in the lower half of the screen)

Decision label: Intro		
Clip name: House-Intro.swf		**Story tree level:** 0
#events: 3	**Clip duration:** 21.8s	
The blue square plays around outside (runs in circles). Discovers the house (box), opens the door and enters the house, while leaving the door ajar. The blue square then explores/inspects the four corners of the house. Happy, the blue square dances around in the house. A yellow square appears outside. The blue square sees the yellow square and curiously approaches the window/wall of the house.		
Main events: Blue arrives; Blue explores house; Yellow arrives;		

Decision label: Hide		
Clip name: House-Stay.swf		**Story tree level:** 1
#events: 3	**Clip duration:** 10.8s	
Yellow approaches the window while Blue moves back a little. Blue approaches and closes the door, before Yellow has the chance to enter. Blue moves back and retreats (hides) in the back of the house (left bottom corner). Yellow tries to get into the house and knocks on the door to get in. Nobody opens, so just waits there. Then the green square appears. (Which surprises the yellow square.)		
Main events: Blue locks Yellow out; Blue hides in the corner; Green arrives;		

Decision label: Play		
Clip name: House-Play.swf		**Story tree level:** 1
#events: 2	**Clip duration:** 15.4	
Yellow approaches the window while Blue moves back a little. Blue goes to the door, opening it for Yellow and greets Yellow. Yellow responds and they move inside and dance for a bit. Then they stop and Yellow tags/pokes Blue and runs away. Blue follows Yellow. They are playing catch. Then the green square appears and Yellow sees it through the window and stops running. Blue continues running although and bumps into Yellow. Then both Blue and Yellow move towards the window to look at Green.		
Main events: B&Y play inside the house; Green arrives;		
Decision label: Play		
Clip name: HStay-PlayWBoth.swf		**Story tree level:** 2
#events: 4	**Clip duration:** 16.1s	
Green and Yellow greet/talk then dance in circles around outside. Blue runs to the door, opens it and watches the two other squares play. Yellow sees Blue and stops dancing, leaves Green one its own, and approaches Blue to say hello. After a short chat, Blue and Yellow dance. Green still alone. Blue wants to invite Green to dance as well but Green hits Blue. (Green is jealous)		
Main events: Y&G dance; Blue goes out; B&Y dance; G hits B;		

Decision label: Watch them		
Clip name: HStay-YellowVSGreen.swf		**Story tree level:** 2
#**events:** 4	**Clip duration:** 15.3s	
Yellow greets Green, but Green is silent and hits Yellow hard. Yellow is hit a few times and tries to run away from Green. They chase around the house. Yellow reaches the door, opens and enters the house. Green is locked out and knocks against the door / tries to open the door but it is locked. Blue approaches Yellow a bit (Blue is surprised by Yellow's sudden appearance)		
Main events: Green attacks Yellow; Yellow escapes; Y enters house; Green tries to enter house;		

Decision label: Protect Yellow		
Clip name: HPlay-StayWYellow.swf		**Story tree level:** 2
#events: 5	**Clip duration:** 13.5s	

Green approaches the door and slams it shut. Blue and Yellow are startled and jump to the left side. Yellow want to confront evil Green, but Blue tells Yellow off trying to block Yellow's way. Yellow pushes Blue to the side and leaves the house and talks to Green. Green is silent and punches Yellow back hard. Blue catches Yellow and protects Yellow from Green by moving in between both of them.
Main events: Door slam; Blue blocks Yellow/Yellow wants outside; Yellow talks to Green; Green hits Yellow; Blue defends;

Decision label: Talk to Green		
Clip name: HPlay-TalkToGreen.swf		**Story tree level:** 2
#events: 3	**Clip duration:** 10.04s	
Blue leads on and goes outside to greet Green. (Blue pushes the door open slightly) Yellow follows Blue a bit slower and pushes the door open. Green laughs (big wiggles), then Blue laughs (big wiggles). Then Green tags / pokes Blue and runs away to the left side, Yellow runs away to the right side. Blue is confused who to follow first. Starts off following Green, then reconsiders. (The initiated a game of “Tick” or “Catch”)		
Main events: Talk to Green; Tick; Run;		

Decision label: Fight		
Clip name: PwB-fightGreen.swf		**Story tree level:** 3
#events: 4	**Clip duration:** 15.9s	
Blue gets back up from Green's hit, but Green hits Blue again. Blue is pushed back by the blow and Green moves up (in an arc). Yellow is scared by the fighting and moves away from the two, hiding behind the door, considering going inside if the fight gets worse. Blue tries to punch Green a few times but Green is unaffected by the hits and throws Blue against the door. As Blue hits the door Yellow is thrown back as well. Green approaches Blue, Blue steps back (scared). Green kicks Blue through the door and to the other side of the house. Then Green approaches Yellow who doesn't move. Only when Green is close Yellow tires to step away, but Green grabs Yellow and drags Yellow behind it. When they pass the broken door, Yellow tries to run to Blue but is held on by Green. Green moves on and Green and Yellow disappear. Blue remains in the broken house and doesn't move. (Knocked Out.)		
Main events: big Fight; Blue looses; kidnap Yellow; G&Y disappear;		

Decision label: Escape		
Clip name: PwB-escape.swf		**Story tree level:** 3
#events: 4	**Clip duration:** 22.5s	
Green hits Blue another time, Blue flies back against the wall of the house. As Green moves back to hit again, Yellow comes up to Green and embraces Green, preventing it to hit Blue. Yellow hugs Green and slowly moves away from Green but encourages Green to come along. Yellow moves to the door, opens it a bit and moves further inside the house, telling Green to follow. Green follows Yellow inside the house, where Yellow hugs and moves around Green, and suddenly runs out of the house. Blue jams shut the door as soon as Yellow was out, trapping Green inside. Blue and Yellow embrace and hug, and disappear together, while Green remains in the house knocking on the door to get out.		
Main events: Fight; Yellow's coax; Green is trapped; B&Y hug+disappear;		

Decision label: Attack		
Clip name: YvsG-HuntGreen.swf		**Story tree level:** 3
#events: 3	**Clip duration:** 14.3s	
Blue and Yellow move together (team up) and position in front of the door. Green is on the other side of the door. Together, Blue and Yellow throw open the door, the door hitting Green and knocking Green out. They then move behind Green and push Green into the house and close the door, locking Green in. Blue and Yellow embrace and disappear. Green remains still inside the house.		
Main events: team up; knock Green out + trap; hug+disappear;		

Decision label: Defend		
Clip name: YvsG-Escape.swf		**Story tree level:** 3
#events: 4	**Clip duration:** 17.2s	
Yellow hides behind Blue, while Green breaks through the door. The door is broken and Green moves towards Yellow, but Blue stands in Green's way. Green tries to move around Blue, but Blue doesn't let Green do so. Green gets angry and hits Blue who is thrown back by the hit. Yellow sits in the corner. Blue hits Green back, pushing Green back into the room, while Yellow moves from the top left corner of the house behind Blue and towards the door. (pushing a piece of the broken door beside). Green tries to get at Yellow again but is pushed back against the wall by Blue. Blue slowly follows Yellow. Green starts another attempts to get to Yellow but Blue throws a piece of the broken door at Green who is knocked out. Blue and Yellow, no longer threatened, embrace and disappear together.		

Main events: Yellow hiding behind Blue / Blue defending Yellow; Green destroying the door/breaking in; Green tries to get to Yellow; Yellow and Blue embrace + disappear;

Decision label: Fight Green		
Clip name: SwY-Fight.swf		**Story tree level:** 3
#events: 3	**Clip duration:** 17.8s	
Blue hits Green, who is knocked back by the blow, Blue follows and hits again. Green surrounded by Blue and Yellow is intimidated and forced to go into the house. Blue shuts the door on Green. Blue and Yellow are happy and embrace, while Green is trapped in the house and knocks against the door to be released. Blue and Yellow disappear together. Green is still trapped and goes to a corner (is angry and sad).		
Main events: Blue wins fight; Green got trapped; B&Y embrace + disappear;		

Decision label: Escape with Yellow		
Clip name: SwY-Escape.swf		**Story tree level:** 3
#events: 5	**Clip duration:** 24.5s	
Green punches Blue, Blue is knocked back but Yellow catches Blue. Yellow then grabs / supports Blue and they flee/run away around the house. Green follows them but is a bit slower. All around the house Yellow stops and closes the door, but Blue keeps on moving to the side of the house and is later joined by Yellow. Green comes around the corner and sees the closed door and assumes the two other squares are inside, and Green goes inside the house. As soon as Green enters, Blue and Yellow jump out of their hiding place and run for it (disappear), while Green is inside the house and realises its mistake and is angry. Green storms out and disappears as well (although goes into the other direction)		
Main events: Green hits Blue; Blue and Yellow flee; Door trick; B&Y escape; G angry + disappears;		

Decision label: Catch Green		
Clip name: TtG-ChaseGreen.swf		**Story tree level:** 3
#events: 3	**Clip duration:** 15.4s	
Blue chases Green; Green and Yellow run behind the house, meet and laugh. Yellow sees Blue and moves backwards while Green continues laughing. Blue tags/ticks Green, which surprises Green. Yellow laughs and Blue and Yellow run to the front of the house while Green follows Yellow. They all enter the house and run in circles, then stop and laugh.		
Main events: Blue chases Green; Green is caught; they all play inside the house;		

Decision label: Catch Yellow		
Clip name: TtG-ChaseYellow.swf		**Story tree level:** 3
#events: 5	**Clip duration:** 17.9s	
Blue chases Yellow; Yellow and Green run away around the house and meet behind the house. Green blocks Yellow's way so Blue catches Yellow. Green laughs but Yellow then tries to catch Green and all the squares run to the front of the house. Blue is the first and grabs the door, Blue waits for Green to pass and then quickly shuts the door on Yellow who is now locked out. Blue and Green laugh. Yellow gets inside and they all run in circles, then stop and laugh again.		
Main events: Blue chases Yellow; Green blocks Yellow's path; Blue catches yellow; Blue locks Yellow out; play inside in circles;		

I. Act Log File

1,House-Intro.swf,Rephrase,House-Stay.swf,HStay-PlayWBoth.swf,PwB-FightGreen.swf,House-Intro.swf,House-Play.swf,HPlay-StayWYellow.swf,SwY-Fight.swf,House-Intro.swf,House-Stay.swf,HStay-YellowVSGreen.swf,YvsG-HuntGreen.swf;

2,House-Intro.swf,Replay,House-Stay.swf,HStay-YellowVSGreen.swf,YvsG-HuntGreen.swf,Replay,House-Intro.swf,House-Play.swf,HPlay-TalkToGreen.swf,Replay,TtG-ChaseYellow.swf,Replay,Replay,House-Intro.swf,House-Stay.swf,HStay-YellowVSGreen.swf,YvsG-Escape.swf,Replay;

3,House-Intro.swf,House-Stay.swf,HStay-YellowVSGreen.swf,YvsG-Escape.swf,House-Intro.swf,House-Play.swf,HPlay-TalkToGreen.swf,TtG-ChaseYellow.swf,House-Intro.swf,House-Stay.swf,HStay-PlayWBoth.swf,PwB-FightGreen.swf,House-Intro.swf,House-Play.swf,HPlay-TalkToGreen.swf,TtG-ChaseGreen.swf,Replay;

4,House-Intro.swf,House-Play.swf,HPlay-TalkToGreen.swf,TtG-ChaseYellow.swf,Replay,House-Intro.swf,House-Stay.swf,HStay-YellowVSGreen.swf,Replay,YvsG-Escape.swf,Replay;

5,House-Intro.swf,Rephrase,House-Play.swf,HPlay-TalkToGreen.swf,Rephrase,TtG-ChaseGreen.swf;

J. Formatted Data

Experiment Scores

Experiment #1:

<table>
<tr><td>Decision label: Intro</td><td>Rephrase: False</td><td>#Play-through: 1</td></tr>
<tr><td colspan="3">lip name: House-Intro.swf</td></tr>
<tr><td rowspan="2">P: So.. Okay. Well, what seams to happen was that ..
the blue square has appeared on the screen first, moved about outside the box,
and then, part of the box opened,
the blue square went inside
and then the box partly closed again.
The blue box moved about inside the other box for a moment
and then a yellow box appeared outside
and then it stopped, and this came up.</td><td colspan="2">Blue arrives;
2, action

Blue explores house;
1, action

Yellow arrives;
1, action</td></tr>
<tr><td colspan="2">Avg: 4/6 = 0.66, action</td></tr>
<tr><td colspan="3">Score: 0.66, action</td></tr>
</table>

<table>
<tr><td>Decision label: Hide – Blue square</td><td>Rephrase: True</td><td>#Play-through: 1</td></tr>
<tr><td colspan="3">Clip name: House-Stay.swf</td></tr>
<tr><td rowspan="2">Well, then what seams to happen was
Blue actually closed that box
and Yellow tried to get in but kept bouncing of it.
And Green.. and then a green cube appears.. it's outside.</td><td colspan="2">Blue locks Yellow out;
1, interaction

Blue hides in the corner;
0, action

Green arrives;
1, action</td></tr>
<tr><td colspan="2">2/6 = 0.33 , interaction</td></tr>
<tr><td colspan="3">Score: 0.33, interaction</td></tr>
</table>

<table>
<tr><td>Decision label: Play</td><td>Rephrase: False</td><td>#Play-through: 1</td></tr>
<tr><td colspan="3">Clip name: HStay-PlayWBoth.swf</td></tr>
<tr><td>Well, Green and Yellow are playing, like moving like that, and Blue went out and first he went like that with Yellow and then it spoke to Green, and Green bumps it. And we are at the end of the film story.</td><td colspan="2">Y&G dance;
2, mentalising

Blue speaks to B & G;
2, interaction

B&Y dance;
1, interaction</td></tr>
</table>

	G hits B; 2, interaction
	7/8 = ; mentalising
Score: 0.875, mentalising	

Transition: Play w both – fight green	**#Play-through:** 1
okay so green was fighting	
Informative: True	**Score:** 2, priming
Comment: demonstrates priming effect through the decision labels	

Decision label: Fight	**Rephrase:** False	**#Play-through:** 1
Clip name: PwB-FightGreen.swf		
Well what.. then... then Green and Blue started attacking .. and Blue got battered by Green flew into the box which part of it broke, and Blue was at the bottom of the box and didn't move and then green and yellow started fighting and faded away.	big Fight; 2, interaction Blue looses; 2, interaction kidnap Yellow; 1, interaction G&Y disappear; 2, interaction	
	7/8 = 0.875; interaction	
Score: 0.875, interaction		

Decision label: Play	**Rephrase:** False	**#Play-through:** 2
Clip name: House-Play.swf		
Well yeah then a green thing appeared. And I think Yellow and Blue were about, to get into a fight there.	B&Y play inside the house; 0, interaction Green arrives; 1, action	
	1/4 = 0.25 ; interaction	
Score: 0.25, interaction		

Decision label: Protect Yellow	**Rephrase:** False	**#Play-through:** 2
Clip name: HPlay-StayWYellow.swf		
Well then Green, Blue's tried to protect Yellow, and Blue got batted back to Yellow.	Door slam; 0,action Blue defends; 1, mentalising Blue blocks Yellow/Yellow wants outside;	

	0, action Yellow talks to Green 0, action Green hits Yellow; 0, action	
	1/10 = 0.1; mentalising	
Score: 0.1, mentalising		

Decision label: Fight Green	**Rephrase:** False	**#Play-through:** 2
Clip name: SwY-Fight.swf		
I see, I think they trapped Green in the box.	Blue wins fight; 0, action Green got trapped; 2, interaction B&Y embrace + disappear; 0, action	
	2/6 = 0.3; interaction	
Score: 0.3, interaction		

Transition: Play again - Intro		**#Play-through:** 2
One more time.. then.. play... leave I think..		
Informative: False	**Score: -**	

Decision label: Intro	**Rephrase:** False	**#Play-through:** 3
Clip name: House-Intro.swf		
What happened.. what happened there was ... same thing as all the other .. the other two times. The Yellow.. The Blue cube went inside the box moved around there for a while and then the yellow cube appeared outside.	Blue arrives; 0, action Blue explores house; 1, action Yellow arrives; 1, action	
	2/6 = 0.33, action	
Score: 0.33, action		

Transition: Intro – decision 1 (Hide/Stay)		**#Play-through:** 3
I don't see what happens if I press 'hide'.		
Informative: True	**Score:** 2, Question	
Comments: the functioning of the experiment is unclear		

to the participant even after two play-throughs.	

Decision label: Hide	**Rephrase:** False	**#Play-through: 3**
Clip name: House-Stay.swf		
P: Well I think Blue went inside the box and shut the door there.	Blue locks Yellow out; 1, action Blue hides in the corner; 1, action Green arrives; 0, action	
	2/6 = 0.33, action	
Score: 0.33, action		

Transition: decision 1 (Hide/Stay) – decision 2 (Play/Watch)	**#Play-through:** 3
E1: mhm. So is that what happened when you pressed the button? P: Yeah. E1: Was it what you expected? P: Well.. Yeah.. because I expected in 'hide' means going into the box.	
Informative: True	**Score:** 1, action

Decision label: Watch them	**Rephrase:** False	**#Play-through:** 3
Clip name: HStay-YellowVSGreen.swf		
apparently Green is rather hostile. There. And Yellow shut the door on Green. Apparently Green is rather hostile. Well .. E1: Why do you think that? P: Well that Yellow shut the door on Green. So, then the Yellow one and after seeing the Green was hostile going, went inside the box and shut the door.	Green attacks Yellow; 1, mentalising Yellow escapes 1, action Y enters house; 2, action Green tries to enter house; 0, action	
	4/8 = 0.50 ; mentalising	
Score: 0.50 , mentalising		

Transition: decision 2 (Play/Watch them) – decision 4 (Attack/Defend)	**#Play-through:** 3
..defend.. .. attack.. Two against one they should have a chance.	
Informative: True	**Score:** 2, mentalising

Decision label: Attack	**Rephrase:** False	**#Play-through:** 3
Clip name: YvsG-HuntGreen.swf		
I see they are shutting the door on him.	team up;	

E1: So why did you choose attack? P: I chose attack because seeing what Green would do, I knew they would have no chance at defending. E2: Why did you think that Robin? P: Because when.. because when Yellow tried top fight Green. In.. He had to.. Yellow had to retreat and shut the door, so Green is quite good at f.. at attack and but apparently he's not so good at defends	2, mentalising knock Green out + trap; 2, interaction hug+disappear; 0, action
	4/6 = 0.66, mentalising
Score: 0.66, mentalising	

Experiment #2

Reactions / Confusions - Evaluation		**#Play-through:** 1
Transition: Intro – decision 1		
E: Good. When you are ready just click on the 'Start'. P: [clicks and watches intro] Uhh! I am flying! Uhh it's locked in the box! E: So if you can describe that little piece of story for me? Just describe what happened. .. .You can watch it again if you want to. P: Well can you just... do the thing.. to know.... what.. actually is happening? E: Do you want to rewatch? E2: This is, P: Yeah. E2: This is the film. Yeah. This is the film, it's just a little animated … P: Hm. [watching the intro again] E: So what did the blue square do? For example. [user clicks on continue] Oh, can you describe it first please? P: What? What?		
Score: 2, stress		

Decision label: Intro	**Rephrase:** False	**#Play-through:** 1
Clip name: House-Intro.swf		
E: What did the blue square do in that little film? P: It went into the … other shape. E: and then what happened? P: An orange square came. E2: And then what happened? E: Is that all you see? What happened? P: mhh.	Blue arrives; 0, action Blue explores house; 1, action Yellow arrives; 1, action	
	2/6 = 0.33; action	
Score: 0.33, action		

Transition: decision 1 (Play/Hide)	**#Play-through:** 1
E: If you are unsure just press 'I don't know'. P: Hide? [clicks] P: [laughs]	

Informative: True	**Score:** 1, stress
Comments: Useful emotional regulation / stress insight.	

Decision label: Hide	**Rephrase:** False	**#Play-through:** 1
Clip name: House-Stay.swf		
E: So what happened in the film? P: Ehm. The orange one tapped to the lid and the green one came. E: What did the blue square do? P: He went into a corner at the bottom. E2: Okay E: Good. Do you want to re-watch the film or do you want to continue? P: continue	Blue locks Yellow out; 1, action Blue hides in the corner; 2, action Green arrives; 1, action 4/6 = 0.66, action	
Score: 0.66, action		

Transition: decision 2	**#Play-through:** 1
E: okay. So what should Blue do now? Should he play? Watch them? Or are you not sure? [click] P: [laughs]	
Informative: True	**Score:** 1, stress
Comments: emotional regulation – useful for software interface evaluation.	

Decision label: Watch them	**Rephrase:** False	**#Play-through:** 1
Clip name: HStay-YellowVSGreen.swf		
E: So what happened now? P: Ehm. The orange one got in and the green one was tapping.	Green attacks Yellow; 0, action Yellow escapes 0, action Y enters house; 1, action Green tries to enter house; 1, action 2/8 = 0.25 ; action	
Score: 0.25, action		

Transition: decision 4	**#Play-through:** 1
E: okay. What should Blue do now? E2: Attack or Defend, or you don't know? P: Attack!	
Informative: False	**Score:** -
Comments: decision exclamation	

Decision label: Attack	**Rephrase:** False	**#Play-through:** 1
Clip name: YvsG-HuntGreen.swf		
E2: What's happening. [while film is playing] P: Now he's trapped! E2: Okay, play it again! E: so what happened in the.. E2: Play it again! E: want to play it again? E2: Play it again and tell us what you think is happening. P: Hmm. I think he got him and they pushed him in and closed the lid and.. went away.	team up; 1, interaction knock Green out + trap; 1, interaction hug+disappear; 1, action	
	3/6 = 0.5; interaction	
Score: 0.5, interaction		

Transition: decision 4 - Intro		**#Play-through:** 1
E2: Yeah, so do you want to play it again and see if different things happen this time? P: Yay!		
Informative: True	**Score:** 1, Engaged	
Comments: Proof of engagement, other side of stress		

Decision label: Intro	**Rephrase:** False	**#Play-through:** 2
Clip name: House-Intro.swf		
P: He really is hiding! And the orange square is coming.	Blue arrives; 0, action Blue explores house; 0, mentalising Yellow arrives; 1, action	
	1/6 = 0.16; mentalising	
Score: 0.16, mentalising		

Transition: decision 1		**#Play-through:** 2
E: Do you want to redo it? E2: no. So what should Blue now do this time? Hide or Play? P: Play		
Informative: False	**Score: -**	

Software Evaluation		
Transition: Intro – decision 1		**#Play-through:** 2
E: What happened in the film? E2: if you play it again? [P clicks continue] Oh! To late.		

Decision label: Play	**Rephrase:** False	**#Play-through:** 2

<table>
<tr><td colspan="2">Clip name: House-Play.swf</td></tr>
<tr><td rowspan="2">E: What happened in the film? Can you please describe it?
P: ehm.
E: What did the blue square do?
P: He let the orange one in and played with the orange square.
E: And then what happened?
P: The green square arrived.
E: And then?
P: I don't know?
E: okay that's fine.</td><td>B&Y play inside the house;
1, interaction

Green arrives;
1, action</td></tr>
<tr><td>2/4 = 0.5; action</td></tr>
<tr><td colspan="2">Score: 0.5, action</td></tr>
</table>

<table>
<tr><td>Decision label: Talk to Green</td><td>Rephrase: False</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">Clip name: Hplay-TalkToGreen.swf</td></tr>
<tr><td rowspan="2">P: They are speaking
and the orange one's coming,
and they're running away from each other.</td><td colspan="2">Talk to Green;
2, mentalising

Tick;
1, action

Run;
2, interaction</td></tr>
<tr><td colspan="2">5/6 = 0.83, mentalising</td></tr>
<tr><td colspan="3">Score: 0.83, mentalising</td></tr>
</table>

<table>
<tr><td colspan="2">Transition: decision 5</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">E: what should Blue do now? Catch Green or catch Yellow or I don't know?
P: Catch Yellow!
E: okay..
P: Why are they not catching Yellow?</td></tr>
<tr><td>Informative: True</td><td colspan="2">Score: 2, Stress</td></tr>
</table>

<table>
<tr><td>Decision label: Catch Yellow</td><td>Rephrase: False</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">Clip name: TtG-ChaseYellow.swf</td></tr>
<tr><td>E2: So play it again. Do: play it again. And then tell us what you think is happening.
P: They are running away and, I don't know what's happening now.. I think they are getting.. Yellow trapped, I don't know what they want to do.
E2: Okay.
E: So what did the green square do?
P: Ehm.
E: you can re-watch it if you want.
P: the green square went the same way as the orange. The green square went in. so as the blue square.</td><td colspan="2">Blue chases Yellow;
0, action

Green blocks Yellow's path;
1, interaction

Blue catches yellow;
0, action

Blue locks Yellow out;
1, action</td></tr>
</table>

	play inside in circles; 0, action
	2/10 = 0.20 ; interaction
Score: 0.20 , interaction	

Transition: decision 5 - end	**#Play-through:** 2
E2: So do you want to continue? P: continue.	
Informative: False	**Score: -**

Transition: end - Intro	**#Play-through:** 2
E: So that again was the whole story. So if you want to watch another story, play again, or if you want... P: Yay! E: Okay!	
Informative: True	**Score:** 1, Engaged

Decision label: Intro	**Rephrase:** False	**#Play-through:** 3
Clip name: House-Intro.swf		
P: The blue one went into the shape and the lid was little open, the orange one came and the lid was still a wee bit open.	Blue arrives; 1, action Blue explores house; 1, action Yellow arrives; 1, action	
	3/6 = 0.5; action	
Score: 0.5, action		

Transition: decision 1	**#Play-through:** 3
E: okay, continue. So you know hide and play already, but just choose one. P: Hide!	
Informative: False	**Score: -**

Transition: decision 1 (Hide) – decision 2	**#Play-through:** 3
E: Can you describe what's going on, while you watch it? […] and then re-watch the movie. P: No! E: No, you don't want to? P: Continue!	
Informative: True	**Score:** 2, Stress
Comments: ignores the questions and storms on. Task might be overwhelming / intimidating for the participant, or just unclear.	

Decision label: Watch them	**Rephrase:** False	**#Play-through:** 3
Clip name: Hstay-YellowVSGreen.swf		
P: [laughs] Now the green square's stuck.	Green attacks Yellow; 0, action Yellow escapes 0, action Y enters house; 0, action Green tries to enter house; 0, action	
	0/6 = 0 ; action	
Score: 0, action		

Transition: decision 4 - end	**#Play-through:** 3
Well, what did the blue square do? P: Defend. What are they defending?	
Informative: True	**Score:** 2, priming
Comments: Priming example, useful for evaluation.	

Decision label: Defend	**Rephrase:** False	**#Play-through:** 3
Clip name: YvsG-Escape.swf		
E: Let's watch the movie again, it's pretty complicated. Replay. And describe as you see it. So what is the green one doing? P: Attacking the blue one. E: So what is the blue one doing? P: attacking the green one and getting him out of order. E: So what did the yellow one do then? P: it went out of the shape, E: and then? P: I don't know.	Yellow hiding behind Blue / Blue defending Yellow; 1, interaction Green destroying the door/breaking in; 0, action Green tries to get to Yellow; 0, action Yellow and Blue embrace + disappear; 1, action	
	2/8 = 0.25; interaction	
Score: 0.25, interaction		

Transition: description 'Defend'	**#Play-through:** 3
E: Do you want to re-watch it so you can tell me? P: no!	
Informative: True	**Score:** 2, Stress
Comments: evaluation of the experimental method	

Software Evaluation	

Transition: decision 4 - <u>end</u>	**#Play-through:** 3
That's again a whole story, so you can watch another if you want or you can have a break. P: Do you know what will have happened if you get a break?	

Software Evaluation	
Transition: -	**#Play-through:** 3
E2: So did you have some questions you wanted to answer Ruth about playing the game? [directed to E] […] So how did you find that Ruth? Was that easy or difficult? P: Easy! E2: Easy! And what things did you like about it? P: Ehm. When they tried to defend. E2: Ah was there any thing that you didn't like about it? P: Eh I don't know.	

<u>Experiment #3</u>

Decision label: Intro	**Rephrase:** False	**#Play-through:** 1
Clip name: House-Intro.swf		
So a box is moving in circles then it goes into a bigger box, then it just moves around randomly, then another box comes and it stops. E: Okay so, do you want to replay it or do you just want to say what happened, or you think you said what you think happened? P: I think I said what I think happened.	Blue arrives; 1, action Blue explores house; 1, action Yellow arrives; 1, action	
	3/6 = 0.5 ; action	
Score: 0.5, action		

Decision label: Hide	**Rephrase:** False	**#Play-through:** 1
Clip name: House-Stay.swf		
P: What happened is, Blue closed the box he was in, and so Yellow is trying to come in and then Green comes over.	Blue locks Yellow out; 1, action Blue hides in the corner; 0, action Green arrives; 1, interaction	
	2/6 = 0.33 ; interaction	
Score: 0.33, interaction		

Transition: decision 1 – decision 2	**#Play-through:** 1
E: Okay, do you want to replay it? And, oh that's ok. What should Blue do?	

P: Watch them. [giggle]	
Informative: True	**Score:** 1, Engaged

Decision label: Watch them	**Rephrase:** False	**#Play-through:** 1
Clip name: HStay-YellowVSGreen.swf		
E: Okay. So you can say while it is doing it what you think it is doing. P: Oh okay. So, Green and Yellow were having a fight, then Green is chasing Yellow, then Yellow opens the box and comes in with Blue.	Green attacks Yellow; 2, interaction Yellow escapes 2, interaction Y enters house; 2, interaction Green tries to enter house; 0, action	
	6/8 = 0.75; interaction	
Score: 0.75, interaction		

Decision label: Defend	**Rephrase:** False	**#Play-through:** 1
Clip name: YvsG-Escape.swf		
Then Blue is defending Yellow from Green, who is trying to get to Yellow. And then Yellow comes out of the box with Blue and then they go away with Green still in the box.	Yellow hiding behind Blue / Blue defending Yellow; 2, priming Green destroying the door/breaking in; 0, action Green tries to get to Yellow; 2, mentalising Yellow and Blue embrace + disappear; 2, interaction	
	4/8 = 0.50; mentalising	
Score: 0.50, mentalising		

Decision label: Intro	**Rephrase:** False	**#Play-through:** 2
Clip name: House-Intro.swf		
P: So again Blue moves around randomly, then into the box, leaving it open then goes in the corner to the next corner and to the middle and Yellow comes around.	Blue arrives; 1, action Blue explores house; 2, action Yellow arrives; 1, action	
	4/6 = 0.66; action	

Score: 0.66, action

Decision label: Play	**Rephrase:** False	**#Play-through:** 2
Clip name: House-Play.swf		
So I choose play and then Yellow comes into the box and they start chasing each other then Green comes around and they stop.	B&Y play inside the house; 2, interaction Green arrives; 2, action	
	4/4 = 1; interaction	
Score: 1, interaction		

Decision label: Talk to Green	**Rephrase:** False	**#Play-through:** 2
Clip name: HPlay-TalkToGreen.swf		
So they talk to Green and then, so they are talking, and then,..., they all walk away from each other.	Talk to Green; 2, primed Tick; 0, action Run; 2, action	
	2/4 = 0.5; action	
Score: 0.5, action		

Decision label: Catch Yellow	**Rephrase:** False	**#Play-through:** 2
Clip name: TtG-ChaseYellow.swf		
P: and then Blue is trying to chase Yellow while Yellow's going away with Green. And then Yellow comes back into the box with Green and Blue and they dance around.	Blue chases Yellow; 1, priming Green blocks Yellow's path; 0, action Blue catches yellow; 0,action Blue locks Yellow out; 0, action play inside in circles; 2, interaction	
	2/10 = 0.2, interaction	
Score: 0.2, interaction		

Transition: end	**#Play-through:** 2
E: Do you want to play again? Try a different variation?	

P: ehm, try a different variation.	
Informative: False	**Score:-**

Decision label: Intro	**Rephrase:** False	**#Play-through:** 3
Clip name: House-Intro.swf		
P: So once again Blue moves randomly around in the circle, opens the box and comes in, leaves the lid open, and runs around in circles from one corner to the next again in circles then Yellow appears.	Blue arrives; 1, action Blue explores house; 2, action Yellow arrives; 1, action	
	4/6 = 0.66; action	
Score: 0.66, action		

Transition: decision 1		**#Play-through:** 2
E: So why do you want.. what should Blue do? P: Ehm, hide. E: why? P: Because Yellow is trying to chase him.		
Informative: True	**Score:** 2, mentalising	

Decision label: Hide	**Rephrase:** False	**#Play-through:** 2
Clip name: House-Stay.swf		
And so Yellow is stuck outside the box with Blue in the corner and then Green appears.	Blue locks Yellow out; 1, mentalising Blue hides in the corner; 2, action Green arrives; 1, action	
	4/6 = 0.66; mentalising	
Score: 0.66, mentalising		

Transition: decision 2		**#Play-through:** 2
E: Okay, and what should Blue do now? P: He should play. E: Because? P: Because it would be nice.		
Informative: True	**Score:** 2, mentalising	

Decision label: Play	**Rephrase:** False	**#Play-through:** 2
Clip name: HStay-PlayWBoth.swf		

And so Yellow and Green were dancing around, then Yellow and Blue dance around, then Yellow and.. then! Blue and Green start dancing around.	Y&G dance; 2, interaction Blue goes out; 0, action B&Y dance; 2, interaction G hits B; 0, interaction
	4/8 = 0.5; interaction
Score: 0.5, interaction	

Transition: decision 3	**#Play-through:** 2
E: So what next? P: and so Blue is going to have a fight. E: Because? P: Because, because the others might just want to play, ..with ..ehm ..him, to trick him and then beat him up.	
Informative: True	**Score:** 2, mentalising

Decision label: Fight	**Rephrase:** False	**#Play-through:** 2
Clip name: PwB-fightGreen.swf		
So he gets into a fight and he gets pushed into the box and then Green and Yellow leave together.	big Fight; 1, priming Blue looses; 2, interaction kidnap Yellow; 0,action G&Y disappear; 2, action	
	4/8 = 0.5; interaction	
Score: 0.5, interaction		

Transition: end	**#Play-through:** 2
P: and i'll play a different variation.	
Informative: True	**Score:** 1, Engaged
Comments: independent navigation through the software – no prompt from instructor. Might be only a hint for automatism and familiarity with the experiment.	

Decision label: Intro	**Rephrase:** False	**#Play-through:** 3
Clip name: House-Intro		
So Blue is moving around in circles randomly for no	Blue arrives;	

apparent reason, opens the box, leaves the lid open, moves around into one corner, to the other, in circles, then Yellow appears. And he.. Blue looks at her.	1, action Blue explores house; 2, action Yellow arrives; 2, interaction
	5/6 = 0.83; interaction
Score: 0.83, interaction	

Transition: decision 1	**#Play-through:** 3
Which choice? P: Then Blue is going to play. E: Why? P: Because he was looking at her which probably means he likes her.	
Informative: True	**Score:** 2, mentalising
Comments: very strong ToM attribution here.	

Decision label: Play	**Rephrase:** False	**#Play-through:** 3
Clip name: House-Play.swf		
So they play around in the box chasing each other. Then they come look outside the box at Green.	B&Y play inside the house; 2, interaction Green arrives; 2, interaction	
	4/4 = 1; interaction	
Score: 1, interaction		

Transition: decision 3	**#Play-through:** 3
And Blue is going to, talk to green. E: Because? P: Because... just to chat because he could be bored.	
Informative: True	**Score:** 2, mentalising
Comments: very strong ToM attribution!	

Decision label: Catch Green	**Rephrase:** False	**#Play-through:** 3
Clip name: TtG-ChaseGreen.swf		
And then Blue is going to catch Green, because Green seems bored and so after chatting he still seams bored and so they play 'it', then Blue gets Green and they start moving around in circles and jumping about.	Blue chases Green; 2, mentalising Green is caught; 2, interaction they all play inside the house; 1, action	
	5/6 = 0.83; mentalising	

Score: 0.83, mentalising

Reactions / Confusions		**#Play-through:** 3
Transition: replay of Catch Green		
E: You can replay the same one, yeah. P: So I am replaying it. And, is this going backwards? E: No, it just showed you the same one again, P: Yeah		
Score: 2, Question		

Software Evaluation	
Transition: -	**#Play-through: 3**
E: So what did you think of that? P: Ehm. It was kind of boring. E: uh-huh, but it was kind of easy for you I guess? P: Yeah E: Okay. I think it would probably maybe younger children that would be doing it. Do you think, in terms of how it worked, how easy it was to use, did you find it easy enough? P: It was quite easy. E: Yeah, nothing you'd change in the design of it to make it an easier game or ..? P: Not really...	

Experiment #4

Decision label: Intro	**Rephrase:** False	**#Play-through:** 1
Clip name: House-Intro		
P: The blue square went into the box, then the yellow square … ehm.. went near the blue square.	Blue arrives; 0, action Blue explores house; 1, action Yellow arrives; 2, interaction	
	3/6 = 0.5 , interaction	
Score: 0.5 , interaction		

Transition: description – decision 1	**#Play-through:** 1
E: Would you like to watch the film again or do you want to continue? P: Continue.	
Informative: False	**Score: -**

Transition: decision 1	**#Play-through:** 1
E: okay, so if you just press on 'continue'. Awesome. So what should the blue square do? Should it hide, play or you are not really sure what it should do? or are not sure what that means?	

P: Uhm, Play.	
Informative: False	**Score: -**

Transition: decision 1	**#Play-through:** 1
E: Mh-hm. Do you want to re-watch it or just continue? P: Continue.	
Informative: False	**Score: -**

Decision label: Play	**Rephrase:** False	**#Play-through:** 1
Clip name: House-Play.swf		
P: uhm. The two squares played then a green square appeared and both of them the ..uhm the blue square and the.. orange square went together and went near the green square.	B&Y play inside the house; 1, interaction Green arrives; 2, interaction	
	3/4 = 0.75; interaction	
Score: 0.75, interaction		

Transition: decision 5	**#Play-through:** 1
E: okay. If you ever want to have a break just tell me, we can stop then. Alright so what should the blue one do? Should protect Yellow? Or talk to Green? Or I don't know? P: uh, talk to Green.	
Informative: False	**Score: -**

Decision label: Talk to Green	**Rephrase:** False	**#Play-through:** 1
Clip name: Hplay-TalkToGreen.swf		
P: uhm on of the squares talked to the green square and then they seperated. E: Mhm, nothing else? P: Uhm. Both of the squares went on the side and one went on in the middle. E: well, that's quite a long video, would you like to re-watch it and explain as we go? P: okay. E: Yeah? Just click Watch again and you just can explain. P: uhm it went around, and the Yellow.. the Green went together, then they went into the box, then went .. went into the circle. I think.	Talk to Green; 2, primed Tick + Run; 2, action	
	2/2 = 1; action	
Score: 1, action		

Transition: decision 7	**#Play-through:** 1
E: Okay. let's see, what should Blue do now? Catch Green, catch Yellow or I don't know? P: uhm, Catch Yellow.	
Informative: False	**Score: -**

<table>
<tr><td>Decision label: Catch Yellow</td><td>Rephrase: False</td><td>#Play-through: 1</td></tr>
<tr><td colspan="3">Clip name: TtG-ChaseYellow.swf</td></tr>
<tr><td rowspan="2">E: well, that's quite a long video, would you like to re-watch it and explain as we go?
P: okay.
E: Yeah? Just click Watch again and you just can explain.
P: uhm it went around, and the Yellow.. the Green went together, then they went into the box, then went .. went into the circle. I think.</td><td colspan="2">Blue chases Yellow;
0, action

Green blocks Yellow's path;
1, interaction

Blue catches yellow;
0, action

Blue locks Yellow out;
0, action

play inside in circles;
1, interaction</td></tr>
<tr><td colspan="2">2/10 = 0.2; interaction</td></tr>
<tr><td colspan="3">Score: 0.2, interaction</td></tr>
</table>

<table>
<tr><td colspan="2">Transition: description Talk to Green replay</td><td>#Play-through: 1</td></tr>
<tr><td colspan="3">E: Do you want to watch it again, or continue?
P: continue.
E: did you describe everything.
P: mh-hm.</td></tr>
<tr><td>Informative: False</td><td colspan="2">Score: -</td></tr>
</table>

<table>
<tr><td colspan="2">Transition: end</td><td>#Play-through: 1</td></tr>
<tr><td colspan="3">E: yeah? okay. So now the story is finished so if you want to try and see another story you can choose play again, or if you have enough, just have a break.
P: mhh. Play again.</td></tr>
<tr><td>Informative: False</td><td colspan="2">Score: -</td></tr>
</table>

<table>
<tr><td>Decision label: Intro</td><td>Rephrase: False</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">Clip name: House-Intro.swf</td></tr>
<tr><td rowspan="2">P: So uhm, the blue square went into the box and then uhm, went around it in a circle then the orange square appeared and then they went near each other</td><td colspan="2">Blue arrives;
0, action

Blue explores house;
2, action

Yellow arrives;
2, interaction</td></tr>
<tr><td colspan="2">4/6 = 0.66; interaction</td></tr>
<tr><td colspan="3">Score: 0.66, interaction</td></tr>
</table>

<table>
<tr><td colspan="2">Transition: decision 1</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">E: Mh-hm good, so what should Blue do now? Hide or play? Or I don't know?
P: uhm. Hide.</td></tr>
<tr><td>Informative: False</td><td colspan="2">Score: -</td></tr>
</table>

<table>
<tr><td>Decision label: Hide</td><td>Rephrase: False</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">Clip name: House-Stay.swf</td></tr>
<tr><td rowspan="2">P: Uhm the blue square shut the door to the big square then the orange square tried to get in, and then the green square appeared.</td><td colspan="2">Blue locks Yellow out;
1, mentalising

Blue hides in the corner;
0, action

Green arrives;
1, action</td></tr>
<tr><td colspan="2">2/6 = 0.33; mentalising</td></tr>
<tr><td colspan="3">Score: 0.33, mentalising</td></tr>
</table>

<table>
<tr><td colspan="2">Transition: decision 2</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">E: okay, continue. Mhm. So what should Blue do now? Play or Watch them? Or I don't know?
P: Watch them.</td></tr>
<tr><td>Informative: False</td><td colspan="2" rowspan="2">Score: -</td></tr>
<tr><td>Comments: solely a decision exclamation</td></tr>
</table>

<table>
<tr><td>Decision label: Watch them</td><td>Rephrase: False</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">Clip name: HStay-YellowVSGreen.swf</td></tr>
<tr><td rowspan="2">E: So what happened in the film?
P: Uhm,
E: You can re-watch it if you want.
P: uh mhm. The Yellow and the Green... the green one pushed the yellow one and bashed the door, then he went around and the yellow one square went inside and closed the door, and the green one tried to get him.
E: What did the blue one do?
P: Uhm. Stayed inside the box.</td><td colspan="2">Green attacks Yellow;
1, interaction

Yellow escapes
2, mentalising

Y enters house;
2, interaction

Green tries to enter house;
2, mentalising</td></tr>
<tr><td colspan="2">7/8 = 0.875; mentalising</td></tr>
<tr><td colspan="3">Score: 0.875, mentalising</td></tr>
</table>

<table>
<tr><td colspan="2">Transition: decision 4</td><td>#Play-through: 2</td></tr>
<tr><td colspan="3">E: Okay, so if you want to continue. What should Blue do now? Attack, Defend or .. ?
P: uhm, Defend.</td></tr>
<tr><td>Informative: False</td><td colspan="2">Score: -</td></tr>
</table>

Decision label: Defend	**Rephrase:** False	**#Play-through:** 2
Clip name: YvsG-Escape.swf		
The Green one.. square.. went.. got in.. a.. then the two up-there squares went out, then disappeared. E: Do you want to watch it while you are explaining it again? P: mhm. E: or if you can tell me what the blue one did? P: uhm it, tried to … defend. E: Should we maybe re-watch it? So what did the blue square do? P: uhm.. he.. uh.. tired to get away from Green. Uhm.	Yellow hiding behind Blue / Blue defending Yellow; 1, priming + 2, interaction Green destroying the door/breaking in; 1, action Green tries to get to Yellow; 0, action Yellow and Blue embrace + disappear; 1, action 3/8 = 0.375, interaction	
Score: 0.375, interaction		

Transition: end	**#Play-through:** 2
E: Mhm okay cool. Continue. Okay well that's the end of the story again so if you want to watch another story press play again, if you want to have a break, press have a break. P: have a break.	
Informative: False	**Score: -**

Experiment #5

Reactions / Confusions		**#Play-through:** 1
Transition: Intro		
E: So when you're ready just press on start, and that's the first bit of the story. So what happened in that little film? P: Ehm E: You can re-watch it if you want. P: What are you meant to do? E: Hm? Sorry? P: Ehm.		
Score: 2, Stress		

Decision label: Intro	**Rephrase:** False	**#Play-through:** 1
Clip name: House-Intro		
E: What did the blue square do, for example? P: So wait, went to that trap door thing E: mhm. P: and then went down into the room or something [laughs/sighs] E: mhm, and then what happened? P: and then another guy came in, he, the yellow square [laughs/sighs]	Blue arrives; 0, action Blue explores house; 1, action Yellow arrives;	

	1, mentalising
	2/6 = 0.33; mentalising
Score: 0.33, mentalising	

Transition: decision 1	**#Play-through:** 1
So now you can decide what Blue should do, either hide, play or if you are not sure press I don't know. P: [laughs/sighs] I don't know. E: Just choose one of them. P: [laughs/sighs]	
Informative: True	**Score:** 2, Stress
Comments: to much stress, the task is hard. Frustration and irritation comes up. This is what we were looking for. Hints to problems in mentalising. Not a proof but a hint.	

Decision label: Play	**Rephrase:** True	**#Play-through:** 1
Clip name: House-Play.swf		
E: [laughs] So what happened in this film? P: So the other guy went in as well, so the orange one went in as well and then a green guy came [laugh/sigh] E: Okay, what did the blue square do? P: Chasing him? Chasing the yellow square.	B&Y play inside the house; 2, mentalising Green arrives; 1, action	
	3/4 = 0.75; mentalising	
Score: 0.75, mentalising		

Decision label: Talk to Green	**Rephrase:** False	**#Play-through:** 1
Clip name: HPlay-TalkToGreen.swf		
What happened in the film? P: So, So yeah, the trap door was open and they went out and the Blue talked to the Green and the Yellow went [uff] there.	Talk to Green; 2, primed Tick; 0, action Run; 0, action	
	0/4 = 0; action	
Score: 0, action		

Transition: decision 5	**#Play-through:** 1
We can re-watch the film. P: I don't know. E: Well if you don't know, it's okay you can re-watch it as many times as you want. P: yeah.. I don't know. I don't know [laughs] [clicks continue/I don't know] E: well okay.. Just pick one. P: Uhm. Green.	
Informative: True	**Score:** 2, Stress

Comments: High stress!	

Decision label: Catch Green	**Rephrase:** True	**#Play-through:** 1
Clip name: TtG-ChaseGreen.swf		
So, so they all went around and the Blue I think chased the Green and then they chased each other around and the Blue and the Yellow one was hitting the Green and then they all went in again. Through the trap door again. I think.	Blue chases Green; 2, interaction Green is caught; 1, interaction they all play inside the house; 1, action	
	4/6 = 0.66; interaction	
Score: 0.66, interaction		

Transition: description – Catch Green	**#Play-through:** 1
E: mhm. Do you want to re-watch it? P: mhh. E: Are you sure? P: Yeah.	
Informative: True	**Score:** 1, Stress
Comments: Refuses watching the film again even though the participant would find the answer to the questions by re-watching the film.	

Transition: end	**#Play-through:** 1
E: Alright, so that was a whole story. And now you can play again and watch another story, or have a break. P: Have a break.	
Informative: True	**Score:** 2, Stress
Comments: Stressed participants wont play the game more than they have to. ie. Take the break when they can.	