Development and preliminary application of an instrument to detect partial dissociation of emotional mental state knowledge and non-emotional mental state knowledge

by

Stefan Scheepers

University of KwaZulu-Natal, Durban, RSA

November 2010

DECLARATION

Submitted in partial fulfilment of the requirements for the degree of

……., in the Graduate programme in

…….,

University of KwaZulu-Natal, South Africa.

I declare that this dissertation is my own unaided work. All citations, references and borrowed ideas have been duly acknowledged. I confirm that an external editor was not used. It is being submitted for the degree of …… in the Faculty of Humanities, Development and Social Sciences, University of KwaZulu-Natal, South Africa. None of the present work has been submitted previously for any degree or examination in any other University.

_____

Student name & surname

_____

Date

Abstract

Theory of mind is the ability to have mental states about mental states. Among theories concerning the structure and role of theory of mind is the view that theory of mind is the cognitive component of empathy. It is proposed that there is partial dissociation within theory of mind between emotional state representation and non-emotional state representation. In trying to test this hypothesis, an instrument was developed and implemented in a pilot study. Current theory of mind tests are reviewed and design features discussed in relation to the new hypothesis. The instrument aims to measure emotional and non-emotional state representation on separate subscales, as well as coding representations from emotional stories and non-emotional stories separately. The instrument was administered to 33 third level or higher students from the University of KwaZulu-Natal. Groups were chosen from science major (n = 9) and humanities major (n = 24) students. The findings fail to show the group performance patterns reported in literature, for example that humanities students tend to score higher in ToM tests than science students. A number of factors might contribute to the finding, but principally, low sample size and unequal general cognitive ability between groups are proposed as vital. Problems with the pilot study are identified and improvements suggested for subsequent testing.

Acknowledgements

With sincere thanks to David Spurrett for constant and patient guidance, to Jeff Martin, whose assistance in designing the electronic version of the instrument and conducting the empirical work was essential, and to Douglas Mansfield for introducing me to theory of mind.

Table of Contents

Chapter 1: Introduction

Attributing mental states to other beings, or persons, is part of normal human behaviour. We understand and predict the behaviour of others by seeing them as agents with beliefs, desires, emotions, intentions and more. Attributing mental states, to others or to ourselves, is to have a mental state about mental states. This ability to have mental states about mental states, or have meta-representations of mental states is called theory of mind (ToM). ToM can be described as a theory because the mental states represented are not directly observed. This ability is part of our 'folk psychology'; our way of making sense of other people's behaviour by reasoning about their mental life.

The term 'theory of mind' was first used in a study (discussed in further detail in section 2.1.1) by Premack and Woodruff (1978) in which they interpret the behaviour of a chimpanzee, Sarah, as most plausibly explained by Sarah having a theory of mind. Since the study extensive research has been done into many aspects of theory of mind including which non-human animals display evidence of the ability, how different populations of humans display the ability, what constitutes a test of theory of mind, what neural structures subserve the ability and what theory can best describe it.

One of the most active areas of research into theory of mind is Autism research. Autistic individual often show extreme impairment in social interaction. It has been proposed that this may be in part due to a lack on theory of mind; that Autistic people have an impaired ability to represent the mental states of others and understand them as mental agents. Others disagree and view Autism primarily as an affective impairment whereby sufferers are unable to relate emotionally to others.

A theory put forward by Baron-Cohen (2003) is that the theory of mind deficits in Autism are part of a wider deficit in empathy or empathising. Baron-Cohen claims that empathy is the ability to not only be aware of the mental states (thoughts, feelings, desires, etc) of others, but to respond appropriately to them. His view is that there are two components to this; a cognitive component and affective component. Theory of mind, Baron-Cohen claims, is the cognitive component – the ability to be aware of, or to represent the mental states of others. This component doesn't include an appropriate emotional response to others' mental states, which falls under the affective component of empathy.

This research considers the question of how theory of mind works in relation to other abilities – as a component of a more general ability or an ability more general to specific sub-components. The hypothesis is considered that within theory of mind, at least two sub-abilities can be conceptually defined – 1) the ability to represent emotional states and 2) the ability to represent non-emotional states. Further, a first, exploratory attempt is made to design and implement an instrument that could show these two abilities to be at least partly dissociated.

Although the current dissociation suggested is between the representation of emotional and non-emotional states, it is acknowledged that there is a strong possibility that if there is such dissociation, there may be further components that fall under non-emotional state representation.

In ToM literature, a number of terms are used to refer to theory of mind. Unless otherwise stated 'theory of mind', 'mentalising', and 'mind-reading' are used fairly interchangeably while I take the terms 'affect sharing' and 'empathising' to not share that meaning. However, there is disagreement between authors about the properties of these abilities and hence definitional differences. Part of the current project is to orientate the hypothesis and research in relation to existing research and theories, which requires comparing differences in their use of key terms. When doing this, alternative meaning of key terms are sometimes adopted.

Chapter 2: Literature Review

2.1 What Theory of Mind is

Theory of mind can be defined as the ability to have mental states about mental states. The mental representations taken to indicate having theory of mind (ToM) are a type of meta-representations (representation of a representation) and having 'mental states about mental states' is a good gloss of what is meant by having ToM. The discussions of ToM, are often focussed on knowledge of the mental states of *others* since most of the literature and research focuses on ToM as part of social cognition. However, most of what is said applies equally to representations of our own mental states, as opposed to simply having those states without being aware of them. Unless stated otherwise, 'theory of mind' refers to both abilities.

Part of the current task is to discuss debates about how this ability works and what it does and does not include. Although existing ToM models are discussed the main aim of this project is not to take a position on any complete model of ToM, but to focus on a specific part of ToM; the respective roles of emotional and non-emotional state knowledge in ToM tests.

2.1.1 Early theory of mind

The term 'theory of mind' originates in primate studies by Premack and Woodruff (1978) who tried to find out whether non-human primates interpret others' behaviour in terms of mental states. They tested a fourteen year old chimpanzee named Sarah, using a primary experiment and a number of further experiments. In the experiments, Sarah was shown videos of a human actor in various situations in response to which Sarah was required to respond by selecting one of a series of photographs. Sarah had

previous training in a simplified visual language between 4.5 and 6.5 year of age. She had also been previously tested on a variety of cognitive tasks for 5 days a week for 10 years before the study. The authors note that this could explain why the format of the video tasks didn't present a problem.

In the primary experiment, Sarah was shown four videos. In each video an actor encountered a problem when trying to get to bananas (for example the bananas hanging from above, out of his reach). Trainers showed Sarah each video, but paused it before the end and gave Sarah a box containing two photographs. In each picture the actor was doing something different, one of the two alternatives showing an action that would solve the problem in the video (for example showing the actor stepping onto a box) and the other not. The trainer left the room after giving Sarah the photographs. Sarah's task was to pick one of the two alternatives and put it next to the television, and ring a bell to signal her trainer to come into the room. The trainer recorded her choice and told her (using the appropriate tone) that the choice was right or wrong. Researchers showed Sarah each video a number of times, each time giving Sarah two different alternatives. Sarah correctly chose the option which represented the solution to the problem in 21 out of 24 trials. A second set of videos was made using 'broader' problems (such as an actor who is cold because a heater isn't lit) and used in a similar test. First, distracters (photographs not showing a solution to the problem) were used that were correct options to trials other than the one being tested at the time (for example the picture of the actor stepping onto a box is the correct answer in one trial, and it is used as a distracter in a trial where the problem is different). Next they used more complex incorrect alternatives, for example showing an unlit piece of paper, a burnt out piece, and (the correct option) a burning piece of

paper following the video of an actor 'being cold' in a room with an unlit heater. In this variant, Sarah picked the correct option 11 out of 12 times (Premack and Woodruff, 1978). The authors speculate that the results can be best understood by assuming Sarah inferred the actor's intentional mental state and the problem presented to this intention, and that she chose the alternatives that best fit the actor's goal (Premack and Woodruff, 1978).

It is worth noting that in addition to ToM, the authors considered two further possible explanations of Sarah's behaviour. One was associationism, which explains problem solving as completing familiar patterns. Sarah's behaviour would thus be explained as producing the next step (or indicating which would be the next step) in the sequence of events. The second alternative to ToM was empathy, whereby the animal 'puts itself in the actor's position' and reasons about what she (Sarah) would do if in the same situation. The authors claim that what sets empathy apart from ToM is that empathy doesn't assume Sarah can infer an actor's knowledge; her representation of the other's mind is limited to their *intention*. Subsequent mental states are 'what Sarah would do' if she herself had those intentions; her prediction of her own mental states given the intention of the actor within the situation, rather than a prediction of the actor's further mental states.

Three additional experiments were done to help the authors decide on which of the three interpretations was most plausible. I discuss these in minimal detail, partly for brevity and partly because the conclusions drawn from them are, according to the authors, speculative. In the first experiment, the problems depicted in videos were broader than a desirable object being physically inaccessible. The problems were: the

actor being locked in a cage, cold due a heater not functioning, the actor trying to play an unplugged phonograph and an actor trying to use a hose that isn't properly connected. In each case Sarah chose the photograph that solved the problem. The task was made more difficult by adding more difficult distracters such as an unlit piece of paper as a distracter when the correct option is a lit piece of paper and the problem is a heater that isn't lit. Out of 12 trials, Sarah made only one mistake. The authors see this experiment as reason to prefer ToM as an explanation for Sarah's behaviour over associationism, since the problems are too novel to be part of an event sequence with which she would be familiar. A further experiment was done in which the problem scenarios from the primary experiment were used, but two sets of videos were made, each using a different actor. According to the authors, they had reason to believe that Sarah liked the one actor but not the other. Two sets of possible responses were also used; a 'good' set – the same ones used in the primary experiment – and a 'bad' set – where the solution goes wrong (for example the actor falling over the box.) In trials using the actor Sarah liked, she chose the 'good' option 8 out of 8 times, while for those using the actor she disliked, she chose the 'good' option only 2 out of 8 times. Thought the authors admit that there could be multiple explanations of this, they note that the different responses for different actors suggests that Sarah does not simply adopt the actor's intention and pick what she would do if she were in the situation – empathy is not the best explanation of her behaviour. The authors don't reject the alternatives as explanations, but give reasons why they think ToM is a more plausible explanation. They claim that at least some of the test tasks Sarah performed are novel enough that she would not be familiar with the problems as part of a sequence and thus that her behaviour can be best understood as producing the next step in a known sequence. (Premack and Woodruff, 1978)

This study first named the skill in question 'theory of mind' and it, with the debate that followed, led to more detailed accounts of, and tests for, ToM. In response to the authors' claims, critics pointed out that Sarah's performance doesn't necessarily mean that she represented the actor's mental state. Only evidence of others' *false* beliefs could show mental state knowledge since someone (or something) else acting on their true belief is the same as them simply acting on the way the world is (Dennett, 1978, Pylyshyn, 1978). It is not necessary to represent their *belief* – one could simply represent how the world is and use this to predict their actions.

2.1.2 False belief representation as a test for Theory of Mind.

From these insights the 'false belief test' (FBT) was developed by Wimmer and Perner (1983) as a ToM test. It tests whether a subject is able to show knowledge of someone else's belief that is inconsistent with how the subject being tested believes the world is. This is done by showing subjects a short puppet show in which one character (commonly named Sally) places an object in a given location, for example a cupboard. This character then 'leaves' and a different character (commonly named Anne) moves the object to a second location, such as a basket. The first character then returns, and the subject is asked where this character would look for the object.[1] If the subject predicts behaviour based on the *other person's* belief about the world, which is different from how the world actually is, then the subject must be reasoning about the other person's mental states (intentions or beliefs) and not only relying on their own beliefs about how the world is. The FBT is discussed further in section 2.3.2.

[1] The original test was done using puppets but the modality of false belief tests varies extensively. Some false belief tests use puppets, acted scenes, some verbal or text based descriptions and some use cartoons. The constant is that a subject is somehow informed of another agent that believes something, which the subject knows to be false.

Initial findings showed that there is a sharp improvement on FBT performance between ages 3 and 5 with children younger than 3 years normally failing the false belief test and 5 year olds typically passing. Simon Baron-Cohen has written extensively about empathy (Baron-Cohen, 2003; Baron-Cohen & Wheelright, 2004) and ToM (Baron-Cohen, 1995). Baron-Cohen proposes that a deficit in ToM is characteristic of Autistics (Baron-Cohen, 1995) which could be the reason why Autistics have difficulty with social interaction (Frith, 2003). In a study comparing normal, Down's Syndrome and Autistic children's performance on the FBT (*n = 27, 14 and 20* respectively), Baron-Cohen et al found that 85% of the normal children and 86% of the Down's Syndrome children answered correctly when asked where the character that didn't see the object being transferred thinks the object is. By contrast, 80% of the Autistic children did not answer correctly, despite being older[2] than both other groups (Baron-Cohen et al, 1985).

Baron-Cohen also claims that it is more difficult for Autistic children to predict emotion based on beliefs than to predict emotion based on situations or on desires. Baron-Cohen compared the ability of normal children, mentally handicapped children and Autistic children to predict a story character's emotion. All three groups were told stories designed so that a story character's emotion can be inferred from one of three test conditions; the situation (e.g. it being their birthday), a desire (e.g. getting something that they say they want) or a belief (e.g. falsely believing that they are getting something they want). Baron-Cohen found that Autistic children predict emotions less accurately when the emotion to be predicted is the result of a belief than when it is the result of a desire or of a situation. In all cases, subjects were asked to

[2] Mean ages: Normal group = 4years 5 months; Down's syndrome = 10 years; 11 months; Autistic = 11 years 11 months.

identify the emotion and say why the character feels that way. Answering correctly on both these questions meant passing the test. 17.6% of the Autistic group passed the tests in the belief condition, compared to 56.3% of the mentally handicapped group and 73.7% of the normal group. For the desire condition, 52.9% of the autistic group, 50% of the mentally handicapped group and 89.5% of the normal group passed. Baron-Cohen takes this to support the view that Autistics have a specific deficit in representing beliefs (Baron-Cohen, 1991).

2.2 Deficits in Theory of Mind

2.2.1 Deficits in different aspects of Theory of Mind

This section discusses ToM deficits in more detail by looking at ToM functioning in a range of pathological conditions.

The failure of normal 2 year-olds and older autistic children on the FBT could be explained by the distinction Abu-Akel (2003) draws between the total absence of ToM – a conceptual/representational deficit – and the inability to implement it – an application/performance deficit (Abu-Akel, 2003). Thus, while older Autistic children might lack the ability to mentalise, normal young children could simply lack the processing, language or cognitive skills to apply their knowledge of mental states to a task such as the false belief test. This is given some support by findings that simplifying the task can lead to it being passed by 3 year olds (German & Leslie, 2000). It has been suggested that a conceptual deficit is present in autism and that Asperger's syndrome, an Autistic spectrum disorder less extreme than Autism, presents a performance deficit (Bowler, 1992).

In addition to the ToM deficits associated with Autism and Asperger's, schizophrenics with passivity phenomena[3] seem to have an impairment in representing their own mental states (Daprati et al, 1997) but no impairment in representing mental states of others (Pickup and Frith, 2001). Abu-Akel also suggests that positive symptom schizophrenia, such as paranoid or delusional schizophrenia, present instances of overactive mentalising (Abu-Akel, 2000) and this differs dramatically from the deficit in mentalising observed in Autism and Asperger's syndrome (Abu-Akel, 2003).

Abu-Akel (2003) claims that theory of mind is not something that has to be totally absent or defective but that a variety of impairments can be noted among sufferers of Asperger's, Autism and positive and negative symptom schizophrenia within the ToM domain (Abu-Akel, 2003). He suggests that the differences in deficits observed in a range of pathological conditions give insight into how ToM works and that we can identify a range of profiles of ToM impairment. This includes individuals with a lack of conceptual or representational understanding of mental states as observed in Autism. Those with negative symptom schizophrenia and Asperger's (and possibly normal young children) are proposed to have a representational understanding of mental states, but are unable to apply this understanding. Then there are those in which both an understanding of mental states and an ability to apply this knowledge is present, but the attribution of mental states functions abnormally – as in positive symptom schizophrenia – and those who seem to have an intact functioning of the understanding of the mental states of others, but impaired functioning of understanding their own mental states – as in schizophrenic passivity phenomena

[3] Passivity phenomena are delusions associated with schizophrenia whereby sufferers believe their behaviour is controlled by aliens or some other force outside of themselves.

(Abu-Akel, 2003). Neuro-anatomical evidence indicates that the same brain regions, specifically the inferior parietal lobe, function abnormally in patients with Autism and patients with schizophrenia. Additionally severe limbic system abnormalities have been found in patients with Autism and patients with schizophrenia (Abu-Akel, 2003).[4] This and the difficulties people with autism and people with schizophrenia have in responding to social and emotional cues (Abu-Akel, 2003) further supports the view that both conditions involve ToM dysfunction. Aside from the suggested role of empathising deficiencies observed in Asperger's, Autism and schizophrenia, psychopathic individuals have been characterised by an absence of empathy regarding the emotional states of others (Blair, 2003), as assessed in adults by the revised psychopathy checklist (Hare, 1991). What is most important in the above literature for my research is the view that ToM itself is an ability that comprises of various functions, and that dysfunction in parts can lead to a variety of impairments in what is called theory of mind.

The literature discussed above gives reason to believe ToM dysfunction plays a part in a variety of pathological conditions. Knowing how ToM combines with other abilities (e.g. empathy) and what sub-components it may have will help in understanding its role in normal human interaction. Exploring the possibility of operationally dissociable skills that fall under what is called ToM may be a useful step toward gaining such knowledge.

[4]Abu-Akel argues for the involvement of the limbic system in mentalising. He doesn't, however, make the distinction between mentalising and empathising and much of what he refers to as mentalising contains an emotional element.

2.2.2 Systemising and Empathising

Deficits in ToM in different populations may be interesting, but without a theoretical framework it is difficult to see how they are to help improve understanding ToM. In this section I consider a theory that proposes a possible underlying factor that might account for a number of ToM and empathising group differences.

The debate about whether Autism is primarily a cognitive or an affective deficit has been ongoing throughout the development of the ToM hypothesis of Autism (Tager-Frusberg et al, 1993). Baron-Cohen characterises empathising as being aware of someone's thoughts and feelings or 'taking their perspective', being aware that these might be different from your own and appropriately responding to others' internal states with one's own emotions (Baron-Cohen, 2003, p. 23-24). He views ToM as being the cognitive, in contrast to the affective, component of empathy which in total is seen as a broad awareness of and response to the mental states, including emotions, of others. He claims that the affective component of empathy is required for a person empathising with another to respond to the other person's emotional state with the appropriate emotion. The cognitive component is needed to *understand* another's emotions and take on their perspective, with no implication of an emotional state being *present* in the person taking on the others' perspective. This cognitive component encompasses awareness of not only emotional states but also non emotional states such as beliefs, intentions and desires. Though distinct, there is an area of possible overlap between the two components which falls under a 'mixed component' (Baron Cohen, 2003. p. 28-29). This might be where others' mental states are simultaneously understood/known (the cognitive component) and appropriately responded to emotionally (the affective component) or where one component depends

on the other, for example where appropriate emotional response depends on knowing certain mental states, or knowing another's mental states depends on having an appropriate emotional response.

Baron-Cohen claims that ToM, or mind reading,[5] is the core deficit in Autism and that Autistics are 'mind blind' or unable to mind-read. Autistics are not only bad at empathising (due to ToM deficits) but also display pockets of ability that typically involve a great deal of systemising – things based on input-output rules (Baron Cohen, 2003). The competence in systemising-focused tasks and the overwhelming prevalence of male sufferers of autism spectrum disorders – there are roughly four times as many male as female Autistics (Holtmann et al, 2007; Newschaffer et al; 2007, Hill & Frith, 2003) – suggests, according to Baron-Cohen, that generally the male brain is more specialised for systemising, and the female brain more specialised for empathising. Comparing male and female test scores and behaviour in multiple areas that involve a high degree of either systemising or empathising Baron-Cohen claims statistically significant differences along numerous variables, including the following empathic variables; females (compared to males) show more sharing and turn taking, empathic responses to others' distress and higher scores on empathy questionnaires. He also claims that they show greater competence in and earlier development of ToM, higher sensitivity to vocal and facial expressions, more emphasis on the value of social relationships and more developed and emotionally orientated language abilities (Baron-Cohen, 2002). Males show more rough playing and aggression and increased incidence of empathy disorders as well as being more likely to establish social dominance hierarchies (Baron-Cohen, 2002). Some

[5] The term refers to the ability to make inferences about others' mental states – having theory of mind.

systemising variables Baron-Cohen notes are as follows; males (compared to females) score better on mathematical tests, are better at construction of 3-D models and mental rotation of objects, better at tasks that require judging position and volume, better at map reading and recalling detail and score higher on systemising tests (SQ) (Baron-Cohen, 2002). This is further supported by data suggesting males are better systemisers than empathisers, and that they are better systemisers than women while the reverse is true of empathy; women are both better empathisers than men, and are better empathisers than systemisers (Baron Cohen, 2002; Lawson et al, 2004; Carol & Kim Yung, 2006).

Lawson et al (2004) compared performance on a systemising and an empathising task between a group of males with Aspergers syndrome (n=18), a group of males without Aspergers (n= 44) and a group of females from the general population (n= 45). Each group completed the social stories questionnaire (SSQ) and the physical prediction questionnaire (PPQ), which were posted to subjects. The SSQ is similar to the faux pas task (see section 2.4) and consist of 10 stories, each divided into 3 sections. Of the 30 story sections, 20 contain utterances by one character that could upset another character. Of the 20 'target' utterances that could upset other story characters, 10 are 'blatant' (easy to infer) and 10 are 'subtle' (less easy to infer). After each story section, subjects were asked if anything was said that could potentially upset any of the characters and to identify the line in which it was said. Each story also included a control question and only data from participants that answered all control questions correctly was included in the analysis. Scores were calculated from the number of correctly identified target utterances, with a maximum of 20. Group mean scores for the SSQ were 9.22 for Aspergers syndrome (AS) group, 12.02 for non-AS males and

13.62 for non-AS females. The authors report that non-AS females scored significantly higher than non-AS males ($p < .02$), and that non-AS males scored significantly higher than AS-males ($p < .002$).

The physical prediction questionnaire (PPQ) tests understanding of physical systems by presenting subjects with 40 items, each of which shown a mechanical diagram. The diagram shows the direction of movement of an input lever and an arrangement of connected objects with the 'device' depicted in the diagram. The subject is then asked to predict the output movement of connected levers or bobs by selecting one of five possible options all showing different output movement. One point is scored for each correct answer, resulting in a score out of 40 for each subject. Group mean scores for the PPQ were 28.18 for AS group, 28 for non-AS males and 16.18 for non-AS females. Non-AS females score significantly lower than their closest group, non-AS males ($p < .001$) but there was no significant difference between non-AS male and AS male scores ($p < .87$).

Evidence indicating higher competence in empathising is usually accompanied by lower competence in systemising and vice versa (Baron-Cohen, 2003) and the empathising and systemising differences proposed between males and females indicates, Baron-Cohen claims, different 'brain profiles'. Male brains tend to be more specialised for systemising, while female brains tend to be more specialised for empathy. He proposes that Autism is an extreme case of a systemising-specialised brain. Baron-Cohen refers to Autistic brain as the 'extreme male' brain (Baron-Cohen, 2003); a view proposed by Asperger in 1944 (Asperger, 1944). Research shows further group differences in systemising and empathising ability between

science and humanities students, with science students being better systemisers than empathisers and humanities students being better empathisers than systemisers (Focquaert et al, 2007). The Empathy Quotient (EQ) (Baron-Cohen & Wheelright, 2004) and the Systemising Quotient (SQ) (Baron-Cohen et al, 2003) were developed to measure empathising and systemising abilities respectively. Though the content differs, both tests present subjects with 60 statements, which they then rate on a likert scale to indicate how strongly they agree with each statement. Focquart et al tested 214 science students and 137 humanities students using the EQ and the SQ. Comparing the difference between EQ and SQ scores for each participant and performing analysis of variance showed a highly significant difference ($p < .001$) between the groups (Focquaert et al, 2007). It has also been found that Aspergers sufferers are stronger systemisers than empathisers (Lawson et al, 2004).

Singer (2006) argues that although the same thing is often meant by the terms 'theory of mind' and 'empathy', they are in fact two different abilities that rely on different neural structures. Based on her review of neuro-scientific and developmental literature on the topic, she argues that empathy (seen as the ability to share others' emotional states) and ToM (seen as the ability to understand mental states like intentions and beliefs) differ both neurally and developmentally (Singer, 2006). Though there are terminological differences between Singer's view that not all mental state representations have the same underlying neural structure or development and the dissociation I argue for, there is a strong similarity. This is that there is more to mental state representation than the ability to homogenously represent all mental states. Further, this is at least partly supported by the fact that representing emotional

states is importantly different (whether cognitively, neurologically, developmentally or some combination of these) from representing non-emotional states like beliefs.

2.3 Belief focussed Theory of Mind testing

In this section literature is discussed on ToM tests that focus on subjects' ability to have knowledge of other people's non-emotional states, the testing and scoring methods used, and some results from studies in which they were used. Tests that focus on the ability to know people's emotional states are discussed in section 2.4

A number of tests are labelled ToM tests insofar as they test subjects' ability to have mental states about mental states and require the subject to understand, infer or remember a mental state, either of another person (real or fictional) or of themselves. Various media are used in ToM testing, including cartoons (Gallagher et al, 2000), photographs (Mitchell et al, 2005; Baron-Cohen), text based stories/scenes (Kinderman et al, 1998; Stiller & Dunbar, 2007), acted scenes and puppet shows (Wimmer & Perner, 1983). Acted scenes, such as (some versions of) the false belief test described below, involve a scene being acted out complete with accurate facial expression, body language, gaze direction, etc. all of which can aid forming mental state representations. Alternatively, describing such scenes in text, as is done in the faux pas task (FPT) and imposing memory task (IMT), omits many of these aspects that are present in online, non-test situation where ToM might be used. It also limits testing to literate subjects, but makes tests easy to administer to many subjects and to control the information to which subjects have access. Even when stories are read to subjects by a researcher there are language demands (such as vocabulary) that make the task difficult for some populations, such as young children or those with language

impairment. ToM cartoons and written stories have been used to compare visual and verbal ToM tasks (Gallagher et al, 2000). Photographs have been used to show facial expressions to test recognition of emotions (Baron-Cohen, 2003). Further details about these differences in testing, including different modalities, are covered in more detail in section 4.1 both to gives a more complete idea of what constitutes ToM testing and to form the basis of discussion concerning research design decisions for my own research.

2.3.1 Implicit Theory of Mind tests

In this section, studies are discussed that focus on behaviour which implies having ToM, without requiring subjects to explicitly show awareness of another person's mental state by naming the state itself, answering a question or giving an explanation or prediction of behaviour that relies on the presence of any mental state. ToM tests do often require subjects to explicitly name a mental state or predict behaviour dependent on the mental state and examples of explicit ToM are discussed in the next section (2.3.2) focussing on kinds of mental state knowledge they test and how they are scored.

Pretend play and gaze (usually direction, target and duration) are taken to indicate something about a child's theory of mind without explicitly testing mental state awareness. According to Leslie's theoretical analysis on the subject (Leslie, 1987) the ability to engage in pretend play emerges at around 18 months at which age children can 'play along' with pretence. Leslie claims that the concept of pretence is in apparent contradiction to, or a distortion of, reality. Leslie (1987) uses an example of a mother pretending a banana is a phone by holding it up to her ear and talking.

Children seem to find pretend play interesting without showing evidence of confusing pretended states and real states; they know that the banana is not really a phone, but that we are just pretending it is. Leslie argues that meta-representation underlies the ability to make sense of such behaviour because it requires both the normal representation of a banana, and the meta-representation that 'banana-behaviour' is being (temporarily) substituted by 'phone-behaviour'. Making sense of others' pretence and pretending oneself involves behaving in some sense contrary to a held belief such as the belief that belief that the banana is not a phone[6]. While the false belief test requires recognising that someone can act on a false belief because they think it's true, understanding pretence requires that a child recognises that someone else can pretend something is true, while knowing they don't actually believe it, for example seeing an adult pretending the banana is a phone even though you know that that the adult knows it isn't. Though pretend play doesn't conclusively demonstrate mental state knowledge, it is according to Leslie the most plausible explanation of pretence.

In an fMRI study, German et al (2007) showed 16 normal adult subjects video clips of actors either performing, or pretending to perform, simple actions such as getting a book from a shelf or pouring tea. Comparing the degree of neural activity during each condition (pretended or real action) revealed that the areas that show the most activation during pretend actions compared to real actions were the medial prefrontal areas, inferior frontal gyrus, temporo-parietal regions and the amygdla. The authors claim a large body of ToM imaging research gives evidence that these areas are the same areas that show more increased activation than other areas during ToM tasks

[6] Friedman and Leslie (2007) emphasise that pretending is importantly different from behaving 'as-if' something is the case e.g. the mother doesn't behave as if the banana is actually a phone, but simply *pretends* it is.

(German et al, 2004). Autistic children are also known to show less spontaneous pretend play and the DSM IV-TR uses the lack of spontaneous pretend or imitative play as one of the criteria for diagnosing Autism (American Psychiatric Association, 2000).

Clements and Perner (2001) asked 3-year olds, who normally fail the false belief test (see 2.3.2), where the character in a false belief play would look for an object that had been moved without their knowledge. The authors found that despite most 3 year olds getting the answer wrong and answering that the character will look where the *subject* knew the object to be, the 3-year olds reliably look first, and for a longer period of time, at the location where the character would think the object is (Clements & Perner, 2001). Grice et al (2005) suggest, based on Event-Related Potential imaging data that in Autistic children the neural correlates of processing gaze direction may be delayed. Impairment in eye-to-eye gaze is also a diagnostic criterion for Autism as defined by the DSM IV-TR (American Psychiatric Association, 2000).

2.3.2 Explicit Theory of Mind tests

In this section I discuss explicit ToM tests; tests that require a subject to explicitly show mental state knowledge through an answer to a question or in some other report or behaviour.

If person A shows explicit knowledge of person B's false belief (which contrasts with A's belief) A shows knowledge of B's mental states without leaving room for the possibility (as mentioned in 2.1.1) that instead of representing B's belief, A might simply be acting on how A believes the world is. A test for recognising the possibility

of false beliefs in others, the false belief test is described in section 2.1.2. The participant watches a scene in which one person (Sally) place an object somewhere, for example a cupboard. Sally then moves away or looks away so she can't see what is happening to the object, and a second person (Anne) moves the object to a second location, a basket. Sally then returns and the participant is asked whether Sally would look for the object in the cupboard or the basket. To pass the test, the correct answer must be given; that Sally will look in the cupboard. This is taken as evidence that the participant recognises that others can have false beliefs, and thus have mental states – the participant is seen as having ToM.

Testing a subject's awareness of their own (previous) false belief typically involves presenting them with a container that normally contains something specific (e.g. a Smarties tube would normally contain Smarties). Researchers ask the participant what he/she thinks is inside and they usually say it contains Smarties. A researcher then shows the subject that it contains something other than Smarties (e.g. pencils) and then asks participant the test question, "what did you previously think was in the container?" If they show knowledge of their previous false belief by saying they thought it contained Smarties, they pass the test. If they say, as children younger than 4 typically do, that they previously thought it contained pencils, they fail the test as they fail to show awareness of their previous false belief. The vast majority of 5-year-olds (90%) pass this test (Baron-Cohen et al, 1985). In her review, Singer claims that similar results from studies in various cultures support the hypothesis that this developmental pattern is universal. Close scrutiny of this claim would require extensive review of a number of studies. For brevity, this has been excluded from this literature review.

However, there is some debate about the view that passing or failing the false belief task is a benchmark test of theory of mind. Bloom and German (2000) suggest that the false belief task involves more than just the knowledge that others have mental states. Beliefs normally aim at truth, so the false belief task requires not only an awareness of someone else's mental state, but recognising that this mental state is not performing the functional role it normally does. They also state that there children as young as 2 display knowledge of other minds in cases like pretend play and imitation of completed or intended action, so some that fail the FBT show other signs of mentalising (Bloom and German, 2000). In addition, other ToM tests, such as the 'screen task', are passed by older autistic children (Roth & Leslie, 1998). This task involves the placing of an object in a box which is next to a basket. Researchers place a screen in front of both the box and basket blocking both from view. A new box and basket are put in front of the screen and they then put something in the new box. They then take the object out of the new box and move it to the basket. Researchers then ask subjects if the object behind the screen is in the original box or the original basket. Normal 3-year olds typically fail this task, but as their ability to engage in social situations, pretend play and imitation shows, Bloom and German claim it would be inaccurate to view 3-year-olds and older autistic children as a homogenous group that all lack theory of mind because they fail at some false belief test (Bloom and German, 2000). The criticisms above raise the possibility that ToM tests might require cognitive capacities apart from ToM to be passed and it should not be assumed that these cognitive capacities have essential links with without testing if it is the case.

Wellman et al (2001) consider the impact that various conditions, including non-ToM task demands, have on FBT performance. In a meta-analysis of ToM - specifically

false belief research in children - they compare various study conditions to measure the influence they have on test performance. The analysis was done on reports and articles including 143 studies and 591 conditions (which include demographic details of the participants, type of task required, the nature of the target object used, whether or not there was deliberate deception, etc). Wellman et al found that a combination of four task conditions notably improved children's performance; i) whether or not the deception was the motive leading to the false belief, ii) the level of participation of the child in the scene (e.g. observer or actor), iii) how salient the mental state was in the scene (whether the mental state was inferred, stated or emphasised) and iv) whether or not the object that is the target of a false belief is real and present at the time the false belief question was asked. These conditions, together with the mean age and country of origin[7] of children were found to account for 55% of the variance in correct false belief test answers. This means some but not all of the performance differences can be explained as resulting from the abovementioned variables, undermining the view that eliminating non-test variables and administering a 'pure' false belief tasks will eradicate differences in performance between older and younger children. On the contrary, the authors claim the findings give strong support for the view that children's concept of others' mental states undergoes a change at the preschool stage of life which enables a separation between the real world and someone's 'mental world' or representation of the world (Wellman et al, 2001).

The imposing memory task (IMT) was developed by Kinderman et al (1998) to test mentalising ability. In the IMT, subjects are told a series of seven stories involving

[7] Country of origin analysed alone was found to be highly significant ($p < .001$). For example comparing Australia and USA showed that the odds ratio of passing the tests was 2.27:1 (favoring Australia as country of origin). This is of particular interest for the current study.

'meta-mental states', or mental states about mental states. For example, while I can have an emotional state like happiness, I can also have a belief about the mental state, like believing someone else is happy. I can then have a further mental state about *this* mental state, such as desiring to know if someone is happy or not, and so it can go on with higher levels of mental states about mental states. Each story is followed by a question set specific to that story and for subsequent questions within each set, higher levels of meta-representation are required to answer correctly.

In each question in the IMT, subjects must choose one of two statements as being the correct option. Complexity in the levels of meta-representation increases within each story's question set; the earlier a question appears in a set the fewer levels of meta-representation it requires and later it appears the more levels it requires. In all questions both potential answers involve the same complexity of meta-representation but only one statement is correct. Apart from the mentalising questions there were also 'memory' questions that tested a subject's recall of detail in the stories but didn't involve attributing mental states. Performance on this test gives a discrete score (0-7) using a weighted mean formula that includes how many questions were answered correctly, the levels of intentionality the questions require and the level at which the subject first answered incorrectly. 'Negative marking' (deducting score for incorrect answers) is also used to counter the potential effect of subjects answering correctly by who guessing the correct option by making it equally likely they will lose score for an incorrect guess.

Schenkel et al (2008) investigated ToM in children with bipolar disorder. As part of the research they developed the Affective Story Task (AST) to measure of ToM

performance in situations with differing emotional contexts. In the AST, stories are read to subjects. In all the stories a character is likely to draw an inference about what is happening because they are unaware of some other relevant information; they form a false belief about what is happening. For example in one story a girl laughs as she walks into the house because her friend had told her a joke. Her little sister is wearing a funny mask as she walks in and her mother sees her laughing. The mother is likely to infer that she is laughing at the funny mask, since she doesn't know about the joke her daughter heard earlier. Each story is designed to be positive-, negative- or neutral-emotional 'valenced'; a character has either a positive, negative or neutral emotional state resulting from the events in the story. This is to test whether subjects perform differently on ToM questions in the emotionally charged (positive and negative) stories and the neutral stories. Subjects are asked two questions about each story; a ToM question to test if they know the character in the story formed a false belief and a control question that tests story comprehension. If the control question is answered incorrectly the test question data is not used for that story. The task is designed to measure the effect that positive, negative or neutral affective valence has on recognising a false belief that a character in the story is likely to have (Schenkel et al, 2008). Each item in this task is scored as correct or incorrect, with questions requiring participants to correctly identify a character's false belief in each of the three story types. In the 2008 study each of the 26 subjects with bipolar disorder and the 20 control group subjects were given one story of each type. The bipolar group's mean percentage of correct answers was 64% for the positive and 58% for the negative condition and the control group's mean percentage of correct answers was 95% for the positive condition and 90% for the negative condition. The bipolar group performed significantly worse than the control group on positive ($p < 0.05$) and

negative ($p< 0.05$) story ToM questions. The bipolar group also performed significantly worse on negative story ToM questions than positive ($p< 0.01$) ones while there was no significant difference found between these within the control group ($p> 0.05$) (Schenkel et al, 2008).

Schenkel et al also use the hinting task by Corcoran et al (1995), in which characters in a short passage try to convey a message to another character by hinting rather than expressing their intention directly. For instance in one passage a girl's birthday is coming up and she says to her dad that she loves animals, especially dogs (Corcoran et al, 1995). Subjects were then asked what the character really intended by what they said. In the example above, the intention was to get her dad to get her a dog for her birthday. The task aims to measure the ability to infer meaning not explicitly stated. It contains 10 passages and a correct answer to the first question is awarded 2 points. If subjects answer incorrectly for the first question they are asked again with a further hint as to the correct answer (e.g. asked what the girl wants her dad to do). Correct answers after the second question are awarded 1 point. Subjects score 0 if after the second hint they don't answer correctly and answering questions results in the subject getting a score between the minimum of 0 and the maximum of 20. Schenkel et al found that the bipolar group did significantly worse than the control group on the hinting task ($p <0.001$) with the bipolar group scoring 70% mean percentage correct and the control group scoring 90% mean percentage correct (Schenkel et al, 2008).

The common feature in these tests is that passing them requires subjects to use and/or understand narratives that either describe or permit inferences about the mental states of others. But the tests also differ in a number of ways. They differ in the levels of

meta-representation they require. For example the hinting task tests only the subject's understanding of a character's first order mental state – their intention – whereas the items in the IMT range in mental state order, sometimes requiring a subject to understand, for example, a character's belief about a second person's desire for a third person to have a belief about a fourth person's desire.[8] The tests also differ in what specific mental states are required to be represented. The false belief task only requires representing a belief, while the hinting task requires representing a belief about the cause of a (very likely) known emotion, and the Faux Pas Task requires inferring an emotional state to pass, though it isn't explicitly tested (see section 2.4). Another ToM test, the IMT, doesn't require subjects to represent the emotional state of characters, but the information included in the story allows the subject to infer that a character has a certain emotional state. Another difference between tests, in addition to requiring or simply allowing affect representation, lies in how obvious this emotional state is. For example, in the hinting task, a character is said to be excited, play games, eat nice food and tell jokes (Schenkel et al, 2008)[9] and enters the scene laughing. Her emotion is a simple one and is not only easy for the subject to infer, but the subject's inference about the character's mother's belief about the character's emotional state is also fairly uncomplicated. This representation of the mother's *belief* about the daughter's emotion is not identical to the mother's belief about the *cause* of the emotional state may be false because the subject knows that the mother sees something in both the situation and the character's behaviour that leads her to infer the character's emotional state.

[8] See Appendix A
[9] See Appendix B

As shown, ToM test vary in which kinds of mental states they permit and require test subjects to represent. Which mental states are included and the degree to which they are available or required has also been shown to affect test performance in at least some tests. The effect of including different emotions has also been shown to not be equal for all populations. Especially when testing an ability argued to be central to understanding population differences, as theory of mind is argued to be, it is crucial that the potential of the included design features to affect test performance is not left up to chance. My aim in this project is to test whether a specific variable in ToM testing – inclusion or exclusion of emotional states on test items – will affect test performance in differently in students majoring in Sciences or Humanities. If it does, there is some support for the hypothesis that representing emotional state and representing non-emotional states is at least partly dissociable.

2.4 Emotion-focused Theory of Mind tests

In section 2.3 I discuss ToM tests that only or mostly test a participant's ability to represent another person's belief state. In this section I look at the aspect of ToM testing that focuses on people's ability to represent others' emotional states. I then look at the differences between more emotion-based ToM tests and more belief-based ToM tests.

The Faux Pas Test (Stone et al, 1998, Gregory et al, 2002) tests subjects' ability to identify and understand a faux pas, which is a violation of a norm in a social situation, and explain it in terms of characters' mental states, usually their emotional reaction. The task requires understanding a situation involving a mix of affective, belief and desire states, which are often not stated and need to be inferred by the subject. In the

adult version 20 short stories are read to the subject who then answers questions about these. Half of the stories contain a faux pas, the other half don't. Subjects are asked (1) whether someone said something awkward or something they shouldn't have. If they answer 'yes' they are asked 5 more question, (2) who said it, (3) why should they not have said it or why was it awkward, (4) why did the person say it. They are also asked (5) a mentalising question that relates to the potential faux pas situation and a (6) question about how a character felt. They are also asked (7, 8) two control questions, regardless of their answer to question 1.

Answers to all questions are rated as correct or incorrect and 1 point is given for each correct answer, with the exception of question 1 in control (non faux pas) stories, where a correct answer scores 2 points. In faux pas stories, answering 'no' to question 1 results in a score of 0 for the whole story. All incorrect answers score 0. For question 1 'yes' is scored as correct for faux pas stories and 'no' as correct for control stories. Answers to question 2 are correct if the correct person is identified. Correctly answering question 3 requires subjects to refer to the faux pas, but doesn't require them to identify the mental states involved themselves, though there is room for subjects to indicate this in their answers. Answers to 4 are rated as correct if they indicate that someone didn't realise or know something with reference to the faux pas. Question 5 tests whether participants know the faux pas was unintentional and saying it was unintentional is rated as correct. Question 6 is "[a] test of subjects' empathy for the story characters" (Stone et al, 1998, p. 23) and is rated as correct if it identifies the correct emotional state, such as anger, embarrassment etc. Correct answers for 7 and 8 require being accurate (rated by the scorer) about the details the questions call for and incorrect ones indicate that errors in test question should be interpreted with caution

Questions 7 and 8 are scored separately from other questions so comparisons can be made between ToM errors and non ToM errors. The maximum a subject can score is 60 and the minimum is 0 for faux pas related questions. This test has been used in a number of studies (Stone et al, 1998, Gregory et al, 2002; Stone et al, 2003; Shaw et al, 2007). However, the results from these studies are not as important for the current purpose as the design of the test itself.

Testing people's ability to recognise others' emotional states (often called empathy) is done by measuring the accuracy with which facial expressions are recognised. At about 12 months children show sensitivity to facial expression and respond to a novel object based on the perceived emotional state of another person. Mumme and Fernald (2003) studied the effect of the emotional reaction of an actress to objects on infants' responses to those objects. In two separate studies 10- and 12- month old infants (one age group per study, both =32) were tested in the study by showing the infants videos of an actress reacting to one of two objects in front of her. Her reactions were of 3 kinds that constituted three testing conditions; neutral, positive and negative affect. While the video was being played, the same objects as in the video were also on the table on which the monitor was and in front of which the infant sat. While the video was played, the objects were out of the infant's reach and when the video finished playing, the objects were moved (by a hidden researcher via a concealed mechanism) to be within the infant's reach. A 30 second 'play period' then occurred. The object the infant played with and how long (limited to a maximum of 30s) the infant played with the object and signs of change in the infant's facial expression were recorded. This resulted in two scores. The first of these was object touch, which is the

proportion of the play time that the infant touched the target object.[10] The second score was affect score. Raters judge each infant's facial expression every 3 seconds by playing back infants' recorded facial expression. Infants were given a score for each affect type (positive or negative). Neutral facial expressions were rated as 0. A positive affect score of 1 was given for slight smiles and raised eyebrows and a score of 2 for big smiles. A negative affect score of 1 was given for slight frowns or worries expressions and 2 for crying, scowls or big frowns. A score for each affect type was calculated by summing the scores for that affect and dividing them by the number of 3s intervals in the play period (not all infants played for the full 30s). The authors report that 12-month olds touched the target object significantly less ($p < .01$) during negative affect conditions than in neutral conditions and showed significantly more negative affect after negative affect conditions than after neutral conditions ($p < .01$). 10-month olds showed no significant difference in their affect or object interaction for different emotional conditions (Mumme & Fernald, 2003).

Psychopaths have an impaired ability to recognise facial expressions of fear, according to Blair et al (2004). Blair et al compared 19 control individuals with 19 psychopathic individuals, assessed using the Revised Psychopathy Checklist (PCL-R) (Hare, 1991). Each group did the emotional expression multi-morph task, in which a participant is shown a series of pictures with each of six emotional expressions; happy, surprised, disgusted, angry, sad and fearful. Participants underwent three testing trials for each emotional expression. Each picture starts with an expressionless face which gradually morphs into one of the prototypical emotional expressions

[10] The total time touching the target object divided by the play period duration, the latter being dependent on how much of the maximum allowed 30 seconds the infant spent playing with the objects.

mentioned above. There are 20 stages that each picture progresses through, with each consecutive picture showing the emotion more strongly than the previous one. Participants were asked to say what the expression was as soon as they thought they knew what it was without guessing. Participants were given a score (min. 0, max 20) calculated by the number of stages of the morph they had seen before correctly identifying the emotion. For example, identifying the prototypical emotion after seeing all the stages resulted in a score of 1, but if they got the answer right 7 stages before the prototypical expression, they scored 7. Getting the answer right on the first picture, 20 pictures before the prototypical expression would be shown, would result in a score of 20. Not recognising the emotion correctly resulted in a score of 0. The scores from all subjects on all 3 trials were combined for each emotion to arrive at a mean score for each group. Though the authors found the psychopathic group to be generally less sensitive to emotional expression, the only group scores found to be significantly different were those for expressions of fear ($p<0.01$) where the psychopathic group mean expression recognition score was 4.68 and the comparison group scored 7.87 (Blair et al, 2004).

The 'Reading the Mind in the Eyes' (RME) test was developed by Baron-Cohen and Wheelright as a test of empathising skill (Baron-Cohen et al 2001; Baron-Cohen, 2003). It presents subjects with 36 photographs of the eye region of faces expressing emotions. The subjects are instructed to pick the word (out of four possibilities) that best describes what the person in the picture is thinking or feeling. A subject's total number of correct answers is a score out of 36. The normal range is 22-30 range with a higher score showing the subject is good at decoding eye-region facial expression, and less indicating the subject had difficulty with the task (Baron-Cohen, 2003).

Despite the instructions asking subjects to indicate what someone is thinking or feeling, the task differs from the non-emotional tests discussed in 2.3 insofar as the states of others that the participant is required to show knowledge of are free of the kind of representation found in beliefs where the belief represents some propositional attitude. So even if a judgement is made about what someone is thinking, it is a judgement about the kind of mental state they are in, rather than a representation of the content of their belief, as is found in false belief tests.

Baron-Cohen et al (2001) administered this test to four groups. Group 1 were adults with Aspergers Syndrome or High functioning Autism ($N = 15$). Group 2 were normal adults ($N = 122$). Group 3 were normal adult Cambridge University undergraduate students ($N = 103$). Group 4 were randomly selected from the general population who were IQ matched with group 1 ($N= 14$). Mean group scores for the RME were as follows; Group 1 = 21.9, Group 2 = 26.2, Group 3 = 28, Group 4 = 30. One way ANOVA comparing the RME scores of the four groups showed a significant main effect of group ($p = .0001$). The authors also claim that group 1's performance was significantly worse than the other three groups, whose performance did not differ significantly, but effect sizes were reported.

Groups 1, 3 and 4 also completed the Autism spectrum quotient (AQ) – a measure of autistic traits in adults. Mean group AQ scores follow; group 1 = 34.4, group 3 = 18.3 and group 4 = 18.9. For all three groups combined, data analysis showed no significant correlation between AQ score and IQ score ($r = .05$, $p = .77$) or IQ score and RME score ($r = .09$, $p = .6$). As the authors suspected, there was significant inverse correlation between AQ score and RME score ($r = -.53$, $p= .004$). The authors

take the results to validate the RME test as a test for social intelligence (Baron-Cohen et al, 2001).

2.5. Neuro-imaging and Neuro-anatomy

In this section I discuss neuro-imaging and neuro-anatomical research that focuses on the abilities discussed so far. I aim to show a few of the important areas of agreement and disagreement and orientate my research in relation to these. I aim to show that despite disagreements about the precise neural structures subserving ToM, the view that it is a complex skill is common and there is some evidence that non-overlapping or partially non-overlapping neural circuitry sustains abilities that have been researched under the heading 'theory of mind'. If there are non-overlapping substrates that underlie abilities claimed to fall within ToM, then there is some reason to think these abilities might be partly dissociable by suitable testing.

In a review of literature on the neural basis and development of ToM and empathy Tania Singer (2006) argues that the neural substrates of mentalising and empathising are non-overlapping and have different developmental trajectories (Singer, 2006). Singer sees mentalising as the understanding of intentions, goals and beliefs or attributing propositional attitudes to people. She views empathising as the ability to share emotions and allows us to experience what someone else feels (Singer, 2006). The view that empathising and mentalising rely on different neural circuitry is largely based on differences in brain areas implicated by research in each of the two abilities separately since there is little (or no) research that directly compares neural activity during empathising and mentalising. These studies typically involve brain imaging while a ToM or empathising task is being performed, or comparing task performance

with or without the proper functioning of certain areas, as in lesion studies or neurosurgical studies. Singer's view is that mentalising relies mostly on the medial prefrontal cortex (mPFC), superior temporal sulcus (STS) and temporal lobes, while the limbic/paralimbic system, including the amygdala, has been linked to the ability to recognise and respond to the emotions of others, i.e. empathise (Singer, 2006). I now discuss some evidence for this view.

2.5.1 Imaging studies

2.5.1.1 Imaging studies in Theory of Mind

Gallagher et al (2000) used fMRI scanning to compare neural activity during verbal and non-verbal ToM tasks in 6 normal subjects. Subjects did a story comprehension task (the verbal task) in which they were show texts, each followed by a question which they were instructed to 'answer internally' without saying their answer out loud. There were three story conditions; (1) ToM story, (2) Non-ToM story or (3) unlinked sentences and each story was preceded by a prompt to indicate to the subject which of the 3 conditions was being tested and each subject completed 4 tasks under each of the 3 conditions. After scanning each subject, researchers showed them the text again, repeated the question and asked what they answered, giving a score of 1 for a correct and 0 for an incorrect answer. All subjects scored the maximum possible.

The subjects also completed a cartoon task (non-verbal) in which they were shown a series of caption free pictures. Each picture was a ToM cartoon (where the humour depends on attributing either a false belief or ignorance to one or more characters), a non-ToM cartoon or a jumbled picture. The jumbled pictures were an ensemble of cartoon animals, people and objects randomly positioned and unrelated. Each picture

was preceded by a non verbal prompt indicating of which of the three types it was. For jumbled pictures, subjects were instructed 'just to look'. For ToM and non-ToM cartoons while they were instructed to consider the meaning silently. After the fMRI scanning researchers again showed the pictures to subjects and asked them to explain the meaning; correct answers scoring 1 and incorrect ones scoring 0. Scores were close to the maximum of 28 for all subjects on all conditions. FMRI data showed the mPFC and the temporo-parietal junctions (TPJ) to be the only regions showing significantly increased activity from baseline in both the verbal and non verbal ToM items and without also showing increased activity in non-ToM items.

Mitchell et al (2005) researched the effect of perceived self/other similarity on brain activity while mentalising. 18 subjects with no history of neurological problems were shown photographs of faces and were asked to either make ToM or non-ToM judgements while undergoing fMRI scanning. Each subject was shown 240 photographs of faces (all with neutral expression) and asked either how symmetrical the face was (non-ToM) or how pleased the person was to have their picture taken (ToM). Which condition (ToM or non ToM) was asked about which faces was randomly decided and after 30 minutes subjects were again shown the same faces and asked to rate how similar each target was to them on a 4 point scale (1 being 'very dissimilar to me' and 4 being 'very similar to me') to obtain a measure of the subject's perceived similarity with the target. Areas in which greater activation was found in ToM than in non-ToM tasks included the mPFC and TPJ the right STS. However the amygdala also showed increased neural activity. When neuro-imaging data for trials was analysed together with rated similarity of trial targets, the authors

found that only a region in the mPFC showed increased activation on ToM over non-ToM trials for targets rated as highly similar. (Mitchell et al, 2005)

2.5.1.2 Imaging studies in Empathy

Carr et al (2003) used fMRI to record the neural activity of 11 healthy subjects during two experimental conditions; observing facial emotional expressions and imitating such expressions. Subjects were screened for medical or behavioural disorders with a questionnaire and neurological exam. Subjects were shown 3 sets of picture of emotion expressing faces (happy, sad, angry, disgusted, surprised, and afraid). In one set whole faces were shown while in the second and third only the cropped eye or mouth regions (respectively) were shown. In each set shown there were 24 pictures and for each picture subjects were instructed to either imitate and internally produce the emotion target emotion, or to simply observe. The two conditions (imitate/observe) were counterbalanced in all sets, and there was a rest period (24s) between sets. Imaging data showed no significantly different neural activity in the three imitation tasks or the three observation tasks. Activation patterns were largely similar during observation and imitation conditions with the premotor face area, the superior temporal sulcus, the insula and the amygdala showing common activation patterns but greater activity during imitation than observation (Carr et al, 2003). The authors take the greater activation in these areas as support for their hypothesis that we understand others' emotions through action representations, but it by extension supports the view that these areas are used in empathising; mimicking emotions and emotional expressions.

Wicker et al (2003), also in an fMRI study, focus specifically on disgust and pleasure. Their study looks at neural activity while either observing or experiencing an emotion. 14 males were shown short (3s) movie, while in the scanner, of actors smelling the contents of a glass and then reacting with a neutral, pleased or disgusted facial expression depending on the contents of the glass. The glass contained one of three liquids (making three distinct test conditions); water, pleasant odorant or disgusting smelling odorant. Six actors were used in filming. Each actor recorded 6 movies under each of the three conditions, making three condition sequences of the 6 movies (one movie per actor with all actors under the same test condition). Each of these condition sequences were shown twice (consecutively, with a pauses separating them and change of movie order) to each subject. This constituted the visual runs of the three experimental conditions.

Subjects also did olfactory runs in two of the conditions while being scanned. Subjects were given either pleasant or disgusting odorants to smell via an anaesthesia mask. 20 odorants were used in each olfactory run split into 2 sets (pleasant and disgusting). Each set contained 10 different odorants each with a different rating of intensity (1-10). Each olfactory run contained 8 trials of each condition (pleasant and disgusting) and each participant did two runs separated by a break. The imaging data led the authors to conclude that observing disgust in others activates neural substrates that are otherwise show selective activation when feeling disgust; the left anterior insula and the right anterior cingulate cortex (Wicker et al, 2003).

2.5.2 Surgical/Lesion studies

2.5.2.1 Surgical/Lesion studies in Theory of Mind

A number of studies on neural damage or neurosurgery also lend some support to Singer's view of the different neural substrates underlying empathising and mentalising. Apperly et al (2004) researched 12 adults with brain lesions in the frontal, parietal and/or temporal lobes. Subjects were shown a video of an acted false or non-false belief situation where an object is in one of two identical boxes and an actress knows which of the two boxes contains the object, but the subject being tested doesn't know. In false belief trials the woman then leaves the room and a man then swaps the position of the boxes around without the woman knowing. She then returns, gives a clue to the subjects about where she thinks the box is and the video is paused. Subjects are then asked were the object is and then feedback was given by continuing the video, revealing the location of the object. In non-false belief trials, the man does not swap the boxes around when the woman leaves the room. The study also included story based false belief tasks which consisted of 12 short false belief stories, followed by a false belief question and control questions with all questions having a choice of two answers. Subjects did these in 3 sessions of approximately 20 minutes each, separated by 1 or 2 weeks. As part of the study, Subjects also did executive function (working memory and inhibition) and language tests, the details of which are not crucial for the current research. Four patients did not perform significantly above chance on the video FBT, and also made errors on control trials. Three other patients did not perform significantly above chance on the video FBT, but made no errors on control trials. The authors see this as suggesting a specific difficulty in reasoning about beliefs (Apperly et al, 2004). All three these patients also had lesions in the left TPJ and this, according to Apperly et al, adds to existing evidence of the region's

involvement in false belief reasoning and shows that even lateralised damage to it impairs false belief reasoning. However, only one of the three patients showed the same score pattern in the story based false belief task by scoring above chance on non-false belief questions but below chance on false belief questions (Apperly et al, 2004).

2.5.2.2 Surgical/Lesion studies in Empathy

Several regions that form part of the limbic/paralimbic system have been proposed as key components in ToM including the amygdala and anterior cingulate gyrus (ACG) (Abu-Akel, 2003). The claim is supported by studies such that use tests such as 'reading the mind in the eyes', where participants are required to infer emotional/mental states from pictures of people's eyes and asked what the person is thinking or feeling, or 'faux pas recognition'(Stone et al, 2003). Stone et al (2003) compare two subjects with acquired bilateral amygdala damage to a two control groups of 10 and 24 subjects. The test group and the control groups did two tests; a faux pas recognition task and a reading the mind in the eyes task (both described earlier). The FPT contained 10 items and the RME of 25 items. Both test participants did significantly worse than controls in FPT test questions, but not control questions. Both also scored below average for the RME, but only one significantly, though authors note that this may be due to the low number of test items (Stone et al, 2003).

Shaw et al (2003) compared 19 epilepsy patients with anterior temporal lesions and subsequent surgical lobectomy with 19 healthy controls with no history of disorder. All subjects in the test group had either a right or left anterior lobectomy (LATL or RATL) which included the removal of the left (LATL, n= 9) or right (RATL, n= 10)

amygdala. Control subjects did ToM tests, in the form of a faux pas task and a strange stories task (Happe, 1994), and a facial emotion recognition test, while test subject did these 1-3 months before and 4-6 months after having an anterior temporal lobectomy. The FPT was similar to that discussed in section 2.4. In the strange stories task, characters in a story say something not meant literally. The subject's understanding of why the character said it is then tested. Correctly identifying a mental state scored 2 points, an implicitly correct or contextual mental state answer scored 1 point and answering incorrectly or not referring to mental states scored 0 points.

In the facial emotion recognition task, subjects were shown pictures of 4 people (2 men, 2 women) with pictures of each showing 6 different facial expressions; anger, sadness, disgust, happiness, surprise and fear. A face would be shown to subjects along with an emotional term naming one of the six emotions. Below this was a rating scale ranging from 1 (not at all) to 10 (very much) which subjects had to use to rate the intensity of the face with respect to the emotional term shown is to the face being shown. Each face was shown for the same emotional term before the face cycle started again, but with a different emotional term. This repeated until the face cycle had been shown with each of the emotional terms. Two score ratings were allocated to this task; a congruent rating for question where the emotional term correctly described the face shown and an incongruent rating for question where the emotional term did not correctly describe the face shown and a discrimination index was calculated by finding the difference between congruent and incongruent ratings. A high discrimination index indicates a subject generally rates a particular emotional expression as highly proto-typical, i.e. dissimilar from other emotional expressions (Shaw et al, 2007).

The authors' main findings were that pre- and post-operative ToM scores did not differ significantly and that pre-op bias for misinterpreting fear is corrected after a LATL. This is suggested by the near significant difference in rating ($p= 0.06$) found between pre operative ($M= 4.68$) and post operative ($M= 2.85$) trials where a face showing fear was shown, and subjects were asked to rate the intensity of other emotions (incongruent trials) and the significant difference between pre-operative and post-operative discrimination index ($p= 0.002$) for the LATL group. The authors also note that other studies show the amygdala to be a vital component in ToM and speculate that the this discrepancy may be due to their small sample size or the fact they use two ToM measures that are not typically used in the imaging studies where the amygdala is found to be active. They maintain though that their study shows that at least some aspects of social cognition are not harmed by the surgical removal of these areas of the brain (Shaw et al, 2007).

Singer (2006) states that the ability to infer the cognitive states of others is predicted to mature later in the individual's development than the ability to share emotional states, as affective states rely on structures that mature earlier in ontogeny than the structures recruited when mentalising (Singer, 2006; Gogtay et al, 2004). Studies on cortical development over time support this statement. Gray matter density (GMD)correlates with cognitive functioning and maturation (Frangou et al 2004) and differences in GMD were compared across participants aged 7 to 87 by Sowell et al (2003). They found that the first areas to fully myelinate were those in the visual, auditory and limbic cortices. These areas also showed a more even, linear decline in GMD compared to the frontal and parietal neocortices (Sowell et al, 2004). The lateral temporal lobes and prefrontal cortex mature late, with the dorsolateral

prefrontal cortex (DLPFC) only fully matures at the end of adolescence and the STS (Singer, 2006), constantly implicated in mentalising, matures last (Apperly et al, 2004). The different areas also show different rates of decline in old age, with the PFC (Singer, 2006) and temporal regions (Sowell et al 2003) showing sharp deterioration in old age, congruent with the decline in mentalising abilities during old age while empathy seems to remain intact (Singer, 2006).

Though there is some evidence that empathising and mentalising rely on non-overlapping neural substrates, there are many situation in which both are useful and both form part of more general social cognition. Baron-Cohen notes that there is a difference between having an appropriate emotional response to someone else's emotional state and having knowledge of someone's mental state (Baron-Cohen, 1995). Baron-Cohen uses the term 'empathy' to refer to the whole skill set of understanding someone's mental state and responding to it with an appropriate emotion. He then sees empathy as comprising of two components; one being an 'affective component' that allows us to respond to someone's emotional state with an appropriate emotional of our own, and one being a 'cognitive component' which lets us understand and/or predict the mental states of others. It is important to note that the representation of others' emotional states falls within the cognitive component. This is so because our simply knowing their emotional state doesn't necessarily mean we respond with an (appropriate) emotion of our own. For example if something terrible happened to my close friend, I would both be aware of his emotion and have my own emotion in response to his pain. Yet if the same thing happened to my worst enemy, I can be completely aware of his pain through the cognitive component of empathy, but

fail to respond with any emotion at all. Baron-Cohen refers to this cognitive component as mind reading or ToM.

Thus his view is compatible with the evidence presented by Singer for there being different areas specialising in representing belief/intentional states and affective mirroring. These very neatly map onto his cognitive and affective components of empathy and the difference is purely terminological. What is worth making clear though is that Singer's view doesn't discount the fact that *knowledge of* emotional states, without an emotional reaction to these states, falls under ToM as Baron-Cohen states.

2.6. Literature Suggesting Possible Dissociation.

My research aims to answer some of the questions raised by Singer 2006 and take up her recommendation that research is needed that investigates in further detail than current tests when mentalising and empathising abilities play common roles and when they can be separated (Singer, 2006).

Existing disagreement regarding the regions activated in theory of mind tasks might be a result of the lack of a common understanding regarding exactly what theory of mind is. Some see it as the representation of the states of others, including affective states (Abu-Akel, 2003; Stone et al, 2003) while others think it is the ability to understand our own and other people's mental, or cognitive states such as thoughts and beliefs (Bloom and German, 2000; Singer, 2006) also referred to as propositional attitudes (Singer, 2006). It seems as though a distinction between the representation of mental and emotional states of others (and ourselves) would be helpful. I will use the

terms *mentalising* and *theory of mind* to refer to the understanding, inferring or use of other people's cognitive states such as beliefs thoughts and intentions. *Empathy* can be seen as the ability to recognise and share the emotions of others, even in absence of direct emotional stimulation though the term is often used in everyday speech to describe a mix of emotions and mental states (Singer, 2006). Though these two terms have at times been taken to mean much the same thing and are undoubtedly related, treating them as identical in research and discussion without considering how they differ is a simplification which if avoided may be helpful for research. It is the possibility of a partial dissociation between these two abilities that is the focus of the current research project.

Taking into account the role of the amygdala and other parts of the limbic system in emotional processing in ourselves and others, it is not surprising that tasks claiming to require theory of mind, but failing to distinguish between the cognitive and emotional states of others and often containing some emotional component, report activity in the limbic areas during these tasks. Thus, limbic activity reported during such tasks might indicate that a specific ToM task recruits emotional cognition, rather than indicating that limbic areas are fundamental to all mental state knowledge.

While the false belief test focuses on epistemic mental states (believing and knowing) this is not the case with all ToM tests. The faux pas task requires representing the emotional states of the characters in order to recognise that something awkward has been said and this may or may not depend on the epistemic states of the characters. The affective story task involves both epistemic and emotional states, but the participants are tested on inferring epistemic states about another state, either

emotional or non-emotional such as an epistemic, intentional or desire state. It doesn't require participants to infer an emotional state that is the dependent on subtle non-emotional states such as in the faux pas task – the emotional states are 'simple' (anger, sadness, happiness, etc) and are explicitly stated either in the story or in the behaviour of the characters. The hinting task requires representation of intentions but not emotional states. Furthermore the 'reading the mind in the eyes' test involves no inference of a person's direct epistemic state, but test for recognition of emotional states, both simple (e.g. anger, sadness) and belief dependent (e.g. surprise, regret) (Baron-Cohen, 2005).

2.7 Conclusion

In literature on theory of mind, there is support for the view that not all mental state knowledge is the same. The role of emotional states in some ToM tests affects performance and some of these effects differ between subject groups and there is reason to think that emotional states and non-emotional states don't rely on identical neural substrates.

This raises the question of whether there may be dissociation between emotional state (*ES*) knowledge and non-emotional state (*NES*) knowledge. This dissociation might be show to exist (and begin to be measured) by measuring performance on an instrument that aims to measure independently the ability to have *NES* knowledge on its own and the ability to have both *ES* and *NES* knowledge.

This test might show that not all people are equally skilled at both kinds of mental state knowledge and, potentially, that some subjects that score similarly on a ToM test

that doesn't aim to measure the two skill independently may show non-identical performance on independently measured *NES* and *ES*/*NES* subscales of ToM; that testing the two subscales separately may measure a difference that other tests don't.

Chapter 3: Theoretical Framework

3.1 Theory of Mind Theories

In this section I briefly restate two theories about theory of mind that were described in chapter two; Singer's view about the differences between empathising and mentalising and Baron-Cohen's view that ToM is the cognitive part of empathy (see section 2.2.2 for detail). I argue that despite initially seeming inconsistent they are compatible if emotional and non-emotional mental states are not conflated.

3.1.1 Empathising and Mentalising are non-identical

Singer defines empathising as the ability to share and understand others' emotions and ToM or mentalising as the ability to understand and attribute propositional attitudes such as beliefs and intentions. She argues that the two abilities rely on largely non-overlapping neural structures and that these, and the cognitive abilities they are proposed to underlie, show distinct developmental paths. It is worth noting that a large part of Singer's proposed distinction relies on emotional states and non-emotional states (beliefs and intentions) being different in an important way.

3.1.1.1 Emotional state and Non-emotional state properties

Apart from *ES* and *NES* possibly differing in the neural areas that specialise in the representation of each, as Singer suggests, the two kinds of mental states also seem to have other distinguishing properties. Belief states necessarily have representational content – a representation of something other than itself. Emotional states, if they have any representational content at all, have it in a different sense. As is seen in some ToM tests, beliefs are easily meta-representational; the representational content of a belief can be another belief. Beliefs can thus be 'stacked' and we can have not only 2nd, but 3rd, 4th

and n^{th} order beliefs.[11] Emotions are not as open to meta-representation. It is difficult to imagine what a meta-emotion would be because they don't have obvious representational content. Even though we can be happy *because* someone else is happy, or be happy *about* their happiness, the relationship between the simple state and the meta state is not the same as in non-emotional states and seems more like the initial emotion plays some causal role rather than a representational one. A higher order belief is a different belief altogether from the belief that is meta-represented, while this doesn't seem to be the case for emotions. If emotional states do in some sense have representational content, it differs from that of beliefs at least insofar as the kind of representation in belief is truth apt; it is capable of being true or false. This is just as much the case for a first order belief as a belief of any higher order. Again, if there can be a 'false' emotional state, it is unclear what this would be.[12] There seems to be some theoretical distinction between emotional and some non-emotional states, namely beliefs. However the focus in this, and most, theory of mind discussions is on how people *use* mental state knowledge and the theoretical differences between the states are of secondary importance when compared to empirical data.

3.1.2 Theory of Mind as the cognitive part of empathising

Baron-Cohen views ToM as the cognitive subdivision within ToM which might seem that this is at odds with Singer's view of the two abilities as largely non-overlapping. As mentioned in 2.2.2, what Baron-Cohen means by the cognitive component of empathy is the ability to be aware of the mental and emotional states of others, while he sees the ability to respond appropriately to their emotion as the affective component of empathy.

[11] See the test items in the Imposing Memory Task in Appendix A for an example of this stacking.

[12] 'False' meaning 'not true', as opposed to 'not genuine'

So both Baron-Cohen and Singer see *awareness* of others' non-emotional states as part of a component of the ability in question. Both also see this component as distinct from one which responds appropriately to the emotions of others, or 'shares emotions' (as Singer puts it) with others. Baron-Cohen calls this the affective part of empathy and Singer calls it empathising. The difference is that Baron-Cohen includes *awareness of* (as opposed to *response to*) emotional states in the 'cognitive' component along with non-emotional state awareness, while Singer groups it with emotion state 'sharing' in empathising.

Part of Baron-Cohen's theory (see 2.2.2) is that there is evidence that empathising ability and systemising ability relate in such a way that increased empathising ability generally correlates with decreased systemising ability, and vice versa.

3.2 Systemising and Empathising in *ES* and *NES* knowledge

If it is accepted that *ES* and *NES* have different properties and that *NES* are more open to recursive meta-representation than *ES*, then there is reason to think that *NES* are more open to a systematic explanation than *ES*, since levels of meta-representation are systematically linked. Beliefs are also more obviously open to a systematic explanation, whereby at least some beliefs can be inferred given a simple rule such as 'seeing leads to believing. For example, person B can infer that person A thinks it is raining from the fact that B sees that A sees it raining. Further, person C can infer that B believes A believes it is raining by seeing that B sees that A sees its raining. There is no theoretical limit to this 'belief stacking' and it can be understood as a recursive system that can use any given belief as the content to represent in a different belief.

If *NES* states can more easily be understood in a systematic way than *ES* state, then it seems reasonable that people that are stronger systemisers than empathisers will find it easier to understand *NES* than *ES*. It also seems reasonable to expect that where such states involve a high degree of meta-representation, strong systemisers might be able to more easily understand the meta-representational state because of its possible underlying system.

3.2.1 Singer and Baron-Cohen's views are compatible

If *ES* and *NES* knowledge differ in their 'system-friendliness', then it seems useful to draw a further distinction within Baron-Cohen's view of ToM as the cognitive component of empathy. Specifically, within this cognitive component there is *ES* knowledge and *NES* knowledge.

3.2.1.1 *ES*/*NES* dissociation viewed as Systemising/Empathising

It would then be expected that the proposed inverse relationship between systemising and empathising is not one where high systemising ability is necessarily accompanied by a uniform decrease in all aspects of empathising. An overall difference in empathising ability might be the result of combined differences in ES and NES knowledge ability within empathising. More specifically, someone with high systemising ability would likely be better at representing 'systemising-friendly' mental states rather than those that are less open to systematic understanding.

The possibility of different kinds of mental states being unequally open to systematic understanding, and thus known more or less easily depending on one's

empathising/systemising ability,[13] leaves open the possibility that with reported group differences in empathising lie further differences that current tests are not designed to measure. This is partly because current test don't score tasks that require *ES* and *NES* knowledge separately. ToM research results indicating that test groups who are typically stronger systemisers and weaker empathisers (male groups, science groups) perform worse than groups who are typically weaker systemisers and stronger empathisers (female groups, humanities groups) raise two at least two questions. The first is to what degree the mental states used in the ToM test are 'systemising-friendly'. The second is whether or not the task is open to multiple strategies that don't rely equally on systemising and empathising. These questions are elaborated below; first schematically and then with reference to an existing ToM test, the Imposing Memory Task (IMT).

3.3 The hypothesis

3.3.1 The possibility of multiple cognitive paths

My general hypothesis is that there could be a partial dissociation between representing emotional states and non-emotional states.[14] Given the possibilities argued above that *ES* and *NES* might allow for different levels of systemising explanation and that people vary in their systemising (and empathising) ability it seems possible, even probable that some people are better at *NES* than *ES* representation while for others the opposite is true. If so, it is possible that situations (either real life or in a test setting) in which both *ES* and *NES* knowledge is available the same situation could be represented differently, or different aspects focused on, by different people. The possibility resulting from my hypothesis is

[13] defined as the *overall* ability to represent mental states

[14] As mentioned belief dominate the discussion as examples of non-emotional states, but it applies to other non-emotional states.

that in such cases two,[15] at least partly non-overlapping, cognitive paths may be available; one where understanding and/or responses rely highly on *ES* knowledge, the other where they rely highly *NES*.

The suggested process can be described like this: A person is presented with a situation. From this situation they take some combination of *ES* and *NES* representations, depending on how available these are given the situation and on how easily the person forms each kind of representation. They are then presented with something that prompts them (call it a probe) to respond or report in some way; to give some kind of feedback either to another person in a real life situation or via a procedure in a test setting. They then evaluate their report options, possibly but not necessarily consciously, based on the representations they formed from the situation and pick their report based on these representations. This is illustrated in Figure 3.1 on the next page over.

[15] The number of cognitive paths here is a result of the current focus on ES and NES. It is possible, but outside of the scope of this project, that there are multiple paths that rely on ES and NES representations or other kinds of mental state representations.

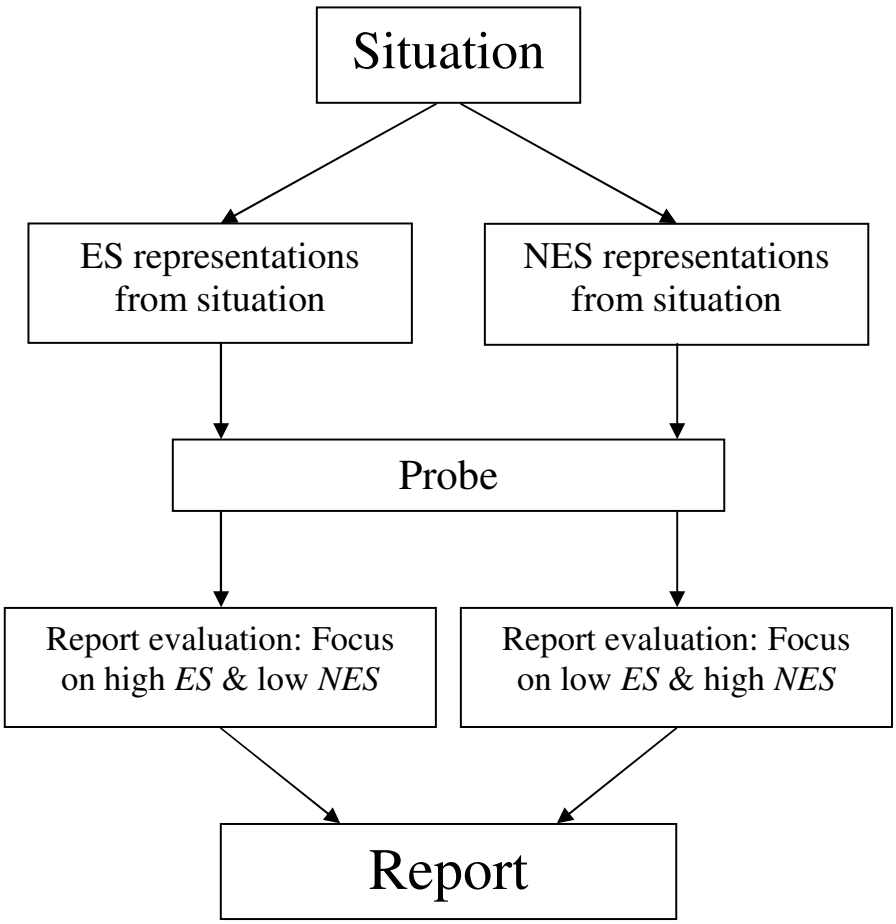

Fig. 3.1: A schematic diagram of two cognitive paths for understanding or responding to a ToM situation; One relies on representing emotional states (*ES*) and the other relies on representing non-emotional states (*NES*)

If this is the case, it could have two possible outcomes. If report options are evaluated in light of the representations formed from the situation, then the chosen report could vary depending on what kind of representations were made. Otherwise, there is the possibility that the same report could be given, despite there being more than one potential cognitive path if the two different kinds of representations, and the report evaluations based on these respectively, lead to the same chosen report. In such a case, if people did use different cognitive paths and give the same report, they would not appear to be using different cognitive paths; the same situation and the same probe led to the same report. In

many simple situations this is likely to make no difference to the behaviour (reports) of normal people who have no particular difficulty with either kind of mental state representation. However, in situation where either cognitive path is at some threshold beyond which the mental representations used in that path become sufficiently challenging, the possibility of multiple cognitive paths influencing behaviour becomes a stronger possibility. In cases where populations are thought to have some kind of representational cognitive deficit, as with Autistics, there is reason to think the possibility of such a threshold being reached is increased. It should also be noted that this suggestion is only one possibility and that it is more likely to be plausible in relatively 'slow' processes and not extremely fast online processes.

3.3.2 Multiple cognitive paths in Theory of Mind tests.

The possible outcomes of different cognitive paths have different consequences for ToM testing. If the situation is such that different reports are given due to different kinds of mental state representations, then a test that treats one report as a correct ToM gives preference to one kind of mental state knowledge over another. If this is a deliberate part of the test design it is not problematic, but if the aim is to test mental state representation in general, it has the potential to undermine the validity as a ToM test by missing an important difference in sub-competencies.

If two different cognitive paths potentially lead to the same report, there is also the potential to undermine some of the assumptions made by some ToM tests. If the test assumes that the task is certain difficulty because of the complexity of forming one kind of mental state representation, there is the possibility that a cognitive path relying on mental state representations of a different kind might not be equally difficult. People who give the same report could then have arrived at it with greater or lesser difficulty,

depending on how easily they form specific kinds of mental state representations and how demanding doing so is in the task in question. It is worth noting that in both in general, and in the specific example discussed in section 3.3.2.1, it is probable that most normal people probably rely on some mix of both kinds of mental representation rather than only on one of the possible cognitive paths. Below I discuss an example of this possibility from a prominent ToM test; the Imposing Memory Task (IMT).

3.3.2.1 Multiple cognitive paths in the Imposing Memory Task

In the IMT (see appendix A) some questions are structured in a way that if a story characters' emotional state is inferred, subjects can use this emotional state knowledge to correctly answer the question, rather than the stacked intentional and belief states that the test focuses on (See Fig. 3.2). For instance, in the story "Where's the post office?" Henry tells Sam where to find a post office but it turns out the post office is somewhere else. Question 4[16] gives a choice of two claims; in one Sam thinks Henry knows where the post office is (and Henry's knowledge is correct); in the other Sam thinks Henry knows it was in the location he told Sam which turned out to be wrong. But if Sam thought that Henry knew where the post office was (option *a*), yet Henry told Sam an incorrect location, this is likely to be seen (by the subject) as Sam thinking he's been deliberately deceived by Henry. More importantly, this is likely to lead the participant to infer a consequent emotional state in Sam (such as anger or disappointment). The important thing is that only the option in which Henry did not intentionally deceive Sam in the question (option b) would lead to this emotional state, so the participant could pick the answer that best fits the characters' emotional state inferred initially inferred from the story, rather keep track of the levels of meta-representation of intentions and beliefs. In

[16] Question numbers are specific to the story and in my discussion always refer to the story being discussed.

this case, the correct answer (that Sam thought Henry knew the location, but in fact Henry's belief was false) doesn't lead to Sam being unhappy (or whatever emotion Sam is inferred to have from being deceived). So, if the character's emotional state was inferred when the story was initially read then the correct answer can be arrived at without remembering the meta-levels of intentional and belief states. Answering this question by relying mostly on beliefs requires a participant to represent Sam's belief about Henry's belief about the location of the post office, whereas answering it by relying mostly on emotional states allows a 'shortcut' - representing Sam's emotional state that he wasn't unhappy with Henry, and then inferring that the correct report must be the option where Sam is not unhappy with Henry; statement *b*.

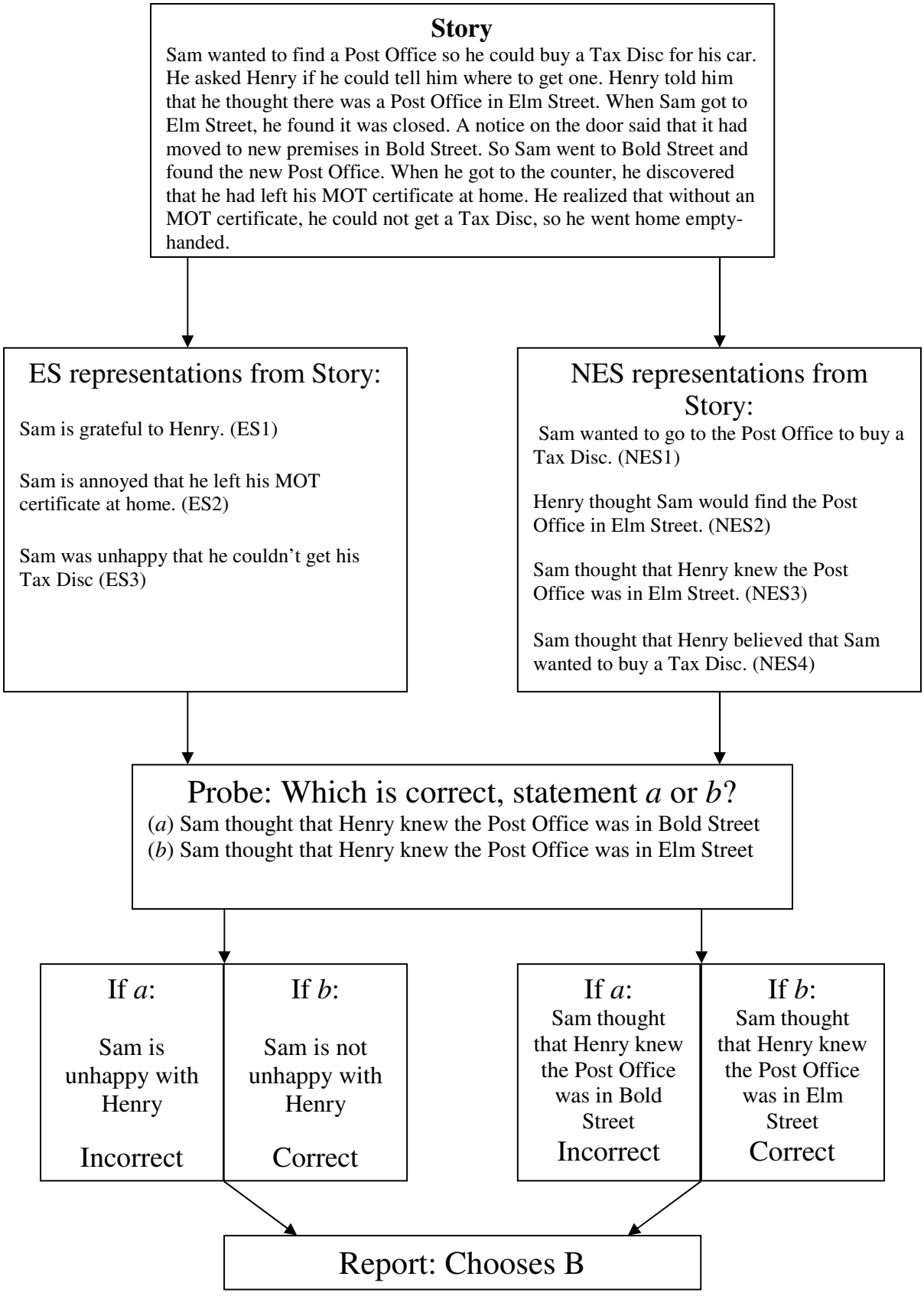

Fig. 3.2: A diagram showing how two cognitive paths may be used to answer a question from the Imposing.Memory Task. (Question 4 from the story "Where's the Post Office")[17]

[17] Items in the box '*NES* representations from story' are taken directly from, and are the correct answers to, ToM questions for the story "Where's the post office" from the Imposing Memory Task – See Appendix A.

There are several examples of emotional representations possibly making alternatives cognitive paths available in the IMT. In "John's problem" question 4 gives a choice of two claims. In one he thinks Sheila would like to go for a drink with him, in the other he thinks she would not like to. Given that he 'fancies' Sheila, the participant is likely to represent his emotional state in different ways depending on which one is correct. If John's emotional state was inferred from the story, only one of these will be coherent. The participant's representation of this emotion can be used as an alternative to their representation of John's belief about Sheila's hypothetical preference when deciding on an answer. The same applies in question 10 where the statements involve a 5^{th} order mental representation (a belief about a belief about a desire about a belief about a desire), but differ concerning John's belief about Penny's belief; one states she believes John wants to go for a drink with Penny and Sheila, the other states John wants to go for a drink with Sheila alone. In this case, the participant might have inferred John's or Penny's emotional state or both when the story was read. Either one of these can be clue to the answer, given that John's emotional state might differ depending on whether there is a chance Sheila, who he likes, and him might go for a drink alone (or he thinks there's such a chance). His emotional state might also differ according to whether or not he thinks Penny might feel used if he wants to go for a drink alone. Again, what is important is not that this also involves mentalising, but that the initial mentalising doesn't need to be remembered or used to answer the question, but that the represented emotional state (maybe with some new on-line mentalising) can be remembered or represented instead, and that this will only fit with one of the options given if they allow the subject to infer different emotional. This is even more likely in possible individuals who are bad at representing intentional or belief states, or any high order states, but are good at representing and inferring emotional states.

In “Simon” question 4, the two alternatives differ about Simon’s belief about Jim’s belief about whether or not Susan would or would not marry Jim. However, this difference also means that Jim’s emotional state will differ, depending on which of the two scenarios described by the two statements is the case. Additionally, given the difference in Simon’s belief over the two options, he would also have different emotional states if different options were true since he cares about Simon. Thus the difference between the two options can be seen as resulting from either’s representation of the other’s emotional state, which also differ according to which statement is taken to be true. Again, the participant can use a match between the actual or represented emotional state of the characters in the question and the story as a substitute for representing their belief and intentional states from when the story was read. It is worth noting that no information about Susan’s actual intention needs to be remembered from the story by a participant basing his/her choice on which one matches the emotional states from the story, reducing the complexity of the question. Question 6 is similar insofar as both Jim and Susan’s emotional states are likely to be (and be represented as) different depending on which answer is correct. There are then two possible emotional representations that can each (or both) act as a representational substitute for representing the complete sequence of stacked mental states. Question 8 is the same as question 6 apart from one more layer of stacking and the same problems apply if participants infer the emotional state of the characters. If both question 6 and 8 are answered based on emotional state inferences then the increase in difficulty (due to the increased stacking) falls away.

In “Emma’s Dilemma” the alternatives available as answers to question 4 differ on whether or not Jenny thought Emma’s boss would believe the lie she advised Emma to

tell him. Though the two scenarios may lead participants to have different emotional towards Jenny, there is no emotional state of any *character* that can be inferred by the participant to be used instead of inferring their belief states. Though the subject's emotional state towards the Jenny can lead the subject to the answer that fit the emotional state they had while reading the story, this is different from inferring someone else's emotional state as a substitute for inferring their belie/intentional state(s). In question 6, the claims differ on whether or not Emma thinks the boss knows she lied. This difference is likely to also make a difference to her emotional state (e.g. being anxious), so in this question an inference about her emotional state can help a subject in choosing an answer.

The points noted above do not cover all possible aspects of *ES* and *NES* representation. In some cases *ES* knowledge might be dependent on *NES* knowledge and so an *ES* cognitive path may not (and in most cases probably isn't) wholly dissociable from NES knowledge. However, at least part of what is being tested in the IMT is the recall of the states. It is possible then that those who are less good at *NES* representing can use *NES* to arrive at ES during the initial presentation of the situation but then be less good at 'online' representing of *NES* when the report options are being evaluated. In this case being good at *ES* representation, even if initially relying on *NES*, could be used beneficial. My aim is not to say that there is an ability that is at work that is completely separate from the one being tested. I aim to show that, given that belief states and emotional states differ conceptually and neuro-anatomically, and that the current ToM tests clearly differ in how much of each they require subjects to represent, without indicating awareness of this difference or at least not implementing such an awareness in the way which I argue would be useful. This means that the current tests are open to at least the possibility of participants who are better at representing either emotional or non-emotional (intention,

belief, etc) states scoring at a level on tests that aim to measure one of these abilities, on the basis of their being good at a skill that isn't being tested, but can be used in answering the questions. The most obvious example would be the IMT, that aims to test the ability to represent or remember stacked intentional and belief states, but because it is blind to the dissociation argued for, is open to participants who are not very good at remembering stacked non-emotional states but good at remembering emotional states getting questions right that involve a high degree of stacking, on the basis of their emotional content or some mix of this and the stacked states.

3.3.1 *ES*/*NES* distinction and Theory of Mind test scores.

Although the IMT is just one of many ToM tests, most ToM tests don't draw a distinction between *ES* and *NES*, and when they do indicate awareness of each kind of mental state, they don't show evidence of deliberately including either or both in the test. It is the possibility that people aren't equally good at representing the two kinds of mental states that is seemingly absent from all existing ToM tests. This is not necessarily a problem if the aim is to test overall ToM and not subcategories of it. However, it is potentially problematic if the possibility of a specific kind of mental state representation is assumed when there is the possibility of performing a task in a ToM test relying on a kind if mental state representation other than the one assumed. More specifically, this could affect scores by not giving a comprehensive score that reflects overall ToM ability where the score *is* assumed to show this. Alternatively there is the possibility that a score is taken to be an indicator of a specific kind of mental state representation (such as stacking of higher order mental states in the IMT) when the same score could reflect a different aptitude for a different kind of mental state representation (such as keeping track of characters' emotional states). So recognising that there may be variability in people's

ability to represent different kinds of mental states has the potential to somewhat improve certain ToM tests or encourage a more explicit stance in ToM test design on this issue.

3.4 The proposed research

Apart from the effect the hypothesis may have on existing test, there is the possibility of a testing instrument specifically designed to measure both *ES* and *NES* ability. Such an instrument will be doubly beneficial; if it shows that people are not equally good at representing *ES* and *NES* then this is both a ToM test that can measure differences not previously measured, and it gives a starting point for identifying which existing ToM tests are, and which are not, undermined by this difference in ability. A large part of the current project is an attempt to construct an instrument that aims to measure this difference. The design of an instrument that aims to do this is discussed in chapter 4.

Chapter 4 – Research Design

4.1 Analysis of Theory of Mind instrument properties

Chapter 2 focuses mostly on the results of theory of mind tests as support for the proposal that it isn't a single, unitary ability. This chapter discusses the motivation for and construction of a research instrument that is designed as a step towards testing the hypothesis. To orientate this research project within existing ToM research, some prominent ToM research instruments are reviewed, with the focus on their design structure and scoring. Where a test is sufficiently described in the literature review (Chapter 2), a brief reminder and cross reference are given.

A ‘pure’ test for the ability to represent beliefs, without including emotional states, creates a situation where a character has a false belief and tests subjects’ knowledge of the false belief. The False Belief Task (Wimmer and Perner, 1983) and Appearance-Reality Task (Perner et al, 1987) are examples of this. In both tests, a subject is asked a question which they will answer one way if they represent the false belief, and another way if they don’t. The report options open to subjects are dichotomous and the test is either passed or failed with little interpretation required by researchers. ‘Pure’ tests for the ability to represent emotional states without relying on belief states have been achieved by showing subjects photographs of faces bearing expressions of basic emotional states and asking them to identify or rate the emotional state (Shaw et al, 2007) or by neural imaging (Wicker et al, 2003) while viewing the photographs. Though results from these different kinds of mental state tests can be compared, it is difficult to conclude whether differences are the result of differences in ability or of non-mental state factors.

Some theory of mind tests, such as the Affective Story Task (Schenkel et al, 2008), Faux Pas Task (Baron-Cohen et al, 1998; Gregory et al, 2002) and Strange Stories Task (Happe, 1994) include both emotional and non emotional states. However, none of these give a separate score for items requiring or answers indicating emotional states and non-emotional state representation. To test dissociation between these two abilities, an instrument is required that allows both *ES* and *NES* representation, explicitly requires subjects to represent both, and score separately the items that test *ES* and *NES* representation. To construct an instrument capable of doing this, the design features that are needed and are to be avoided from current ToM tests are considered. The table at the end of this section summarises the features of these tests important for the current research. This aids design decisions, both to identify the design features that will help to test the hypothesis (see 4.4.2) and to know what design features should be avoided.

Test	False Belief Task	Appearance-Reality Task	Hinting Task	Imposing Memory Task	Affective Story Task
Origin	Wimmer & Perner, 1983	Perner et al, 1987. (Gopnik & Astington, 1988* applied it to own belief)	Concoran et al, 1995	Kinderman et al, 1998	Schenkel et al, 2008
Used In	Wimmer & Perner, 1983; Baron-Cohen & Frith, 1985; Clements & Perner, 2001	Gopnik & Astington, 1988	Concoran et al, 1995; Schenkel et al, 2008	Kinderman et al 1998, Stiller & Dunbar, 2007* - Expanded version	Schenkel et al, 2008
Presentation Method	Acted scene - Protagonist stores an object. It is then moved without protagonist's knowledge; Subject is asked where protagonist would look for the object	Physical objects presented & deceptive appearance revealed by showing real property	Stories read aloud to subjects	Shown on overhead projector and simultaneously read - Presentation made to all subjects together	3 stories read to subject (positive, neutral and negative affect respectively) each followed by a story relevant false belief question
Testing Method	False belief question asked	Gopnik & Astington, 1988: Question asked to test knowledge of one's own previous false belief	Asked what the hinting character really meant. Further hint in response if incorrect	Subjects are given a questions item booklet - 2 claims per item, Forced choice	False belief question asked for each of 3 stories
Report Method	Verbal answer (gaze, see section 2.3.1 Clements & Perner, 2001)	Verbal answer	Verbal answer	One of 2 options indicated for item - answered in booklet	Verbal answer
Report Options	Pragmatically dichotomous	*Dichotomous: Subject asked: What did you think previously, X or Y?	open - correctness interpreted but based on story	Dichotomous	Open - correctness interpreted by researcher based on story
Scoring Method	Report correct / incorrect	Report correct / incorrect	Report correct/incorrect: Correct = 2; Correct after extra hint = 1; Incorrect = 0	Total calculated from Correct/incorrect reports in 2 ways: 1) Meta-Level of first error 2) Overall correct -incorrect measure using negative marking	Reports correct / incorrect
Scoring Scale	Pass / Fail	Pass / Fail	Max = 20	1) Max = 9, Min = 1 2) Max = 26, Min = -26	Pass / Fail for each valence: Max = 3
Mental States Tested	NES	NES	NES	NES	NES
Salient Mental States	NES	NES	Both ES & NES	Both ES & NES	Both ES & NES

Table 1 A

Test	Strange Stories Test	Faux Pas Task	Social Stories Questionnaire	Reading the Mind in the Eyes	Recognising Facial Expression
Origin	Happe, 1994	Stone & Baron-Cohen in: Stone et al, 1998 and Gregory et al, 2002	Lawson et al, 2004	Baron-Cohen et al, 1997 (Improved version in Baron-Cohen et al, 2001)	N/A
Used In	Happe, 1994; Shaw et al, 2007; Jolliffe & Baron-Cohen, 1999	Stone et al, 1998; Gregory et al, 2002; Stone et al, 2003; Shaw et al, 2007	Lawson et al, 2004	Baron-Cohen et al 1997; Baron-Cohen et al 2001; Stone et al 2003	Wicker et al, 2003; Mitchell et al, 2005; Shaw et al, 2007*
Presentation Method	Stories read by experimenter or subject. Stories are either 'strange' (require ToM) or 'physical' (don't require ToM)	Stories read out to each subject	Booklet containing stories - 3 parts to each story, after each part is a printed question: Was anything said that could have upset anyone?	Photo of eye region shown with mental state terms printed adjacent	Photos of 4 faces - Photo of each face showing: anger, sadness, disgust, happiness, surprise and fear. Emotional state terms printed adjacent
Testing Method	2 Questions: Comprehension question (is what x said true?); Justification question (why did x say that?)	Series of questions asked (see 2.4 for full description)	Test Booklet: Contains stories & report options	Forced choice between mental state terms - Which word in a set best describes what person is thinking or feeling	Subjects rate the intensity of a presented emotional term for a presented face
Report Method	Verbal answers	Verbal answer	Answers filled in on booklet	Verbal report selected mental state	Rating (1-10, presumably verbal) of how much emotional term applies to each facial expression
Report Options	1) dichotomous 2) open	Open - correctness interpreted by researcher based on story	Dichotomous (Y/N) If 'Yes', proceeds to MCQ	Forced choice: Original - 2 options Improved - 4 options	1 to 10
Scoring Method	Justification Question answer rated as 1) Correct or Incorrect (Correct scores 1) 2) Mental or Physical State	Report correct / incorrect	Correct positives = 1 (n = 20), Correct negative = 2 (n = 10)	Report correct / incorrect	Ratings summed for 'congruent' (term matches expression) and 'incongruent' (term doesn't match expression) trials separately
Scoring Scale	Physical Stories: Max = 6 Strange Stories: 1) Comprehension: Max = 24 correct 2) Justification: Max = 24 Mental, 0 Physical Min = 0 Mental, 24 Physical	Max = 60	Max = 40	Max = number of items shown: Baron-Cohen et al, 1997 Max = 25; Baron-Cohen, 2003 Max = 36	Discrimination index: mean difference between incongruent & congruent ratings Calculated for each emotion Max = 9
Mental States Tested	Both ES & NES	Both ES & NES	ES	Both ES & NES	ES
Salient Mental States	Both ES & NES	Both ES & NES	Both ES & NES	Both ES & NES	ES

Table 1 B

4.2 Research question

My research question is, “Is there a (possibly partial) dissociation between having knowledge of emotional states (*ES*) and having knowledge of non-emotional states (*NES*)?” The question is informed by the broad understanding of theory of mind - that it includes the ability to represent all mental states of others. The aim is to contribute to the understanding of ToM by suggesting that it encompasses two separate non-identical but possibly partly overlapping abilities. The details of how this research aims to answer a more specific version of this research question more specific are discussed in section 4.5.1.

4.3 Research Rationale

This section explains why it is worthwhile to do the research. The possible implications that answering the research question might have on more general cognitive theories, such as the inverse relationship between systemising/empathising, and on related research areas, most notably Autism studies, is discussed. Support is also given for why the current project can answer the research question.

4.3.1 General Rationale

An intuitive way in which the research question can be answered is to test how well people perform two tasks, one testing *ES* knowledge competence and the other *NES* knowledge competence, then comparing performance on these. One problem with this is that it would be difficult to tell if the tasks are similarly taxing on the respective skills they test and extraneous variables like working memory, attention and language ability. Therefore it would be unclear whether a performance difference is the result of differing competencies in two abilities or a result of one task being more difficult

than the other in a way that isn't directly related to mental state knowledge. This undermines the goal – to show dissociation by showing a performance difference.

This problem can be avoided by using instruments that keep other variables including difficulty as constant as possible across testing of the two abilities. The current research aims to create a single instrument where one of two distinct subscales tests each ability while keeping these variables as constant as possible. This is attempted by dividing the mental states often represented in ToM tasks into two categories - 'emotional' and 'non-emotional' - then testing the relative contribution of the knowledge of each to the overall performance. If the contributions are not equal, the research supports the view that *ES* and *NES* representation are at least partly dissociated.

If the answer to the research question is that there is some dissociation between different kinds of mental state knowledge, then this is an insight into mental state representation in general. Further research can use this as a basis for trying to find the neural correlates of this cognitive dissociation. If a person or population is characterised by an abnormal functioning of mental state knowledge, we gain further insight by researching whether this abnormality is specific to only one kind of mental state or equal across them all.

The research question is partly informed by the view that systemising and empathising ability are generally in an inverse relationship (as discussed in section 2.2.2). Some dissociation between *ES* and *NES* representation might be expected if there is generally such an inverse relationship because at least some *NES* (most

obviously beliefs) seem more open to an obvious systematic explanation (one based on input-output rules) than emotional states. For example if I see my friend seeing an event, I believe that he thinks it happened. Furthermore, if a third person also sees the event, and sees me seeing him see the event she will think that I think that he thinks it happened. In this way others' mental states can follow from a simple rule, something like ‘seeing leads to believing’. However, there isn't as obvious a systematic explanation of emotional states. It is less obvious how emotions work in such situations. My belief that my friend is sad doesn't seem to follow from as obvious a rule as 'seeing leads to feeling', since it would depend on many other factors, such as his desires or feelings about what he has seen. It seems at that if there is a rule hat could result in my belief about his emotions, it would be significantly more complex than the 'seeing leads to knowing'. So if dissociation is found, it fits with this more established theory about the relationship between empathising and systemising – possibly with belief knowledge accounting for less of the difference in empathising and mentalising skill and emotion knowledge accounting for more.

Establishing dissociation would also give good reason to consider how much systemising and empathising respectively can be relied on by subjects when doing existing ToM tests. It is possible that some ToM questions are passed by strong systemisers because they are more open to systematic explanation, and therefore easier than others that are less open to such an explanation. Given their relative weakness at empathising, strong systemisers may find an equivalent to the ‘emotional shortcut’ that might be used in the IMT (see section 3.1.2.1). A complex systemising explanation might not seem like a shortcut, but the stronger a systemiser and the weaker an empathiser a person is, the easier they are likely to find a ‘systemising

alternative' to a task that would normally (by those with normal empathising ability) be done by empathising. This idea might be useful to study the extent to which tests that measure empathising ability rely on ToM and how prevalent 'systemising-friendly' *NES* states are in the ToM tests used, thus using the dissociation to help identify and minimise potentially unaccounted for variables in the relevant; specifically the extent to which potentially systemising-friendly mental states affect scores on tests that aim to measure empathising ability.

Other, and possibly the most important, reason why answering the research question could be of consequence is the implications of a dissociation would have on areas where ToM is a core ability; most obviously on the view that Autism is characterised by a deficit in ToM. If my research indicates a dissociation, this could be used research whether the ToM deficit in Autism is biased towards specific types of mental state representation, i.e. if it indicates difficulty representing all mental states or if it is a difficulty more specific to *ES* or *NES* or if the subtypes of Autism spectrum disorders show different deficits in this regard. One way this could be tested would be to assess existing ToM instrument tasks for the type of mental state knowledge they require. Data gathered using these might then be reinterpreted as not just showing mental state representation in general for the tested population (as ToM is often currently viewed), but as more *ES*/*NES* specific depending on the degree of *ES* and *NES* found to be required when assessing the instrument.

4.4 Instrument design

4.4.1 Detailed Rationale

As mentioned in 4.2, the research question may be answered by testing subjects' ability to 1) represent *ES* and *NES* together and 2) represent *NES* without *ES*. I refer to these as the two proposed subscales to be tested and for convenience I call 1) the *ES* subscale (since it also contain *ES*) and I call 2) the *NES* subscale (it contains only *NES*). There is no subscale in which only *ES* states are tested without *NES*. There are two main reasons for this. The first is that it is very difficult do construct stories that are complex enough that they are likely to lead to some variation in subject feedback, without including information that makes *NES* inferences likely. It is far easier to describe a situation where characters have no obviously salient emotional state than one where they have no obviously salient belief or intentional state. Furthermore, as discussed in section 3.1.1.1, *ES* are not open to meta-representation in the same way as *NES*, so it is less simple to test meta-*ES* than meta-*NES.* It may be that the lack of a subscale measuring *ES* without *NES* means that the subscales are not as separate as they can be. However, given that this research is in its primary stages it seems reasonable to leave this possibility open for investigation in future research. In this section I give details of how the research question can be made more specific and how the instrument constructed for this project is designed to test the two subscales. Further research design details of and reasons for specific features of the instrument are discussed in sections 4.5 and 4.6.

There are a number of reasons for constructing an instrument that takes what may be two dissociable abilities and constructing an instrument that test these by using subscales rather that testing each ability individually. One reason has been discussed

in section 4.1; that it would be difficult to ensure that the non-test variables like those mentioned in 4.3.1 (such as working memory load, language requirements, scoring method, etc.) are similar enough between the two tests to be confident they won't unduly influence the results of the test. This can be achieved by having near identical test items, and changing only the variables that define the subscale, and leaving such variables as language, memory and attentional demands the same across subscales. The instrument also tests the two skills in an 'integrated' way; subscale test items are not presented as two consecutive sets, but presented in a mixed order. The overall skills tested are tested simultaneously overall, decreasing the chance that fatigue will have an unequal effect on each subscale.

Using two subscales on the same instrument rather than two separate tests may also reduce the risk of potential priming effects from subjects repeatedly doing tasks specifically requiring them, for example, to represent emotional states. Especially if there is reason to believe a population will find a given task more difficult than another population, this effect may diminish if they show improvement through subsequent repeated tasks requiring the same skill. This could skew performance that aims to measure their ability to represent such states in everyday situations. Similarly, subjects are more likely to work out what the instrument aims to measure if it repeatedly and only measures one thing. It is also easier to ensure that the complexity of the two subscales is similar between the two subscales than if two established, but unrelated, tests were used to test the two abilities.

Written stories are used to present the subjects with a situation involving people likely to have certain mental states. This medium is used because it has been used extensively in ToM research (See Table 1 A & B) and enables administering tests to a large sample quickly. It is also enables describing a social situation, as well as control the information presented to the subjects to minimise interference by uncontrolled variables. For example when describing a hypothetical situation in which a character has a certain emotion, the subjects' ability to read facial expressions of emotions won't be a variable – only the features included in the story are available to be taken from the stimulus and these can be adjusted to change the parameters for the mental state to be used in testing, e.g. the level of meta-representation, the kind of mental state, how obvious the inference is to make etc. By intentionally limiting the information in the stimulus variance in performance is more likely to be due to variance in subjects' competence in the abilities being tested.

Science and Humanities major university students are used as sample groups because the established group differences in ToM ability between these groups (discussed in 2.2.2) makes them suitable for testing the current hypothesis. Given that there is a difference in ToM ability, it would be interesting to know if this difference is equal for *ES* and *NES* knowledge within ToM. A potential dissociation between the two abilities (*ES* and *NES* representation) may be visible by testing the inter- and intra-group subscale scores and will support the hypothesis that there is some dissociation between the two (details in Chapter 5). These groups are also less likely than other populations, given their education level, to have difficulty with the cognitive demands of the task that don't form part of the study, such as working memory and language.

4.4.2 Identifying desirable design features of existing instrument.

In this section some design features of existing ToM tests that might be useful for answering the current research question are discussed (see 4.1). I also discuss the features likely to interfere with answering the research question and ways to make sure the instrument is free of them. These are discussed using the design variables from Table 1 A & B as headings.

Presentation Method

Many existing tests – specifically the Faux Pas Task (FPT), the Imposing Memory Task (IMT), the Social Stories Questionnaire (SSQ) and the affective story task (AST), The Strange Stories Test (SST), Hinting Task – use a textual presentation (either read by a researcher or read by the subject) of a social situation followed by questions about the story as a means of testing ToM. This is useful for controlling precisely what information is included in the situation presented. The IMT, SST and Hinting Task use this primarily to present situations in which characters are likely to have salient beliefs. The FPT, SSQ and AST use this format to present situations that are emotionally loaded and in which characters are likely to have salient emotional states. This method of presentation also allows variable ease of availability of the salient mental states by including more or less information relevant to a given mental state. This means a test can be designed to be more or less taxing on the kind of mental representation ability to be tested. Text based presentation where subjects read stories has an advantage that tests can be fairly easily administered to large numbers of literate subjects.

Testing method

In Shaw et al's (2003) version of the task to recognise facial expression (RFE), subjects rate the intensity with which a given face exhibits a given (from an emotional term). This allows a variety of responses from subjects and therefore helps improve variance in data and allows subtle differences in subject response. This is useful for the current research where there is a concern of a ceiling effect.

Report method

The majority of the test reviewed – the False Belief Task (FBT), Appearance-Reality Task (ART), AST, SST, FPT, RME, RFE – let subjects report using verbal responses. While this is unproblematic when subjects are tested individually, it is problematic when testing multiple subjects simultaneously. The IMT and SSQ use non-verbal report, which means multiple subjects can be tested in the same venue simultaneously without concerns of collusion or influence from others' responses during testing. This report method seems most suitable to the current research.

Report options

A number of the ToM tests discussed use dichotomous report options. The benefit of this, compared with open reports (subjects being allowed to answer questions in their own words), is that there is no need for researchers rating responses (for correctness or containing or not certain markers) and they are clearly correct or incorrect. The drawback is that there is decreased flexibility and variability in reports.

Scoring method and Scoring Scale

The FBT, ART and AST are the only ToM tests reviewed that scored reports as pass / fail and not yield a discrete score. This is useful in some settings, such as investigating proposed developmental changes, but give that differences in ability are likely to be subtle it is preferable to use a discrete scoring method.

The RFE scores subjects on congruent and incongruent trials. A discrimination index is then calculated using the difference between these two scores. Given the proposed dissociation, similarly coding items according to subscales to enable comparisons of different scores seems suited to the current research.

Mental states tested and salient mental states

Though the hinting task, AST, SST, FPT, SSQ and RME all include both salient emotional and salient non-emotional states and the SST, FPT and RME either allow or require both kinds of mental states to be represented in testing, none of the tests do this in a way sufficient to test the hypothesis. For dissociation between ES and NES to be researched, both kinds of mental states must be required in different test items, both items must be sufficiently similar to avoided non-test variable interfering with performance.

The emotionally loaded stories from the FPT, SSQ and AST are a design feature useful for parenting emotional states in the instrument. The design includes being selective in inclusion of emotional states, and different versions of stories in the current instrument were constructed, respectively including and excluding emotional states. Further details are in the next section (4.4.3).

4.4.3 Instrument summary

The instrument consists of 20 stories, a claim set for each story, and a feedback and scoring system. The stories all describe scenes of social interaction. The 20 stories comprise of 2 versions of each of 10 stories, with only 1 version of each story used for any given test presented to a subject. The claims in the claim set are relevant to the story and the same claim sets are used for *ES* and *NES* versions. The feedback system is in the form of a 5 point Likert scale. Subjects rate each claim in the claim set of a given story for 'correctness'. The scoring system is then used to change this feedback into a score. Further details about scoring are in section 4.5.3.

There is 1 version (either *ES* or *NES*) of each story in each test – making 10 stories per test run, and no individual test contains both the *ES* and *NES* versions of the same story. The stories are to be shown to subjects on computer screens until the subject responds by pressing a key to indicate they have finished reading it. After this the story disappears and subjects can't go back to view the story again. The claims then appear one at a time. When a claim appears on the computer screen, it appears under the heading (which is present for all claims in the claim set); '*How correct is it to say that...?*' Subjects respond by rating each claim on a 5 point Likert scale with the points labelled in sequence as '1 - correct', '2 - near correct', '3 - 'neither correct nor incorrect' (henceforth 'neither'), '4 - near incorrect' and '5 - incorrect.' After a response is entered there is a short pause, after which the next statement to be rated appears. As with stories, subjects can't go back to a claim once a response has been entered. This process repeats until all the claims in the claim set have been rated after which the next story is shown followed by the corresponding claim set. This process repeats until subjects have worked through one version (*ES* or *NES*) of each of the 10

distinct stories and their corresponding question sets. Below is an example of one of the stories:

Jill had recently moved into a new flat. She thought the outdated curtains and lampshades that were in the bedroom when she moved in were ugly.

She went shopping and bought some new curtains for her bedroom which she thought were beautiful. When she had put them up, her best friend, Lisa, came over. Jill gave her a tour of the apartment and asked, "How do you like my place?"

"It's lovely! And I know of a place where you can get some beautiful curtains for your bedroom," Said Lisa.
"Oh, okay. By the way, I bought the curtains yesterday from the mall," said Jill.

'Ugly Curtains' (ES version) - story from Instrument

4.5. Schematic Instrument details

4.5.1 Stories

As mentioned, there are two versions of each of 10 stories – an *ES* and a *NES* version – totalling 20 stories. All stories describe everyday social situations where the characters have some set of mental states (such as believing, knowing, feeling, wanting, etc) and these are either known or unknown to other characters in the story. The purpose of having two versions of a mostly similar story is to create a variable by having emotional states either present or absent, while changing other features of the story as little as possible. Most importantly, the *NES* are mostly similar in content and

number so that aside from the difference in *ES*, the two stories are close to identical as ToM stories and their claims sets close to identical as ToM tests. In *ES* versions, the situation described by the story is one that will lead to one or more salient emotional states in the character/s in addition to non-emotional states. For example in 'Ugly curtains' (See 4.4.3 above), it is reasonable to think that Jill will be unhappy that Lisa thinks her curtains are ugly if her remark is interpreted as a judgement of the curtains. In the *NES* version, the situation described is one that doesn't lead to any salient emotional state in any of the characters, yet they still have *NES* that can be inferred by other characters. In the *NES* version of 'Ugly curtains' Jill had not recently bought new curtains, so is not upset that Lisa thinks the curtains are ugly, yet all the other belief states, such as Lisa's belief that the curtains are ugly, are still there.

The purpose is to test ToM ability when *ES* are included v ToM ability when *ES* are not included to see if this influences performance. Group and individual performance on *ES* and *NES* stories can be compared helping answer the question of whether or not there is a group difference in ToM tests that include/exclude *ES*.

4.5.2. Claim sets

Below is an example of a claim set. This is the specific claim set for both the *ES* and *NES* versions of 'Ugly Curtains' (see 4.4.3). Below each claim is the response taken to be correct (see 4.6.5 for support). The claims were presented to subjects in random order.

a.) Jill feels as happy at the end as she did when Lisa arrived

Incorrect

b.) When she first arrived, Lisa thought Jill recently bought the curtains in the bedroom.

Incorrect

c.) At the end Lisa thinks that Jill is just as happy as when Lisa first arrived.

Incorrect

d.) At the end, Lisa thought that Jill thought Lisa liked the curtains.

Incorrect

e.) Jill went shopping before Lisa came to visit her flat.

Correct

Claim set for 'Ugly Curtains'

Of the 5 claims in each claim set, 4 are about the mental states of characters (*a* - *d*), and one is a control question about a non mental state *(e).* Of the four test claims (not control claims) two test the subject's representation of either an *ES* or *NES* state directly; one is a claim about an *ES (a),* one about a *NES (b)*. The other two claims have an extra level of meta-representation and are claims about a belief about either an *ES* or *NES;* one claim is about a *NES* about an *ES (c),* one about a *NES* about a *NES (d).* All claim sets follow this schema and the set of claims to be rated will be the same for both the *ES* and *NES* versions of all stories. This is to make sure that claim

sets for *ES* and *NES* versions don't systematically vary in a way that is unaccounted for and that could affect the responses to these claims.

Apart from the *ES* and *NES* test claims, there is one control claim in each claim set. This is a claim to be rated for accuracy which doesn't have reference to any mental state of story characters, but is about some non mental fact mentioned in the story. Control claims test the subjects' comprehension of the stories, and the score from control questions on all 10 stories results in a 'control score'. This is not the focus of the instrument, but it could be useful in comparing comprehension with overall or subscale specific performance, or to set a threshold for comprehension. If subjects score below this threshold, it might indicate that their comprehension of the stories was insufficient for their *ES* and *NES* scores to accurately represent their ability to represent these mental states and that some other factor, for example language or memory demands, had a major influence on their results.

The question subjects are asked when claims are displayed is; "How correct is this statement?" and the intervals on the Likert items are labelled as varying degrees of correctness. The reason for using correctness as a rating variable is that it does not require the subject to attend to their own mental state. Attending to or considering their own mental state would be asked if they were asked *"How much do you agree with this statement?"* instead, and were then given a Likert item with intervals varying in level of agreement. Phrasing the question and response options in a way that avoids asking subjects to attend to their mental state about the claim to be rated makes it less likely that this potential extra level of meta-representation (their belief

about what the claim says) might have a different effect on *ES*, *NES* and control claims.

4.5.3 Scoring

Responses are coded as score by assigning numerical scores to subjects' Likert responses. The maximum score is allocated to the response that most correctly[18] rates the claim in question in the context of the preceding story. For example in response to claim *d* (in 4.5.2) in the context of 'Ugly Curtains *ES* version' (see 4.4.3), 'incorrect' is taken as most correct since at the end Lisa did not think that Jill thought that Lisa liked the curtains. The minimum score is allocated to the response that least correctly rates the relevant claim. In the example above this is 'correct'. For responses between these, proportional scores are allocated as detailed below.

For question items where one of the extreme ends of the Likert scale is the correct response, the most correct response is given a score of 4. This is done to make it easy to give descending scores to incorrect responses the further they are from the most correct one. For instance if rating a claim as 'correct' is awarded a score of 4, a 'near correct' rating scores 3, 'neither' scores 2, 'near incorrect' scores 1 and 'incorrect' scores the minimum of 0.

The scores for control claims are recorded separately and comprise the 'comprehension score'. For test claims the score is allocated to either the *ES* or the *NES* subscale which is discussed in more detail in 4.6.4 below.

[18] See 3.4.2.5 for details

4.5.4 Subscales

The *ES* subscale score is a result of answers to *ES* story questions and the *NES* subscale score is a result of answers to *NES* story questions.

As mentioned, 10 stories and their claim sets are presented for testing to each subject. Of these 10, 5 are *ES* versions and 5 *NES* versions. All subjects are presented with only one version of each story. The subscales of *NES* representation ability and *ES* representation ability will be calculated by summing scores (see section 4.5.3) gained from all responses to *ES* free claims (for example *b* and *d* above) for the *NES* subscale, and summing scores gained from all responses to claims that contain *ES* (for example *a* and *c* above) for the *ES* subscale.

Using the same schema for all claim sets and the same claim set for *ES* and *NES* versions of a story means that on *NES* question subjects will be asked to rate *ES* claims where there is no basis for inferring any character to have a salient emotional state. Although this is a consequence of trying to keep the two tasks (*ES* and *NES* representation tasks) as similar as possible, it has the benefit of allowing a more subtle subscale comparison than the one mentioned above. Subscale score can be separately calculated for *ES* and *NES* stories. This will result in each subscale having two component parts; one for *ES* stories and one for *NES* stories. This difference may prove significant, because the same score on the *ES* subscale may be gained in two ways. Either by failing to responding in a way congruent with correctly representing the character's emotional state when there is a salient emotion, or by responding in a way that shows incorrectly inferring an emotional state when there isn't likely to be one. Put differently, an imperfect score could be the result of false positive

representation of an emotion, or false negative representation of an emotion. Although this further division of the two subscales is not the main effect that will be looked at, the relative contribution to an effect of the subscale components are included in data analysis. This is made easier by the features that already need to be coded for data gathering – which subscale a given claim's response influences, and whether the story is an *ES* or *NES* version.

Given that there were two versions of each story, and only one version of each was to be presented to each subject, two tests were constructed (Test A and Test B). Test A included ES version of stories 3, 5, 6, 8, and 10 while stories 1, 2, 4, 7, and 9 were NES versions. This pattern was reversed for Test B. Having two tests also allowed alternate versions to be run on adjacent computers, meaning that subjects had different test stories to subjects sitting beside them.

4.6 Specific instrument details

4.6.1 Format

The decision was made to administer the test using computers by presenting the stories as text on the screen and having subjects respond using the keyboards. The reasons for this are as follows. Having each subject read stories of his or her own screen allows each subject to proceed at their own pace and accommodates differences in reading and processing time, which would not be as well accommodated if the text was displayed overhead or the story was read by a researcher.

E-Prime was used to construct the computerised version of the instrument. An important part of the research is that subjects can't go back and re-read a story once they have moved on to the claim set on which that story is based. The testing procedure was set up so that subjects were prompted by onscreen instructions (showed on the same screen as the story) to press 'space bar' to continue after they have finished reading the story. 'Spacebar' was the only allowed response and no other response had any effect. After pressing 'spacebar' onscreen instructions appeared on the same screen as one claim form the claim set, instructing the subject to rate the claim, displaying the rating options as well as which responses correspond to which rating. The five rating responses where the only responses that had any effect, and pressing one of these resulted in the next claim in the claim set being displayed, along with the same instructions that appeared with the previous claim. This process repeated until the subject had rated all the claims in the claim set (as they appeared in random order) after which the next story was displayed. This process repeated until subjects had completed the procedure for 10 stories and their claim sets. After the last claim set was rated, a screen appeared thanking subjects for their participation, and asking them to leave the test venue in silence leaving that instruction screen displayed.

4.6.2 Subjects

Given the reported ToM differences between science and humanities students, the decision was made to use science and humanities major university students as groups. To ensure students were rightly classed as either science or humanities students only 3rd year or postgraduate students were used and major and level of study were confirmed using university database information.

4.6.2.1 Recruitment

Physical recruitment notices were posted in areas where likely to be seen by suitable students; Science LANs, ARTS LANs, various departmental LANs in science and humanities departments. Some potential subjects were emailed by obtaining class lists for appropriate postgraduate and third year classes from academic staff. Recruitment was also done at postgraduate humanities recruitment drives on both campuses. University notices were sent out which contained a link to an online form where interested students could complete their details and only those suitable were contacted.

4.6.3 Subject instructions

When entering the testing venue subjects were read a set of instruction as follows. They were greeted and asked to turn mobile phones off and to not communicate in any way with other subjects. Each subject was give two informed consent forms; a copy to sign and a copy to keep. They were asked to take note of the three digit number on their consent form, which was to be used as their subject number to start the testing procedure. They were instructed how to start the testing procedure and asked to follow the onscreen instructions and to respond as honestly and accurately as possible during the procedure.

Starting the testing procedure initiated an instruction screen on the computer. Subjects were instructed how to rate claims and that the first story is a practice story. The practice story (labelled as such) was presented followed by the practice story claim set. After all claims were rated, a screen appeared informing the subject that they

should ask researchers for assistance if they had any trouble with the procedure, and to press a key to commence the testing procedure.

4.6.4 Ethics and Consent

There were no serious ethical concerns involved and the study was given ethical clearance under the normal conditions for research conducted on human subjects. Informed consent was obtained from all subjects, not only to take part in the study, but also for their university database information to be accessed and used.

4.6.5 Validity

I tested the validity of the instrument by testing whether there is some consensus that the Likert responses taken to be correct are correct. This was done by showing both versions of all the stories, and the claim sets for each to a 5 non-target population 'validity testers'. Validity testers were all academics, selected from a range of disciplines. None of the validity testers knew, at the time of testing, what the hypotheses were or what was being tested.

To improve feedback and to gather more information on how validity testers interpreted the stories, a list of mental states was included with the stories and claim sets mental state lists; a list for each story of the salient mental states (*ES* and *NES*) of character in that story (presented by character, under the character's name). These subjects could go back to the story to check relevant details to judge what mental states the characters would have in the situation described. For both versions of each of the 10 stories, validity testers were asked to read through the story, the mental state list and the claim set thoroughly.

Testers were asked to mark each mental state in the list with either a tick or a cross to indicate whether or not they thought it could, or could not be appropriately ascribed to the relevant story character. They were also asked to respond to each claim in the claim set by rating it on a 5 point Likert scale. The 5 points were labelled as 'Very Correct', 'Correct', 'Neither correct nor incorrect', 'Incorrect' and 'Very Incorrect'.

Of the validity testers, three responded only using 'Correct', 'Neither correct nor incorrect' and 'Incorrect' and indicated that they didn't think 'Very Correct' or 'Very Incorrect' were suitable responses. As a result of this, the points were re-labelled on the final instrument to exclude these as extreme responses, labelling points 1 and 5 as 'Correct' and 'Incorrect' respectively and points 2 and 4 as 'Near Correct' and 'Near Incorrect' respectively. Additionally, validity testers' responses to claims in the claim sets were reinterpreted as being on a 3 point, rather than 5 point scale; responses of 'Very Correct' or 'Correct' were reinterpreted as 'Correct' and ; responses of 'Very Incorrect' or 'Incorrect' were reinterpreted as 'Incorrect'.

All claims in the claim set where the response agreement (on the reinterpreted 3 point scale) was less than 4 out of 5 were excluded from the claim set. Thus, all claims in the final claims sets (see Appendix D) were rated the same on the three point scale by at least 4 of the 5 validity testers. See Appendix C for an example of the validity test.

Chapter 5 Results and Discussion

5.1 Method

5.1.1 Participants

The sample consisted of 38 students in total from Howard College and Pietermaritzburg campuses of University of KwaZulu-Natal. All participants were 3rd level or above. Participants were recruited using three methods; the University email notice system, posting notices on campus and approaching participants on campus for their co-operation. The electronic and physical notices contained links to an online form on which participants could sign up if they were interested by providing their contact details, major and level of study. Those that met the requirements of being a 3rd level or postgraduate student in either Science or Humanities were contacted to confirm their participation. All participation was voluntary and the incentive to take part was a chance to win one of four MP3 players.

5.1.2 Groups

Only Science and Humanities students were used in the dataset. Of the 38 participants that took part in the study, one was found to be a Commerce student and his/her data was excluded. Participant's results were also excluded if their score on control questions was less than 24 out of 40 (60%). The data of four participants (one Science major and three Humanities majors) were excluded on this basis. After exclusions, there were 9 participants in the Science group and 24 in the Humanities group.

5.1.3 Materials

Instrument

The development of the instrument used is discussed extensively in chapter 4. Each participant did one of two versions of the instrument. Both version contained half ES versions of stories and half NES versions. All 10 test stories (but not the practice story) were followed by claim sets of which 4 of the 5 claims were ToM claims (claims about characters' mental states). The participants then rated the claims on a 5-point Likert scale for their correctness. Answers were converted into score as discussed in 4.5.3. Participants were presented with 5 claims after each of the 10 stories and a response to each claim resulted in a score from 0 to 4. Each claim set contained two non-emotional state and two emotion state claims. Thus the maximum score a subject could get per story was 16 overall and 8 on each subscale. After responding to claim sets for all 10 stories, participants' score out of 80 was calculated for each subscale, summing to their total score out of 160.

5.1.4 Procedure.

Two testing venues were used; the multimedia learning centre (MMLC) LAN at Howard College campus and the New Arts LAN at Pietermaritzburg Campus.

E-Prime was used to construct the two computer based versions of the instrument. E-run[19] was installed the computers in the LAN and alternate test versions were copied onto adjacent machines.

[19] A 'Subject Station only' installation of part of E-Prime that allows running of scripts constructed using the full version of E-prime. E-run does not allow constructing or changing E-prime scripts.

When participants entered the testing venues, each was seated in front of a computer. On the screen was a text box prompting the participant to enter their subject number. They were asked not to begin until instructed to do so. They were instructed to…

After all the subjects had finished, the researcher ended the testing procedure and collected the data files.

5.2 Results

5.2.1 Group ToM means compared

Prior to all parametric tests, data distributions were tested for normality. All parametric test results reported conformed to the assumptions of those tests, with no Skewness or Kurtosis found in the distribution.

Fig. 5.1 below shows the group statistics for both the Science and the Humanities groups. The table shows each group's total ToM score (*ES* and *NES* scores together), emotional state scores (*ES* total), the non-emotional state score (*NES* total) and control score d from responses to control claims), which all showed normal distribution. Total ToM score was out of 160, *ES* and *NES* scores were both out of 80 and control score was out of 40.The mean score of the Science group was higher than that of the humanities group in all cases; ToM score, *ES* score, *NES* score and control score.

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
TotalToMScore	Science	9	119.33	12.104	4.035
	Humanities	24	103.42	12.039	2.458
ESTotal	Science	9	55.33	8.231	2.744
	Humanities	24	48.08	5.875	1.199
NESTotal	Science	9	64.00	4.873	1.624
	Humanities	24	55.33	8.375	1.710
ContolScore	Science	9	36.22	3.898	1.299
	Humanities	24	33.88	3.745	.765

Fig 5.1 – Group statistics showing Control score, *ES* total score and *NES* total score for the Science and Humanities groups.

Independent samples t-tests were done to compare Science and Humanities group means on each of Total ToM score, *ES* subscale score, *NES* subscale score and control score. Highly significant differences were found for Total ToM score (p = .002), *ES* score (p. = .008), *NES* score (p = .007). No significant difference was found for control score (p = .123).

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TotalToMScore	Equal variances assumed	.070	.793	3.378	31	.002	15.917	4.712	6.306	25.528
	Equal variances not assumed			3.369	14.351	.004	15.917	4.724	5.808	26.026
ESTotal	Equal variances assumed	1.875	.181	2.826	31	.008	7.250	2.566	2.017	12.483
	Equal variances not assumed			2.421	11.206	.034	7.250	2.994	.674	13.826
NESTotal	Equal variances assumed	3.705	.063	2.907	31	.007	8.667	2.981	2.587	14.747
	Equal variances not assumed			3.675	24.907	.001	8.667	2.358	3.809	13.525
ContolScore	Equal variances assumed	.002	.961	1.586	31	.123	2.347	1.480	-.670	5.365
	Equal variances not assumed			1.557	13.918	.142	2.347	1.508	-.888	5.582

Fig 5.2 – T-test for Total ToM Score (*ES* and *NES*), *ES* Score, *NES* Score and Control Score between the Science and Humanities groups

The above scores were compared for mean male and female and no significant differences were found between male/female means on Total ToM score, *ES* score, *NES* score or control score.

5.2.2 ToM score from different versions

The overall ToM (*ES* and *NES* together) means for the two groups were compared using an independent samples t-test. Group statistics showing the mean ToM score gained from *ES* version of stories and from *NES* versions of stories are shown for each of the two groups in Fig. 5 below. Mean ToM scores from *ES* versions and *NES* versions were out of 80 and both were higher for the Science group than the humanities group.

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
ESVersionToMScore	Science	9	61.44	6.287	2.096
	Humanities	24	54.79	8.377	1.710
NESVersionToMScore	Science	9	57.89	9.545	3.182
	Humanities	24	48.63	8.182	1.670

Fig. 5.3 Group statistics showing ToM score from *ES* versions and ToM score from *ES* versions.

The difference between groups mean ToM scores from *ES* version and *NES* version of stories were respectively tested using Independent samples t-tests. Significant difference was found between group scores for ToM score on *ES* versions (p = .039) and highly significant difference was found between the *NES* version scores (p= .009)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ESVersionToMScore	Equal variances assumed	2.537	.121	2.157	31	.039	6.653	3.084	.362	12.943
	Equal variances not assumed			2.460	19.232	.024	6.653	2.705	.996	12.309
NESVersionToMScore	Equal variances assumed	.044	.836	2.771	31	.009	9.264	3.344	2.444	16.083
	Equal variances not assumed			2.578	12.681	.023	9.264	3.593	1.481	17.047

Fig. 5.4 T-test for ToM score from *ES* versions of stories and ToM score from *ES* stories

ES and *NES* version ToM score were also calculated for males and females. Comparing the means showed that females outscored males on *ES* version ToM score (female mean = 58.23, male mean = 55.55) while males outscored females on *NES* version ToM score (female mean = 49.69, male mean = 52.10) as shown in fig. 5.5 below.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
ESVersionToMScore	Female	13	58.23	7.801	2.164
	Male	20	55.55	8.684	1.942
NESVersionToMScore	Female	13	49.69	4.973	1.379
	Male	20	52.10	11.443	2.559

Fig. 5.5 Group statistics for Male and Female ToM scores on *ES* versions and *NES* version of stories

There was no significant difference was found between male and female mean scores on either *ES* version ToM score (p. = .375) or *NES* version ToM score (p. = .415). The group ToM scores for males and females on *ES* versions did not have equal variance, while for group scores on ToM score on *NES* versions the variance in score differed significantly and equality of variance could not be assumed. See fig. 5.6 for details.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ESVersionToMScore	Equal variances assumed	1.057	.312	.901	31	.375	2.681	2.976	-3.389	8.751
	Equal variances not assumed			.922	27.747	.364	2.681	2.907	-3.277	8.639
NESVersionToMScore	Equal variances assumed	5.892	.021	-.713	31	.481	-2.408	3.377	-9.294	4.479
	Equal variances not assumed			-.828	27.914	.415	-2.408	2.907	-8.363	3.547

Fig. 5.6 Independent Samples T-test comparing male and female ToM score from *ES* Story version and *NES* story version respectively

5.2.3 Subscales from different versions compared

From the ToM scores from each version (*ES* version or *NES* version), ToM score was analysed by subscale to see how much of each score on each subscale contributes to the score from each version. Four score types came from this analysis; *ES* score from ES versions of stories (*ES* from *ES*), *NES* score from *ES* versions (NES from *ES*), *ES* score from *NES* versions (*ES* from *NES*) and *NES* score from *NES* versions (*NES* from *NES*). The Science group mean was higher than that of the Humanities group for each of the four score types, as shown in fig. 5.6 below.

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
ESfromES	Science	9	28.89	5.302	1.767
	Humanities	24	26.33	4.931	1.007
NESfromES	Science	9	32.56	2.242	.747
	Humanities	24	28.46	5.082	1.037
ESfromNES	Science	9	26.44	6.766	2.255
	Humanities	24	21.75	4.998	1.020
NESfromNES	Science	9	31.44	4.126	1.375
	Humanities	24	26.88	5.384	1.099

Fig. 5.6 Group statistics or Science and Humanities group *ES* score from *ES* versions, *NES* score from *ES* versions, *ES* score from *NES* versions and *NES* score from *NES* versions.

Independent sample t-tests were done to compare the group means of each of these four kinds, the details of which are in fig. 5.7. Significant differences were found between group means on *NES* from *ES* (p = .027), *ES* from *NES* (p = .037) and *NES* from *NES* (p = .029). Group means from *ES* scores from *ES* versions did not differ significantly. I discuss this further in 5.3.2.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ESfromES	Equal variances assumed	.003	.955	1.300	31	.203	2.556	1.966	-1.454	6.565
	Equal variances not assumed			1.256	13.537	.230	2.556	2.034	-1.821	6.932
NESfromES	Equal variances assumed	6.287	.018	2.318	31	.027	4.097	1.768	.492	7.703
	Equal variances not assumed			3.205	29.906	.003	4.097	1.279	1.486	6.709
ESfromNES	Equal variances assumed	1.377	.250	2.180	31	.037	4.694	2.153	.303	9.086
	Equal variances not assumed			1.897	11.442	.083	4.694	2.475	-.728	10.117
NESfromNES	Equal variances assumed	.787	.382	2.297	31	.029	4.569	1.989	.513	8.626
	Equal variances not assumed			2.595	18.807	.018	4.569	1.761	.882	8.257

Fig. 5.7 T-test comparing Science and Humanities group scores for each subscale on *ES* versions and *NES* version of stories; *ES* score from *ES* versions, *NES* score from *ES* versions, *ES* score from *NES* versions and *NES* score from *NES* versions.

Comparing the above scores (*ES* from *ES*, *NES* from *ES*, *ES* from *NES* and *NES* from NES) of males/females rather than Science/Humanities groups did not reveal a significant difference on any of the scores between males and females.

5.2.4 Matric Score

Participants' matric score was included in the dataset as a indicator of school performance. 5 participants' matric scores were not available, all from the Humanities

group. Science students had a mean matric score of 44.56 (*n = 9,* SD = 4.304) and Humanities students had a mean matric score of 33.68 (*n = 19,* SD = 6.263).

Performing an independent samples t-test showed the difference between Science and Humanities group scores to be very highly significant (p. < .001) as described in fig. 5.8 below. Matric score between males and females did not differ significantly (p. = .07; females: *n* = 12, mean = 34.17, SD = 7.907, males: *n* = 16, mean = 39.44, SD = 6.821).

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
MatricScore	Equal variances assumed	2.307	.141	4.687	26	.000	10.871	2.320	6.103	15.639
	Equal variances not assumed			5.354	22.179	.000	10.871	2.031	6.662	15.081

Fig. 5.8 Independent Samples T-test comparing Science and Humanities group matric scores

5.2.5 Correlation effects

Given that this was a pilot study and there were no comprehensive expectations from the data, Bivariate correlation was performed to check for correlation between variables (see fig. 5.9 for correlation table). Control score did not correlate significantly with Total ToM score (p = .292), Total *ES* score (p = .917) or Total N*E*S score (p = .097). Of all ToM scores and subscale scores, the only one which correlated significantly with control score was *ES* score from *NES* versions (p = .018)

Matric score correlated significantly with ToM score (p = .044), and highly significantly with Control score (p= .004). Matric score also correlated significantly with total *ES* score (p = .047), and *ES* from *NES* score (p = .037), but not with *ES* from *ES* (p = .647).

Correlations

		Group	Gender	MatricScore	ContolScore	TotalToMScore	ESTotal	NESTotal	ESVersionToMScore	NESVersionToMScore	ESfromES	NESfromES	ESfromNES	NESfromNES
Group	Pearson Correlation	1	-.215	-.677**	-.274	-.519**	-.453**	-.463**	-.361*	-.446**	-.227	-.384*	-.365*	-.381*
	Sig. (2-tailed)		.229	.000	.123	.002	.008	.007	.039	.009	.203	.027	.037	.029
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
Gender	Pearson Correlation	-.215	1	.348	.223	-.010	.033	-.044	-.160	.127	-.268	.006	.275	-.075
	Sig. (2-tailed)	.229		.070	.213	.957	.856	.807	.375	.481	.132	.972	.122	.680
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
MatricScore	Pearson Correlation	-.677**	.348	1	.529**	.384*	.378*	.295	.189	.380*	.091	.253	.396*	.224
	Sig. (2-tailed)	.000	.070		.004	.044	.047	.127	.337	.046	.647	.193	.037	.251
	N	28	28	28	28	28	28	28	28	28	28	28	28	28
ContolScore	Pearson Correlation	-.274	.223	.529**	1	.189	.019	.294	.187	.114	-.083	.410*	.096	.094
	Sig. (2-tailed)	.123	.213	.004		.292	.917	.097	.298	.529	.645	.018	.594	.604
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
TotalToMScore	Pearson Correlation	-.519**	-.010	.384*	.189	1	.862**	.901**	.751**	.811**	.494**	.776**	.642**	.718**
	Sig. (2-tailed)	.002	.957	.044	.292		.000	.000	.000	.000	.003	.000	.000	.000
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
ESTotal	Pearson Correlation	-.453**	.033	.378*	.019	.862**	1	.557**	.685**	.666**	.603**	.547**	.718**	.384*
	Sig. (2-tailed)	.008	.856	.047	.917	.000		.001	.000	.000	.000	.001	.000	.027
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
NESTotal	Pearson Correlation	-.463**	-.044	.295	.294	.901**	.557**	1	.644**	.759**	.293	.803**	.437*	.848**
	Sig. (2-tailed)	.007	.807	.127	.097	.000	.001		.000	.000	.097	.000	.011	.000
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
ESVersionToMScore	Pearson Correlation	-.361*	-.160	.189	.187	.751**	.685**	.644**	1	.222	.850**	.831**	.110	.267
	Sig. (2-tailed)	.039	.375	.337	.298	.000	.000	.000		.214	.000	.000	.541	.133
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
NESVersionToMScore	Pearson Correlation	-.446**	.127	.380*	.114	.811**	.666**	.759**	.222	1	-.024	.409*	.849**	.823**
	Sig. (2-tailed)	.009	.481	.046	.529	.000	.000	.000	.214		.896	.018	.000	.000
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
ESfromES	Pearson Correlation	-.227	-.268	.091	-.083	.494**	.603**	.293	.850**	-.024	1	.414*	-.122	.091
	Sig. (2-tailed)	.203	.132	.647	.645	.003	.000	.097	.000	.896		.017	.497	.616
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
NESfromES	Pearson Correlation	-.384*	.006	.253	.410*	.776**	.547**	.803**	.831**	.409*	.414*	1	.320	.366*
	Sig. (2-tailed)	.027	.972	.193	.018	.000	.001	.000	.000	.018	.017		.070	.036
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
ESfromNES	Pearson Correlation	-.365*	.275	.396*	.096	.642**	.718**	.437*	.110	.849**	-.122	.320	1	.399*
	Sig. (2-tailed)	.037	.122	.037	.594	.000	.000	.011	.541	.000	.497	.070		.021
	N	33	33	28	33	33	33	33	33	33	33	33	33	33
NESfromNES	Pearson Correlation	-.381*	-.075	.224	.094	.718**	.384*	.848**	.267	.823**	.091	.366*	.399*	1
	Sig. (2-tailed)	.029	.680	.251	.604	.000	.027	.000	.133	.000	.616	.036	.021	
	N	33	33	28	33	33	33	33	33	33	33	33	33	33

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

*. Correlation is significant at the 0.05 level (2-tailed).

Fig. 5.9 – Correlation table of variables

5.3 Discussion

This section discusses the results reported in section 5.2. The impact of the results on hypotheses and the possible reasons unexpected results are discussed, as well as improvements and extensions of the research.

5.3.1 Lack of established group ToM patterns

The Science group scored higher than the Humanities group on total ToM score as well as on *ES* and *NES* subscales (*ES* total score and *NES* total score respectively). This is contrary to previous findings that Humanities students perform better on ToM tests than Science students.

There were no significant differences between male and female total ToM score, *ES* score or *NES* score, which shows no evidence for either sex performing systematically better than the other in the ToM task.

The main ToM effects reported in literature between Science and Humanities groups and between males and females (see section 2.2.2) – that Humanities students perform better at ToM asks than Science students and that Females perform better than males – were not visible in the results.

5.3.2 Possible reasons for results

5.3.2.1 Sample size and bias

Late recruiting due to administrative delays with the project meant it was difficult to find suitable subjects and that recruitment response was generally poor. Even though the current research is a pilot study, the number of participants in the Science group is low enough that a few outliers might skew the data.

The low sample size might mean that all group difference effects, and particularly Science group scores, are not definitive and all hypotheses and possibilities discussed are somewhat speculative and need further data gathering to be considered as plausible hypotheses.

The assumption was, given that recruitment was aimed at the general student population, that the sample would be representative, within the constraints of the recruitment criteria, and that sample bias would not be problematic. Given the exploratory nature of this stage of the research this assumption *may* be justified if the sample was large enough that a bias in the sampling strategy would not severely affect the results. However, given the low sample size it both difficult to test for sampling bias and unwarranted to assume the absence of it.

5.3.2.2 General cognitive ability

The very highly significant difference found between Humanities and Science group matric scores is some indication that science students have a higher general cognitive ability. This raises the concern that Science students scored systematically higher due to higher general cognitive ability, rather than due to an aptitude specific to a

particular score such as ToM score. This possibility might explain why science students outperform humanities students on ToM score. Science student scoring higher than Humanities students could thus be a result of systematically higher cognitive ability rather than a result of Science students being better at ToM than intelligence matched Humanities students. Matched intelligence can also not be assumed given the matric score difference between the groups and should be corrected in subsequent testing.

5.3.2.3 No evidence of group effects in this population

A further problem for the existing research is that the effects discussed in literature (women scoring higher on ToM tests than men and Humanities students scoring higher than Science students) have not been established in the population from which the participants were selected. It may be that the effects established differ in kind or in magnitude in the current population from the populations in which they were observed. The findings by Wellman et al, 2001 (see 2.3.2) that country of origin played a highly significant role in false belief test performance support the possibility that different populations might show different ToM profiles.

5.3.3 Potentially interesting results

Having noted the problems discussed in 5.3.2, there are findings that might be of interest to investigate in future studies. Though there is significant correlation between matric score and ToM score ($p = .044$) the correlation between matric and control score is highly significant ($p = .004$) and the latter correlation is much stronger than the former. This might indicate that while ToM score is affected by something indicated by matric score like general intelligence, ToM is not affected by this other

factor as much as control score is affected by it. Put differently it might mean that while those with higher matric score perform better overall than those with lower matric score, this is less the case on ToM tasks than non-ToM tasks; there is still significant, but much lesser, impact on ToM score.

The result discussed in 5.2.3 that there is significant difference between Science and Humanities groups ToM score from *ES* version, as well as a highly significant difference between ToM scores from *NES* difference (see fig 5.4) shows something like the expected pattern. Taking into account the significant correlation between matric score and ToM score, the initial assumption might be to expect that this is equally the case for ToM score from *NES* and *ES* versions of stories. However, despite the Science group outscoring the Humanities group on both versions, the likelihood of this result being due to chance is greater in *ES* versions than *NES* versions. Thought this difference is small, given that the major difference between the two versions is that in *ES* versions one or more character has a salient emotional state, but not in *NES* versions, it would be interesting to compare the relative contributions of *ES* and *NES* versions to overall ToM score in future research where some of the concerns raised in 5.3.2 have been addressed.

As noted in 5.2.3 (also see fig. 5.5) female means score was higher than male mean score for ToM on *ES* versions, but for ToM on *NES* versions male mean score was higher than female mean score. Though the differences in ToM score on both versions were not significant, the direction of scoring is in agreement with the abovementioned possibility of Humanities students performing relatively better in ToM tasks that involve a salient emotional state. The suggestion is also a possibility for gender – that

females do relatively better, and males relatively worse, in ToM tasks that involve a salient emotional state. The current dataset is insufficient to propose these as serious hypotheses, but they are possibilities that are potentially interesting and given that this research is in the early stages, should be examined once a larger and more reliable dataset is available.

A further potentially interesting result (see 4.2.4) is that when comparing subscale scores from different story versions. The Science group outscored the Humanities group on all four scores (*ES* from *ES*, *NES* from *ES*, *ES* from *NES* and *NES* from *NES*). However, the difference between Science and Humanities groups *ES* score on *ES* version was not significant, while the group differences on the three other scores were all significant. This indicates that Humanities students do relatively better on the ES subscale when the situation includes salient emotional states, than they do on the NES subscale in such cases, or on the *ES* and *NES* subscales respectively when the situation does not include salient emotional states. Again noting that these are interesting possibilities rather than serious hypotheses, this might suggest that the Humanities group have a, relative to other scores, aptitude for 'positive *ES* knowledge'; that they are better at correctly rating claims about emotional states that are present, rather than correctly rating claims about non-emotional states or claims about emotional states where none are present. Similarly, matric score also correlated significantly with total *ES* score and *ES* score from *NES* versions, but not with *ES* score from *ES* versions. This also suggests that general intelligence, or some similar factor indicated my matric score, influences other kinds of ToM task more than those where knowledge is required and emotional states are present.

This raises the question of how much of existing ToM tests use situations in which the task requires participants to use (report, represent, or use in some way) emotional states that are genuinely present in the testing situation. If these are very common, this result may give some indication of why it has been previously found that Humanities students are stronger empathisers than Science students. There is then the possibility of a dissociation between two different kinds of emotional state knowledge; positive *ES* knowledge – where emotional states are correctly identified – and negative *ES* knowledge – where the absence of emotional states is correctly identified, or claims that they are present are correctly identified as false.

5.3.4 Improving subsequent stages of research

Given the low sample size and limitations this places on interpreting results, the research can be immediately improved with another round of data gathering with bigger samples. This will enable better interpretation of results, specifically give a better idea of overall group performance on the instrument used, which can then be more thoroughly compared to previous research findings and sample group differences. This will also give a better idea of whether or not the possibly interesting patterns in results discussed in 5.3.3 indicate anything important by checking if the patterns remain in bigger groups.

There is a strong possibility, given the correlation of matric score with many important test variables, that non-test factors such as general intelligence or working memory were influential in the results. This could possibly be why results were contrary to those reported in previous literature. To check for this, the likely influential non-test factors can be identified by researching literature on such effects

and the impact of these on matric score. Specifically, the role of highly relevant factors such as systemising ability plays in the matric scoring system should be considered. Based on what is found to be a likely non-test cause of the unexpected effect, further general intelligence or working memory test can either be included in the instrument or administered separately. To create room for correcting for any effects found by these a larger sample can be used to increase the flexibility to exclude outliers and normalise the distribution in both groups for this factor. This will also make it possible to weight scores in a way that corrects for the effects.

A further foundational improvement would be to perform the same, or similar, test in the population from which the present sample was taken that were used to establish the main group effects described in the literature review. Data showing the strength, assuming they are found, of these effects in the local population would also be useful for comparison with data from subsequent data gathering using an improved version of the instrument used in the current project.

Chapter 6: Conclusion

The initial stage of data gathering was disappointing. The focus of the project was, however, on the design of the instrument and what possible design features might adequately test the hypothesis. In this respect the results from the pilot study can be taken as ground for improving subsequent versions of the instrument, thought to do this effectively more data should be gathered using the current version. Surprising negative results, with humanities students failing to outscore science students in a theory of mind test, are also grounds for further research, specifically into possible reasons for this result if it is replicated. It is possible that the reported effects are limited to certain populations or tests, or that the current test lacks validity as a theory of mind test. This would in itself be interesting given the heavy reliance in designing the instrument on design features from existing theory of mind tests.

References List :

Abu-Akel, A. (2003). A neurobiological mapping of theory of mind. *Brain Research Reviews*, 43, 29–40.

Abu-Akel, A., Bailey, A.L. (2000). The possibility of different forms of theory of mind impairment in psychiatric and developmental disorders. *Psychological Medicine,* 30, 735–738.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: Author.

Apperly, I.A., Samson, D., Chiavarino, C., Humphreys, G.W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience,* 16, 1773–1784.

Asperger, H. (1944). Autistic psychopathy in childhood. Translated and annotated by U. Firth (Ed.) In *Autism and Asperger Syndrome* (1991). Cambridge: Cambridge University Press.

Baron-Cohen, S. (1991). Do People with Autism Understand What Causes Emotion? *Child Development,* 62: 2, 385-395.

Baron-Cohen, S. (1995). Mindblindness: an essay on autism and theory of mind. MIT Press/Bradford Books.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *TRENDS in Cognitive Sciences*, 6: 6, 248-254.

Baron-Cohen, S. (2003) The Essential Difference: men, women and the extreme male brain. Penguin, Allen Lane/ Basic Books

Baron-Cohen, S. (2008). Autism and Asperger syndrome: The facts. New York: Oxford.

Baron-Cohen, S., Joliffe, T., Mortimore, C., Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high-functioning adults with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, 38, 813–22.

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

Baron-Cohen, S., Wheelwright, S. (2004). The Empathy Quotient (EQ): An investigation of adults with Asperger Syndrome or High Functioning Autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163-175.

Baron-Cohen, S., Wheelright, 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. *Journal of Child Psychology and Psychiatry*, 42: 2, 241-251.

Blair, R.J., (2003). Facial expressions, their communicatory functions and neuro-cognitive substrates. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences,* 358, 561–572.

Blair, R.J.R., Mitchell, D.G.V., K.S. Peschardt, K.S, E., Colledge, E., Leonard., R.A., Shine, J.H., Murray, L.K., Perrett, D.I. (2004). Reduced sensitivity to others' fearful expressions in psychopathic individuals. *Personality and Individual Differences,* 37, 1111–1122.

Bloom, P., German, T.P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition,* 77, B25-B31.

Bowler, D.M. (1992). 'Theory of mind' in Asperger's syndrome. *Journal of Child Psychology and Psychiatry,* 33, 877–893.

Carol, J., Chiew, K.Y. (2006). Sex and Discipline Differences in Empathising, Systemising and Autistic Symptomatology: Evidence from a Student Population. *Journal of Autism and Developmental Disorders*, 36: 7, 949-957.

Carr, L., Iacoboni, M., Dubeau, M., Mazziotta, J., Lenzi, G.L.(2003) Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Science*, 100: 9, 5497–5502.

Clements, W.A., Perner, J. (2001). When actions really do speak louder than words, but only explicitly; young children's understanding of false belief in action. *British Journal of Developmental Psychology,* 19, 413–432.

Corcoran, R., Mercer, G., Frith, C.D. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research,* 17, 5-13.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., Jeanerod, M. (1997). Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition,* 65, 71–86.

Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1, 568-570.

Focquaert, F., Steven, M.S., Wolford, G.L., Colden, A., Gazzaniga, M.S. (2007). Empathizing and systemizing cognitive traits in the sciences and humanities. *Personality and Individual Differences,* 43, 619–625.

Frangou, S., Chitins, X., Williams, S.C.R. (2004). Mapping IQ and gray matter density in healthy young people. *NeuroImage 23:3*, 800-805.

Frith, U., (2003). Autism: Explaining the Enigma, second edition. Blackwell, Oxford.

Gallagher, H.L., Happe, Â.F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D. (2000). Reading the mind in cartoons and stories: an fMRI study of `theory of mind' in verbal and nonverbal tasks. *Neuropsychologia,* 38, 11-21.

German, T. P., Leslie, A. M. (2000). Attending to and learning about mental states. In Mitchell, P., Riggs, K. (Eds.) *Children's reasoning and the mind*. Hove: Psychology Press.

German, T.P., Niehaus, J.L., Roarty, M., Giesbrecht, B., Miller, M.B. (2004). Neural Correlates of Detecting Pretense: Automatic Engagement of the Intentional Stance under Covert Conditions. *Journal of Cognitive Neuroscience* 16:10, 1805-1817.

Gogtay, N., Giedd, J., Lusk, L., Hayashi, K., Greenstein, D., Vaituzis, A., et al. (2004). Dynamic mapping ofhuman cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 8174–8179.

Gopnik, A., Astington, J. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child Development*, 59, 26-37.

Gregory, C., Lough, S., Stone, V.E., Erzinclioglu, S., Martin, L., Baron-Cohen, S., Hodges, J. (2002). Theory of mind in frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain*, 125, 752-64.

Grice, S. J., Halit, H., Farroni, T., Baron-Cohen, S., Bolton, P., Johnson, M. H. (2005). Neural correlates of eye-gaze detection in young children with autism. *Cortex*, 41, 342–353.

Happé, F.G.E. (1994). An Advanced Test of Theory of Mind: Understanding of Story Characters' Thoughts and Feelings by Able Autistic, Mentally Handicapped and Normal Children and Adults. *Journal of Autism and Developmental Disorders,* 24*,* 129-154.

Hare, R. D. (1991). The Hare psychopathy checklist-revised. Toronto, ON: Multi-Health Systems.

Hill, E.L., Frith, U., (2003). Understanding Autism: Insights from Mind and Brain. *Philosophical Transactions: Biological Sciences*, 358: 1430, 281-289.

Holtmann, M., Bolte, S., Poustka, F. (2007). Autism spectrum disorders: sex differences in autistic behavior domains and coexisting psychopathology. *Developmental Medicine & Child Neurology,* 49, 361-366.

Kinderman, P., Dunbar, R., Bentall, R.P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology,* 89, 191–204.

Lawson, J., Baron-Cohen, S., Wheelright, S. (2004). Empathising and Systemising in Adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders,* 34: 3, 301-310.

Leslie, A.M (1987). Pretense and Representation: The Origins of "Theory of Mind". *Psychological Review* 94: 4, 412-426

Leslie, A.M. (1991). Theory of mind impairment in autism. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 63-77). Oxford: Basil Blackwell.

Mitchell, J.P., Banaji, M.R., Macrae, C.N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience,* 17, 1306–1315.

Mumme, D., Fernald, A. (2003.) The Infant as Onlooker: Learning From Emotional Reactions Observed in a Television Scenario.
Child Development, 74: 1, 221-237.

Newschaffer, C.J., Croen, L.A., Daniels, J., Giarelli, E., Grether, J.K., Levy, S.E., Mandell, D.S., Miller, L.A., Pinto-Martin, J., Reaven, J., Reynolds, A.M., Rice, C.E.,

Schendel, D., Windham, G.C. (2007). The Epidemiology of Autism Spectrum Disorders. *Reviews in Advance,* 7: 53, 21.1 - 21.24.

Perner, J., Leekam, S., Wimmer, H. (1987). Three-year-olds' difficulty with the false belief task: The case for a conceptual deficit. *British Journal of Psychology*, 5, 125-137.

Pickup, G., Frith, C. (2001). Theory of mind impairments in schizophrenia: Unisymptomatology, severity and specificity. *Psychological Medicine,* 31, 207–220.

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

Pylyshyn, Z.W. (1978). When is attribution of beliefs justified? *Behavioural and brain Sciences,* 1, 592 – 593.

Repacholi, B. M. 1998 Infants' use of attentional cues to identify the referent of another person's emotional expression. *Developmental Psychology* 34, 1017-1025.

Roth, D., Leslie, A.M. (1998). Solving belief problems: toward a task analysis. *Cognition,* 66, 1–31.

Schenkel L. S., Marlow-O'Connor, M., Moss M., Sweeney J.A., Pavuluri, M.N. (2008). Theory of mind and social inference in children and adolescents with bipolar disorder. *Psychological Medicine*,38, 791–800

Shaw, P., Lawrence, E., Bramham, J., Brierley , B., Radbourne C., David A.S. (2007). A prospective study of the effects of anterior temporal lobectomy on emotion recognition and theory of mind. *Neuropsychologia, 45*, 2783–2790.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading:Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews,* 30, 855–863.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature,* 439, 466–469.

Sowell, E.R., Peterson, B.S., Thompson, P.M., Welcome, S.E., Henkenius, A.L., Toga, A.W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience,* 6, 309–315.

Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal Mapping of Cortical Thickness and Brain Growth in Normal Children . *The Journal of Neuroscience,* 24: 38, 8223-8231.

Stiller, J., Dunbar, R.I.M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks,* 29, 93–104.

Stone, V.E., Baron-Cohen, S., Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10, 640-656.

Stone, V.E., Baron-Cohen, S., Calder, A., Keane, J., Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions *Neuropsychologia*, 209–220.

Tager-Flusberg, Baron Cohen, S Cohen, D. (1993). An introduction into the debate, in Baron-Cohen, S., Tager-Flusberg, H., Cohen, D. (1993). *Understanding Other Minds*: *Perspectives from Autism*. Oxford Medical Publications.

Wellman, H.M., Cross, D., Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Beliefs. *Child Development,* 72: 3, 655-684.

Wicker, B., Keysers, C., Plailly, J., Royet, J.P., Gallese, V., Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron,* 40, 655–664.

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

Appendix A:

Selection from the Imposing Memory Task (Kinderman et al, 1998)

WHERE'S THE POST OFFICE ?

Sam wanted to find a Post Office so he could buy a Tax Disc for his car. He asked Henry if he could tell him where to get one. Henry told him that he thought there was a Post Office in Elm Street. When Sam got to Elm Street, he found it was closed. A notice on the door said that it had moved to new premises in Bold Street. So Sam went to Bold Street and found the new Post Office. When he got to the counter, he discovered that he had left his MOT certificate at home. He realized that without an MOT certificate, he could not get a Tax Disc, so he went home empty-handed.

Where's the Post Office

Please tick the correct answer to each question:

1. (a) Sam wanted to go to the Post Office to buy a stamp
   (b) Sam wanted to go to the Post Office to buy a Tax Disc

2. (a) Henry thought Sam would find the Post Office in Elm Street
   (b) Henry thought Sam would find the Post Office in Bold Street

3. (a) The Post Office had moved from Bold Street to Elm Street
   (b) The Post Office had moved from Elm Street to Bold Street

4. (a) Sam thought that Henry knew the Post Office was in Bold Street
   (b) Sam thought that Henry knew the Post Office was in Elm Street

5. (a) The Post Office in Elm Street had a notice in the window saying it had moved to Bold Street
   (b) The Post Office in Elm Street had a notice on the door saying it had moved to Bold Street

6. (a) Sam thought that Henry believed that Sam wanted to buy a Tax Disc
   (b) Sam thought that Henry did not know that Sam wanted to buy a Tax Disc

7 (a) When Sam got to Bold Street to buy his Tax Disc, he realised that he wouldn't be able to buy it because he had forgotten his MOT certificate.
   (b) When Sam got to Bold Street to buy his Tax Disc, he realised that he wouldn't be able to buy it because he had forgotten his insurance certificate.

JOHN'S PROBLEM

It was nearly the end of the day, John thought it might be nice to go to the pub for a drink after work. At first, he wasn't sure whom he should ask to go with him. He very much wanted to ask Sheila, whom he fancied, but he thought that she didn't like him enough to want to give up her aerobics class to go drinking with him. He could, of course ask Pete, his usual drinking companion. Pete was always happy to spend an hour or two in the pub before going home. Then he happened to see Penny. He knew that Penny was one of Sheila's friends. Penny might be able to help him out. She would know whether Sheila would be willing to go out for a drink rather than go to her aerobics class. "Listen Penny," he said, "I thought I might go for a drink after work. I was going to ask you and Sheila if you wanted to come. Would you ask Sheila whether she would like to come for a drink with us?" Penny looked surprised. John had never asked her to go out with him before, but she thought that he was very keen on Sheila. She began to suspect that John wanted to find out whether she knew what Sheila might want to do.

John's Problem - Please tick the correct answer to each question:

1. (a) The story was set in the morning
   (b) The story was set in the afternoon

2. (a) John wanted to go home after work
   (b) John wanted to go to the pub after work

3 (a) After work, Sheila was going to an aerobics class
   (b) After work, Sheila was going home

4. (a) John thought Sheila would not like to go to the pub with him
   (b) John thought Sheila would like to go to the pub with him

5. (a) John and Pete often went for a drink together
   (b) John and Pete only rarely went for a drink together

6. (a) John thought that Penny knew what Sheila wanted to do
   (b) John thought that Penny did not know what Sheila wanted to do

7. (a) John's friend, Pete, occasionally went for a drink, but never after work, always going home
   (b) John's friend, Pete, occasionally went for a drink in the evening, after work

8. (a) Penny believed that John thought she would not know what Sheila would want to do
   (b) Penny believed that John was hoping she would know what Sheila would want to do

9. (a) John spoke to Penny, but neither Sheila or Pete, about going for a drink after work
   (b) John spoke to Penny and Pete, but not Sheila, about going for a drink after work

10. (a) John thought that Penny thought that John wanted Penny to find out what Sheila wanted to do because John wanted to go out with Sheila alone
    (b) John thought that Penny thought that John wanted Penny to find out what Sheila wanted to do because John wanted to go out with them both

11. (a) Penny, the woman that John spoke to about asking Sheila about going for a drink after work, after he had thought of asking Pete, was a friend of Sheila's
    (b) Penny, the woman that John spoke to about asking Sheila about going for a drink after work, after he had thought of asking Pete, did not know Sheila

EMMA'S DILEMMA

Emma worked in a greengrocer's. She wanted to persuade her boss to give her an increase in wages. So she asked her friend Jenny, who was still at school, what she should say to the boss. "Tell him that the chemist near where you live want's you to work in his shop." Jenny suggested. "The boss won't want to lose you, so he will give you more money" she said. So when Emma went to see her boss, that is what she told him. Her boss thought that Emma might be telling a lie, so he said he would think about it. Later, he went to the chemist's shop near Emma's house and asked the chemist whether he had offered a job to Emma. The chemist said he hadn't offered Emma a job. The next day the boss told Emma that he wouldn't give her an increase in wages, and she could take the job at the chemist's instead.

Emma's Dilemma

Please tick the correct answer to each question :

1. (a) Emma worked for a greengrocer
   (b) Emma worked in a chemist's

2. (a) Emma wanted more money
   (b) Emma wanted a different job

3 (a) Emma's friend, Jenny, was still at school
   (b) Emma's friend, Jenny, worked in a bank

4. (a) Jenny thought the boss would believe Emma's story
   (b) Jenny knew the boss would not believe Emma's story

5. (a) Emma told her boss, the greengrocer, that she had been offered a job in an bank
   (b) Emma told her boss, the greengrocer, that she had been offered a job in a chemist's

6. (a)Emma thought the boss believed that the chemist wanted her to work for him
   (b) Emma thought the boss knew that the chemist had not offered her a job

7. (a) Emma's boss, the greengrocer, asked the chemist if he had offered Emma a job
   (b) Emma's boss, the greengrocer, asked Jenny if Emma had been offered a job

8. (a) Jenny thought that Emma hoped that the boss would believe that the chemist wanted Emma to work for him
   (b) Jenny thought that Emma believed that the boss knew that the chemist did not want Emma to work for him

9. (a) The chemist's shop, where Jenny had suggested that Emma tell her boss that she had been offered a job, was near where Emma lived

(b) The chemist's shop, where Jenny had suggested that Emma tell her boss that she had been offered a job, was in a different town

SIMON

Simon was 19 years old and worked as a mechanic. His cousin, Jim, was quite a lot older, and worked as a milkman. Because he got up early in the morning, he seldom went out in the evening. Jim's friend, Edward, worked in a bank, and therefore had more opportunity to go out in the evenings. Simon knew that Jim wanted to marry Susan. Simon also knew that Jim believed that Susan wanted to marry Edward. So he thought that if he could convince Jim that Susan thought that Edward wanted to marry Betty, Jim might be persuaded that Susan would say "Yes"if he asked her to marry him.
Simon Thinks

Please tick the correct answer to each question :

1. (a) Simon worked as a mechanic
   (b) Simon worked in a greengrocers

2. (a) Jim wanted to marry Susan
   (b) Jim did not want to marry Susan

3. (a) Jim's friend, Edward, worked in a bank
   (b) Jim's friend, Edward, worked as a mechanic

4. (a)Simon believed that Jim was convinced that Susan would not marry him
   (b) Simon thought that Jim though that Susan would marry him

5. (a) Simon, who was 19 years old, was Jim's cousin
   (b) Simon, who was 19 years old, was Jim's brother

6. (a) Jim believed that Susan thought that Edward would like to marry Betty
   (b) Jim thought that Susan knew that Edward did not want to marry Betty

7. (a)Because Jim worked as a milkman, and Edward worked in a bank, neither went out very often
   (b)Because Jim worked as a milkman, but Edward worked in a bank, Edward went out more often than Jim

8. (a)Simon hoped that Jim would believe that Susan thought that Edward wanted to marry Betty
   (b)Simon thought that Jim would believe that Susan thought that Edward did not want to marry Betty

9. (a) Edward's friend, Jim, who was Simon's cousin, was older than Simon, who was 19
   (b)Edward's friend, Jim, who was Simon's cousin, was younger than Simon, who was 19

Appendix B:

Positive Valence item (Schenkel et al, 2008)

Positive-valenced

Kathy and Mary live a few short blocks from Julie. On weekends, Kathy, Mary, and Julie usually see each other. On Saturday morning, Julie was getting ready to go to a party at Kathy's house. The doorbell rang just after Julie finished getting dressed. It was Kathy and Mary. They were excited to see Julie. 'Hey, are you ready to go to the party?' they asked. 'Yes, let's go! '

Julie said in an excited voice. Julie brought some of her games with her. Before she left, Julie's mother said, 'Have fun at the party. Be back by 6:00 for dinner.' At the party, Julie saw many of her friends from school. The kids in the room were laughing as Julie walked into the room. Julie asked, 'What's going on?' The kids were taking turns telling jokes to each other. Kathy told a joke and Julie laughed so hard she fell on the floor. Julie yelled out 'Good one! ' Julie told another joke and everyone laughed even more. Then Kathy's mom came into the room and said it was time to eat. All the kids raced into the kitchen to eat. Julie was the first one to get a piece of pizza. After the pizza, all the kids played the games Julie brought from home. Julie won all of the games she played. Julie got first prize for winning the games and was given a gift to take home.

At the end of the party, Julie left with her gift and games. She walked home with Mary. Just before Julie got home, Mary told a joke. Julie started to laugh. She was laughing as she walked into her house. She walked into the kitchen and saw her mother with her little sister. Her little sister was getting ready for a costume party and was wearing a silly mask.

False-belief question: Why does Julie's mother think she is laughing?

Control question: Why do you think she is laughing?

Appendix C:

Thank you for agreeing to look at the following questionnaire. Please read each story carefully and then complete the tasks labeled 1) and 2) in the table below the story. Read the story again if you need to when doing the tasks. Take a break if you find you are having difficulty doing the whole questionnaire in one sitting. Please make sure that before you do a given tasks, you have read the story directly above it. Note that there are two versions of all the stories (A & B) and that they are not the same and that each requires you to do the two tasks. When you are done, please return the questionnaire as soon as possible.

Thank you in advance for your time and assistance.

2. Ugly curtains (A)

Jill had recently moved into a new flat. She thought the outdated curtains and lampshades that were in the bedroom when she moved in were ugly. She went shopping and bought some new curtains for her bedroom which she thought were beautiful. When she had put them up, her best friend, Lisa, came over. Jill gave her a tour of the apartment and asked, "How do you like my place?"
"It's lovely! And I know of a place where you can get some beautiful curtains for your bedroom," Said Lisa.
"Oh, okay. By the way, I bought the curtains yesterday from the mall." Said Jill.

1.) Here is list of possible mental states of each character. Please indicate which of these you think the character will have in the situation depicted in the story with a tick, and those the character won't have with an 'x'.	**2.) Please indicate how correct each of the following claims are by circling the corresponding option below each claim.**
Jill - is unhappy. - thinks the curtains are beautiful. - thinks Lisa thinks the curtains are ugly. - thinks she recently bought the curtains. **Lisa** - feels bad for insulting the curtains. - thinks the curtains are ugly. - didn't think before Jill said so that Jill recently bought the curtains. - thought at first the curtains were already in the flat. - thinks at the end Jill thinks the curtains are not ugly. - thought at first that Jill thinks she thinks the curtains are ugly - thinks at the end that Jill recently bought the curtains.	***Jill feels as happy at the end as she did when Lisa arrived.*** **Very Correct Correct Neither Correct nor Incorrect Incorrect Very Incorrect** ***When she first arrived, Lisa thought Jill recently bought the curtains in the bedroom.*** **Very Correct Correct Neither Correct nor Incorrect Incorrect Very Incorrect** ***At the end Lisa thinks that Jill is just as happy as when Lisa first arrived.*** **Very Correct Correct Neither Correct nor Incorrect Incorrect Very Incorrect** ***At the end, Lisa thought that Jill thought Lisa liked the curtains.*** **Very Correct Correct Neither Correct nor Incorrect Incorrect Very Incorrect** ***When Jill went shopping she bought new curtains for her flat .*** **Very Correct Correct Neither Correct nor Incorrect Incorrect Very Incorrect**

Appendix D – Instrument

Each story in the instrument and its claim set:

Practice story (Test A and B)

Duma has recently got a puppy for his tenth birthday. One weekend his parents decided to go to the beach and Duma took his puppy with him. At the beach, Duma got hot and decided to go with his dad for a swim. Duma's mom was reading a magazine and he asked her to look after his puppy for a while. He took off his shirt and he ran to the water. Before Duma's mom could stop it, the puppy ran after Duma. The puppy ran to the water, but stopped just before the water line.Duma's mom got up to go get the puppy, but before she could get to it a big wave broke and washed the puppy off its feet. The puppy came running back to Duma's mom, covered in water and sand.

Duma is nine years old.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Duma's mom got up to get the puppy after it ran after Duma.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Before he went onto the beach, Duma took off his shirt

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

There are many other people on the beach with Duma and his parents.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Duma's dad did not come to the beach.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

1. Surprise Trip (ES) Test B

Dean and Clair are married. For the last month Dean has been planning a weekend trip for Clair's upcoming birthday, but hasn't mentioned it to her or anyone else. He wanted to ask their friends Patrick and Susan if they would house-sit and Dean thought he'd ask that night when they came over for dinner.

After dinner, when Dean and Patrick were alone in the dining room, Dean said to Patrick: "I'm arranging to go away with Clair for her birthday next weekend. Would you and Susan look after the house while we're gone?"
"That sounds good. I'll check my schedule, but I'll let you know tomorrow", Patrick said just as Clair walked into the room. "No problem," Said Dean. "What are you two talking about?" Clair asked.
Dean was about to answer when Patrick said, "Dean was just telling me about the trip you're going on for your birthday next weekend."
"A trip away! What a good idea, Dean." said Clair. After Clair went back to join Susan in the lounge, Dean said to Patrick "I've always liked surprising Clair for her birthday."
A short while later, Patrick and Susan went home.

When Susan and Patrick went home Dean was not just as happy as he was before dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When Susan and Patrick went home, Clair thought she was getting a trip for her birthday

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When Patrick and Susan went home Patrick thought that Clair thought that she was getting a trip.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end of the story Patrick thinks Dean thinks that Clair doesn't think she's getting a trip.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Patrick doesn't think Dean was less happy when they said goodnight than before dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Dean decided to plan the trip when Susan and Patrick came over for dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

1. Surprise Trip (NES) Test A

Dean and Clair are married. For the last month Dean has been planning a weekend trip for Clair's upcoming birthday, but hasn't mentioned it to her or anyone else. He wanted to ask their friends Patrick and Susan if they would house-sit and he thought he'd ask that night when they came over for dinner.

After dinner, when Dean and Patrick were alone in the dining room, Dean said to Patrick: "I'm arranging to go away with Clair for her birthday next weekend. Would you and Susan look after the house while we're gone?"
"That sounds good. I'll check my schedule, but I'll let you know tomorrow", Patrick said just as Susan walked into the room. "No problem," Said Dean. "What are you two talking about?" Susan asked.
Dean was about to answer when Patrick said, "Dean was just telling me about the trip they're going on for Clair's birthday next weekend."
"A trip away! What a good idea, Dean." said Susan. After Susan went back to join Clair in the lounge, Dean said to Patrick "I've always liked surprising Clair for her birthday."
A short while later, Patrick and Susan went home.

When Susan and Patrick went home Dean was not just as happy as he was before dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When Susan and Patrick went home, Clair thought she was getting a trip for her birthday.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end of the story Patrick thinks Dean thinks that Clair doesn't think she's getting a trip.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Patrick doesn't think Dean was less happy when they said goodnight than before dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Dean decided to plan the trip when Susan and Patrick came over for dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

2. Ugly curtains (ES) Test B

Jill had recently moved into a new flat. She thought the outdated curtains and lampshades that were in the bedroom when she moved in were ugly. She went shopping and bought some new curtains for her bedroom which she thought were beautiful. When she had put them up, her best friend, Lisa, came over. Jill gave her a tour of the apartment and asked, "How do you like my place?"
“It's lovely! And I know of a place where you can get some beautiful curtains for your bedroom,” Said Lisa.
“Oh, okay. By the way, I bought the curtains yesterday from the mall.” Said Jill.

Jill feels as happy at the end as she did when Lisa arrived.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When she first arrived, Lisa thought Jill recently bought the curtains in the bedroom.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end Lisa thinks that Jill is just as happy as when Lisa first arrived.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end, Lisa thought that Jill thought Lisa liked the curtains.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When Jill went shopping she bought new curtains for her flat .

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

2. Ugly curtains (NES) Test A

Jill had recently moved into a new flat. She thought the outdated curtains and lampshades that were in the bedroom when she moved in were ugly. She went shopping and bought some new lampshades for her bedroom which she thought were beautiful. When she had put them up, her best friend, Lisa, came over. Jill gave her a tour of the apartment and asked, "How do you like my place?"
“It's lovely! And I know of a place where you can get some beautiful curtains for your bedroom,” Said Lisa.
“Oh okay. By the way, I bought the lampshades yesterday from the mall,” Said Jill.

Jill feels as happy at the end as she felt when Lisa arrived.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When she first arrived, Lisa thought Jill recently bought the curtains in the bedroom.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end Lisa thinks that Jill is just as happy as when Lisa first arrived.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end, Lisa thought that Jill thought Lisa liked the curtains.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When Jill went shopping she bought new curtains for her flat.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

3. Gender issue (ES) Test A

Sally is a six-year-old girl with a round face and short blonde hair and is tall for her age. Her mother Carol is taking her to play putt-putt. Sally runs ahead to look at the course while Carol reads the sign showing the prices: R12 for adults and R8 for children. It also says that children under seven years old can play free if they're with adults. Carol has to wait a short while for a big group of children who are about to start playing. Sally comes back as Carol hands R20 to the attendant. He goes to the back shelf where there's a till as well as pink and blue putters. He pauses for a moment, looks at Sally, and then goes back to gathering the equipment and putting the money in the till.

Sally looks at the group of children and says, "Mommy, do all the boys get blue ones and all the girls get pink ones?"

"That's right," says the attendant handing a blue putter to Sally. He gives Carol a pink putter, two golf balls and her change. Carol looks at all the items and says; "Are you sure that's right? It's me and my six-year-old daughter." The attendant looks at Sally again. "Oh, Sorry! My mistake," he says and gets Sally a pink putter from the shelf.

Carol and Sally leave the counter to start their game.

Carol is not less happy when they start their game than when they first arrived

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When the attendant first gave Sally a putter he thought she was a boy

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Carol thinks that when they first arrived the attendant thought Sally was a boy.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Carol thinks Sally is less happy when they leave the counter than when they first arrived

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

According to the sign showing the prices, Sally could not play for free

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

3. Gender issue (NES)Test B

Sally is a six-year-old girl with a round face and short blonde hair and is tall for her age. Her mother Carol is taking her to play putt-putt. Sally runs ahead to look at the course while Carol reads the sign showing the prices: R12 for adults and R8 for children.It also says that children under seven years old can play free if they're with adults. Carol has to wait a short while for a big group of children who are about to start playing. Sally comes back as Carol hands R20 to the attendant. He goes to the back shelf where there's a till as well as pink and blue putters. He pauses for a moment, looks at Sally, and then goes back to gathering the equipment and putting the money in the till.

Sally looks at the group of children and says, "Mommy, do all the boys get blue ones and all the girls get pink ones?"

"That's right," says the attendant handing a pink putter to Sally. He gives Carol a pink putter, two golf balls and a receipt. Carol looks at all the items and says; "Are you sure that's right? It's me and my six-year-old daughter." The attendant looks at Sally again. "Oh, Sorry! My mistake," he says and gets Carol's change from the till.

Carol and Sally leave the counter to start their game.

Carol is not less happy when they start their game than when they first arrived

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

When the attendant first gave Sally a putter he thought she was a boy

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Carol thinks that when they first arrived the attendant thought Sally was a boy.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Carol thinks Sally is less happy when they leave the counter than when they first arrived

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

According to the sign showing the prices, Sally could not play for free

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

4. Traffic (ES) Test B

Sebastian and his wife Melanie are about to leave for work. Sebastian has recently bought a motorbike after persuading Melanie it would be safe since he only has a short drive to work on the main highway leading into the city centre. They both leave for work, though Melanie has a much longer drive out of the city to the northern suburbs. Only a few kilometres from home, Sebastian gets a flat tyre. He phones his friend Bill, who works as a mechanic close to where Sebastian broke down. Bill takes Sebastian's bike to his workshop and says he'll fix the flat tyre. Sebastian says he will pick it up the following day in his bakkie and he walks to the nearest bus stop.

A short while later, still on her way to work, Melanie turns on the radio just in time for the traffic report. The reporter says that there are serious delays in the city centre due to a truck colliding with a motorcycle on the main highway. She pulls over at the next turnoff and gets her cellphone out of her handbag.

Melanie is just as happy when she turns on the radio as she is after she hears the traffic report.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sebastian thinks there is an accident going into the city centre.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sebastian thinks Melanie thinks there has been an accident involving a bike.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the time that Melanie pulls over Sebastian does not think that Melanie is less happy than when she left for work.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sebastian has a longer drive to work than Melanie.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

4. Traffic (NES) Test A

Sebastian and his wife Melanie are about to leave for work. Sebastian has recently bought a motorbike after persuading Melanie it would be safe since he only has a short drive to work on the main highway leading into the city centre. They both leave for work, though Melanie has a much longer drive out of the city to the northern suburbs. Only a few kilometres from home, Sebastian gets a flat tyre. He phones his friend Bill, who works as a mechanic close to where Sebastian broke down. Bill takes Sebastian's bike to his workshop and says he'll fix the flat tyre. Sebastian says he will pick it up the following day in his bakkie and he walks to the nearest bus stop.

A short while later, still on her way to work, Melanie turns on the radio just in time for the traffic report. The reporter says that there are serious delays on the way to the northern suburbs due to a truck colliding with a taxi. She pulls over at the next turnoff and gets her cellphone out of her handbag.

Melanie is just as happy when she turns on the radio as she is after she hears the traffic report.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sebastian thinks there is an accident going into the city centre.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sebastian thinks Melanie thinks there has been an accident involving a bike.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the time that Melanie pulls over Sebastian does not think that Melanie is less happy than when she left for work.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sebastian has a longer drive to work than Melanie.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

5. New Kid (ES) Test A

Mike, a nine-year-old boy, just started at a new school. Joe and Peter, two other boys, were in the bathroom at school in two of the stalls.
Joe said, "You know that new guy in the class?"
"No. Who?" asked Peter.
"His name's Mike. His hair looks weird. And he's short."
"Oh yes," said Peter, "I know who you mean now."
Just a short while later Mike came into the bathroom and started taking off his shoes to get changed. While Joe was still in his stall, Peter came out of his stall and saw Mike.

Peter was equally happy before and after he saw Mike.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Peter thinks Mike may have heard what Joe said.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Joe thinks Peter thinks Mike is in the bathroom.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

After Peter sees Mike, Peter thinks Mike is unhappy.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Mike was in the bathroom when Joe and Peter were talking.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

5. New Kid (NES) Test B

Mike, a nine-year-old boy, just started at a new school. Joe and Peter, two other boys, were in the bathroom at school in two of the stalls.
Joe said, "You know that new guy in the class?"
"No. Who?" asked Peter.
"His name's Mike. His hair is blonde. And it's short."
"Oh yes," said Peter, "I know who you mean now."
Just a short while later Mike came into the bathroom and started taking off his shoes to get changed. While Joe was still in his stall, Peter came out of his stall and saw Mike.

Peter was equally happy before and after he saw Mike.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Peter thinks Mike may have heard what Joe said.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Joe thinks Peter thinks Mike is in the bathroom.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

After Peter sees Mike, Peter thinks Mike is unhappy.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Mike was in the bathroom when Joe and Peter were talking.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

6. Pie (ES) Test A

Kim's cousin, Scott, was coming to visit her for the weekend and Kim was making a pecan nut pie to have for dessert that evening. After dinner Kim gets up saying, "I'm baking a pie. It's in the oven; let me check if it's done."
"Mmmm," replied Scott, "That sounds good! I don't usually eat pies from shops or restaurants, but that's because I am allergic to nuts."
Kim goes to the kitchen and returns a short while later.
"The pie isn't quite done yet. Should we go out for ice-cream instead?" asks Kim as she returns from the kitchen, "We can have pie tomorrow."
"OK. Ice-cream also sounds good" says Scott.

Kim is just as happy when she comes out of the kitchen as she was just after dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Scott doesn't think that Kim made pecan nut pie.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Kim thinks that Scott thinks she made pecan nut pie.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Scott thinks that Kim is just as happy when she came out of the kitchen as she was just after dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Kim and Scott are related.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

6. Pie (NES) Test B

Kim's cousin, Scott, was coming to visit her for the weekend and Kim was making an apple pie to have for dessert that evening. After dinner Kim gets up saying, "I'm baking a pie. It's in the oven; let me check if it's done."
"Mmmm," replied Scott, "That sounds good! I don't usually eat pies from shops or restaurants, but that's because I am allergic to nuts."
Kim goes to the kitchen and returns a short while later.
"The pie isn't quite done yet. Should we go out for ice-cream instead?" asks Kim as she comes out of the kitchen, "We can have pie tomorrow."
"OK. Ice-cream also sounds good" says Scott.

Kim is just as happy when she comes out of the kitchen as she was just after dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Scott doesn't think that Kim made pecan nut pie.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Kim thinks that Scott thinks she made pecan nut pie.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Scott thinks that Kim is just as happy when she came out of the kitchen as she was just after dinner.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Kim and Scott are related.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

7. Ladies dressing room (ES) Test B

Emily, Tanya and Sam are friends who are out shopping for clothes together. Sam is looking through the dresses on the rack trying to find her size of a dress she likes. Emily and Tanya are in two adjoining dressing room compartments trying on clothes. Sam finds the dress she is looking for and decides to try it on.

"Sam would look so beautiful in that dress," Emily says to Tanya through the wall.
"I know." Says Tanya, "But I don't think they'll have it in the bigger sizes."
When they step out of their compartments they see Sam standing in the dressing room.

Sam is less happy when Emily and Tanya come out of their compartments than she was just after she decided to try the dress on.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Tanya thinks that Sam needs a big size dress.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Emily thinks that Sam thinks Tanya thinks the shop doesn't have the size Sam is looking for.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Tanya thinks that Sam is just as happy at the end as when she decides to try on her dress.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sam couldn't find a dress in her size.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

7. Ladies dressing room (NES) Test A

Emily, Tanya and Sam are friends who are out shopping for clothes together. Sam is looking through the dresses on the rack trying to find her size of a dress she likes. Emily and Tanya are in two adjoining dressing room compartments trying on clothes. Sam finds the dress she is looking for and decides to try it on.

"Sam would look so beautiful in that dress," Emily says to Tanya through the wall.
"I know." Says Tanya, "But I don't think they'll have it in the bigger sizes."
When they step out of their compartments Sam walks into the dressing room.

Sam is less happy when Emily and Tanya come out of their compartments than she was just after she decided to try the dress on.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Tanya thinks that Sam needs a big size dress.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Emily thinks that Sam thinks Tanya thinks the shop doesn't have the size Sam is looking for.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Tanya thinks that Sam is just as happy at the end as when she decides to try on her dress.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Sam couldn't find a dress in her size.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

8. Colleague (ES) Test A

Derek is a manager at an advertising company. He is meeting Tim from the IT department for lunch at a restaurant near their offices. Derek and Tim met through a mutual friend and Derek arranged Tim an interview for the IT position. They've become good friends since working in the same office. Derek says to Tim "There's someone from head office here this week. His name is Alan – I'm having a meeting with him here later. I think we'll have finished lunch by then."
"No problem." Says Tim. During lunch, Tim talks about what he's been doing at work. "Man, you are really into your work aren't you?" says Derek.
"I guess I am," says Tim, "Compared to my last job this one is great."

As they are paying the lunch bill, Derek says "Oh, here's Alan now."
"Hi Alan. This is my friend Tim" Says Derek.
"Hi Tim," says Alan as they shake hands.
Tim leaves and Alan and Derek sit down and start their meeting. Tim is about to leave the restaurant when he realises his jacket is still at the table. When he gets to the table to get his jacket, he hears Alan say, "Head office feels that there is no reason for having an internal IT department. They want you to find an IT company to hire occasionally instead of having full time IT staff."
"Sorry, I forgot my jacket" Says Tim. He gets his jacket and leaves the restaurant.

Tim feels just as happy after Alan spoke to Derek as he did when he left the restaurant.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Alan thinks that Tim works in the IT department.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Tim thinks that Alan thinks Tim works in the IT department.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Derek thinks Tim is less happy after Derek and Alan spoke than when Derek left the restaurant.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Derek and Tim hadn't met each other before they started working in the same office.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

8. Colleague (NES) Test B

Derek is a manager at an advertising company. He is meeting Tim from the accounts department for lunch at a restaurant near their offices. Derek and Tim met through a mutual friend and Derek arranged Tim an interview for the accounting position. They've become good friends since working in the same office. Derek says to Tim "There's someone from head office here this week. His name is Alan – I'm having a meeting with him here later. I think we'll have finished lunch by then."
"No problem." Says Tim. During lunch, Tim talks about what he's been doing at work. "Man, you are really into your work aren't you?" says Derek.
"I guess I am," says Tim, "Compared to my last job this one is great."

As they are paying the lunch bill, Derek says "Oh, here's Alan now."
"Hi Alan. This is my friend Tim" Says Derek.
"Hi Tim," says Alan as they shake hands.
Tim leaves and Alan and Derek sit down and start their meeting. Tim is about to leave the restaurant when he realises his jacket is still at the table. When he gets to the table to get his jacket, he hears Alan say, "Head office feels that there is no reason for having an internal IT department. They want you to find an IT company to hire occasionally instead of having full time IT staff."
"Sorry, I forgot my jacket," Says Tim. He gets his jacket and leaves the restaurant.

Tim feels just as happy after Alan spoke to Derek as he did when he left the restaurant.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Alan thinks that Tim works in the IT department.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Tim thinks that Alan thinks Tim works in the IT department.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Derek thinks Tim is less happy after Derek and Alan spoke than when Derek left the restaurant.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Derek and Tim hadn't met each other before they started working in the same office.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

9. Pensioners' Discount (ES) Test B

Debbie, a 47 year old woman, is going to the movies. On her way to the ticket counter she notices a sign saying it is pensioners' Monday and the cinema offers 5% discount for people over 60. She is about to pay when the cashier, a woman of about 30, says;
"Don't worry about showing me your pensioner's card."
"Excuse me." says Debbie.
"The pensioners discount; I've deducted it."
"Thanks," says Debbie. She pays, picks up her ticket and leaves the counter.

Debbie is less happy when she leaves the counter than before she spoke to the cashier.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

The cashier thinks Debbie is older than 60.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Debbie thinks that the cashier thinks Debbie is younger than 60.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

The cashier thinks Debbie is less happy when she leaves the counter than when she first arrived.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Debbie is older than 60.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

9. Pensioners' Discount (NES) Test A

Debbie, a 67 year old woman, is going to the movies . On her way to the ticket counter she notices a sign saying it is pensioners' Monday and the cinema offers 5% discount for people over 60. She is about to pay when the cashier, a woman of about 30, says; "Don't worry about showing me your pensioner's card."
"Excuse me." says Debbie.
"The pensioners discount; I've deducted it."
"Thanks," says Debbie. She pays, picks up her ticket and leaves the counter.

Debbie is less happy when she leaves the counter than before she spoke to the cashier.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

The cashier thinks Debbie is older than 60.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Debbie thinks that the cashier thinks Debbie is younger than 60.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

The cashier thinks Debbie is less happy when she leaves the counter than when she first arrived.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Debbie is older than 60.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

10. Office Party (ES) Test A

Bob went to the office dinner party where he joined a conversation with two other men from his office; Mr Jones and Mr Baker . It was hard for Bob to hear the conversation because of a very noisy woman in a purple dress standing behind them talking to a lady in a red Jacket.
"Oh my word," said Bob, "that woman in the purple dress has a voice like a vuvuzela."
"You're right!" agreed Mr Jones.
Bob went to the bar to get another drink. While he was waiting, he overheard another colleague saying, "Mr Baker, I still haven't met your wife,"
"Oh. She's right over there" said Mr Baker pointing across the room. "The one in the purple dress."

When Bob went to the bar Mr Baker felt just as happy as when Bob first joined the conversation.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Mr Jones thinks that the woman in the purple dress is Mrs Baker.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Mr Jones thinks that when Bob went to get a drink Mr Baker was just as happy as he was when Bob first joined the conversation.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end of the story Bob thinks that Mr Baker thinks the woman in the purple dress is Mrs Baker.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

The woman in the purple dress was Mr Baker's wife.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

10. Office Party (NES) Test B

Bob went to the office dinner party where he joined a conversation with two other men from his office; Mr Jones and Mr Baker. It was hard for Bob to hear the conversation because of very noisy woman in a purple dress standing behind them talking to a lady in a red Jacket.
"Oh my word," said Bob, "that woman in the purple dress has a voice like a vuvuzela."
"You're right!" agreed Mr Jones.
Bob went to the bar to get another drink. While he was waiting, he overheard another colleague saying, "Mr Baker, I still haven't met your wife,"
"Oh. She's right over there" said Mr Baker pointing across the room. "The one in the red Jacket."

When Bob went to the bar Mr Baker felt just as happy as when Bob first joined the conversation.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Mr Jones thinks that the woman in the purple dress is Mrs Baker.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

Mr Jones thinks that when Bob went to get a drink Mr Baker was just as happy as he was when Bob first joined the conversation.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

At the end of the story Bob thinks that Mr Baker thinks the woman in the purple dress is Mrs Baker.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**

The woman in the purple dress was Mr Baker's wife.

Correct **Near Correct** **Neither Correct nor Incorrect** **Near Incorrect** **Incorrect**