# Normative Data for Two Theory of Mind Tests in a South African Context

Marguerite Loftus

Zayaan Noordien

Department of Psychology

University of Cape Town

Supervisor: Progress Njomboro

Word Count:

Abstract: 221

Body: 9424

## ABSTRACT

Over the years, the utility of socio-cognitive tests in clinical practice and neuropsychological research has been demonstrated. Although most neuropsychological batteries still exclude these tests, social cognition deficits have been demonstrated to play an important role in predicting rehabilitation success and disease outcome. In this study we provide partial normative scores for two tests of an important aspect of social cognition; theory of mind reasoning. One of these tests is the Reality-Unknown False Belief task, which is a video-based belief reasoning test with controls for deficits in executive function, language and memory, and the other is the “Reading the mind in the eyes” test. Normative and baseline cut-off scores were obtained for both tests. Results indicated on the Reading the Mind in the Eyes test a significant main effect of race, where Whites performed better than non-Whites. A pairwise comparison also found that race served as a protective factor for Whites whose first language is not English. Findings from the Reality-Unknown False Belief task showed that three of the four control factors i.e. memory, response inhibition and one filler trial had a statistically significant influence on the test factor i.e. false belief performance. The participants performed at ceiling on the other filler. It can be concluded that further study should include larger and more equal sample sizes to confirm these results.



***Keywords*****:** normative data, social cognition, theory of mind, reality-unknown false belief task, reading the mind in the eyes test.

## Introduction

Humans are social beings that rely heavily on the social world around them for survival. This interaction is made possible through a process of social cognition (Adolphs, 2003). Social cognition is multifaceted and is comprised of different types of social skills, which allow individuals to accurately interpret and respond to social stimuli (Penn, Corrigan, Bentall, Racenstein, & Newman, 1997). One distinct process related to how humans make sense of their social world is through a process commonly referred to as Theory of Mind (ToM). ToM refers to the unique human ability to infer, explain, interpret, share and predict other people's mental states, beliefs, thoughts, feelings, and behaviour (Abu-Akel & Abushua'leh, 2004; Apperly, Samson & Humphreys, 2005; Rieffe, Terwogt, & Cowan, 2005; Stone, Baron-Cohen, Calder, Keane, & Young, 2003). Although some consider ToM as synonymous with social cognition, many regard it as a distinct process (Adophs, 2003, 2009). Its however important to note that a number of socio-cognitive processes such as those involving recognising emotional states from faces, or moral reasoning, significantly contribute to ToM processing (Sabbagh, 2004; Baird & Astington, 2004), hence the talk of a "social brain" in reference to the neural correlates of social cognition (Adolphs, 2009).

Recently, the utility of including socio-cognitive tests in neuropsychological test batteries has been noted (Crawford, Garthwhaite, & Betkowska, 2009; Lee, Farrow, Spence, & Woodruff, 2004). ToM is an important socio-cognitive measure to assess, as ToM impairments are often associated with neurological change or acquired brain damage, such as that following acquired brain injury from road accidents, gunshot wounds, accidental falls as well as infections like HIV/AIDS infection. South African population-based studies have shown that there is a high prevalence of acquired traumatic brain injury (Bruns & Hauser, 2003) and HIV infection (Welz et al., 2007). Thus research in ToM is particularly important in South Africa.

In addition, it is important to identify socio-cognitive deficits, particularly ToM deficits in these patients as these are associated with higher medical costs and significant caregiver burden (Cummings, 1997; Seltzer, Vasterling, Yoder, & Thompson, 1997; O'Shea, 2003). Normative data for most socio-cognitive tests is lacking and the tests are also not standardised. This underscores the need for work in this area if valid and reliable socio-cognitive measures are to be incorporated into neuropsychological batteries. The collection of normative data, in common socio-cognitive tasks is one important step in that direction, as it allows researchers to compare the performance of healthy controls to patients with clinical disorders (Crawford, Garthwaite, & Slick, 2009). Such normative data sets are particularly useful in the South African context, where neuropsychology is just beginning to be appreciated as a distinct clinical discipline in its own right.

## Background

### Historical Overview of Theory of Mind

Developmental psychology has contributed significantly to ToM research (e.g., Wellman, Cross, & Watson, 2001). Consequently, most of the studies have been done on infants and children. It is understood that the basic abilities of ToM are present in infants by 15 months of age when they are able to identify pretend play (Onishi, Baillargeon, & Leslie, 2007). The ability to recognise first-order false beliefs is usually developed between the ages of 3 and 4 years (e.g., Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007, Stone, Baron-Cohen, & Knight, 1998; Wellman et al., 2001). First-order false beliefs (e.g. Tom thinks x, when it is actually y) refers to one's ability to recognize that other people can have a false belief about the world that is different from our own (Abu-Akel & Abushua'leh, 2004). Second-order false beliefs (e.g. Tom thinks Mary thinks x, but actually Mary thinks y, however the correct answer is z and consequently both of them are wrong) develops between the ages of 6 and 7 years, where children grasp that other people also possess ToM abilities (Turkstra, 2008). One therefore has the ability to identify that one can have a false belief about a false belief (Apperly, Samson, Carroll, Hussain, & Humphreys, 2006). Lastly, the recognition of a *faux pas* situation, which develops between the ages of 9 and 11 years, refers to one's ability to identify situations when someone says something inappropriate, without the awareness that what they are saying is inappropriate (Stone et al., 1998). *Faux pas* situations are related to ToM since it involves the feelings and emotions that are associated with social interaction (Adolphs, 2003). Evidence also suggests that the development of ToM capacities involves multiple cognitive skills such as language and memory (Silliman et al., 2003). Furthermore, ToM continues to develop and mature throughout late childhood, and even after reaching adolescence (Homer et al., 2008). Although, ToM research has mostly been done on children it has been found that it is also important to consider ToM in an adult population, since it has been noted that brain damage in adults can impair ToM (Channon & Crawford, 2000; Bibby & McDonald, 2005).

**Theory of Mind and Neurological Disorders**

Consistent studies have shown that damage to certain brain areas in adults particularly the prefrontal cortex impairs ToM functioning (Happé, Malhi, & Checkley, 2001; Stuss, Gallop, & Alexander, 2001). Neuroimaging studies have reported that either the medial (Fletcher, Happé, & Frith, 1995 Gallagher et al., 2000) or the orbito prefrontal cortex (Baron-Cohen et al., 1994) is involved in ToM reasoning.

Neurological damage to the prefrontal cortex have been reported in many psychiatric and neurological illnesses such as schizophrenia (Lee et al., 2004), autism (Barnea-Goraly et al., 2004) and traumatic brain injury (Milders, Fuchs, & Crawford, 2003). These deficits are associated with impaired social functioning, including communication deficits, unemployment and low level of community functioning (Couture, Penn, & Roberts, 2006). Impairments in social behaviour are therefore a feature of neurological change or damage and socio-cognitive tests may help in identifying some of these deficits.

Rehabilitation programmes which target social cognition have been shown to be good predictors of rehabilitation success and disease outcome in patients with socio-cognitive deficits (Pijnenborg et al., 2009). For instance, the Social Cognition and Interaction Training for Individuals with Schizophrenia (Kee et al., 2008; Combs et al., 2007) and the Community-based Psycho-social Treatment for Schizophrenia (Brekke, Long, Nesbit, & Sobel, 1997; Brekke, Hoe,Long, & Green, 2007) have been shown to improve disease outcome, reduce the likelihood of relapse and hospitalization, and provide a better prognosis and quality of life (Brekke et al., 2007; Sergi, Rossovsky, Nuechterlain, & Green, 2006; Zucker et al., 2007). This demonstrates that programmes aimed at improving social cognition are an important part of the rehabilitation process.

Considering the clinical and rehabilitation implications of an impaired social cognition it is important that valid socio-cognitive measures are developed. However there are only a few, if any, neuropsychological batteries that include assessments of socio-cognitive domains. Most traditional neuropsychological batteries such as the Repeatable Battery for the Assessment of Neuropsychological status (Randolph, Tierney, Mohr, & Chase, 1998; Hobart, Goldberg, Bartko, & Gold, et al., 1999), the Luria Nebraska Neuropsychological Battery (Golden et al., 1982) and Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) are all examples of traditional neuropsychological batteries that tend to focus on distinctively six cognitive areas involving attention, concentration, language, memory, visio-spatial and executive functioning (Roos et al., 2010; Baune, McAfoose, Leach, Quirk, & Mitchell, 2009).

### Theory of Mind Tests

There are a wide range of tasks available that test ToM, although most are for use on young children or mentally handicapped adults (e.g., Wimmer & Perner, 1983; Baron-Cohen, Leslie, & Frith, 1985). The most commonly used are story-based ToM tests (e.g. Channon & Crawford, 2000) and those assessing ToM through cartoon-stimuli and facial recognition (e.g. Adolphs, 2003). However, only a few tests are available that is sensitive enough to measure mild social impairments in adults with otherwise normal intelligence.

It has been reported that language has an influence on ToM performance (Figueras-Costa & Harris, 2001; Hale & Tager-Flusberg, 2003; Milligan, Astington, & Dack, 2007). Language processing becomes a problem in cases of neurological change such as that resulting from traumatic brain injury (Kendall & Terry, 1996; Levin, 1995), which then may have an effect on ToM performance. However, due to a lack of adult ToM task studies the relationship between ToM and language in an adult population are inconclusive. Studies have however, shown that there are significant gender differences on ToM test performance, where adult females score significantly higher than male adults (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). It has been suggested that females may perform better on ToM tasks because they are superior to males with regard to empathy and emotional sensitivity (Baron-Cohen & Jolliffe, Mortimore, & Robertson, 1997).

It has been suggested that language may play a critical role in the development of ToM reasoning, but with age ToM capacity is not dependent on language (Apperly et al., 2006). Studies have also shown that poor performance on ToM tasks are associated with executive function deficits (Rowe, Bullock, Polkey, & Morris, 2001) especially those related to inhibition of self-perspective and working memory (Carlson, Moses, & Breton, 2002; Corcoran & Frith, 2003). These processes are independent of ToM deficits, but are recruited in most ToM tasks, e.g. Strange Stories task (Happé, 1994), *Faux Pas* task (Baron-Cohen et al., 1999). Thus, most of these tasks do not take these ancillary processes into account. Since these ToM tests are usually used on neurological patients with a high likelihood of deficits in these ancillary processes, it is crucial to use tests that control for these factors. For instance, Apperly et al. (2006) has developed a non-verbal false belief test that controls for executive processing, language, comprehension, and memory. Such a task can be particularly useful on patients who may have severe executive function, language and memory impairments (Apperly, Samson, Chiavarino, & Humphreys, 2004).

**Theory of Mind and Culture.**

There is conflicting results regarding the role that culture plays in ToM. Some studies have shown that children cross-culturally develop the skill for false-belief reasoning at around the same age (Callaghan et al., 2005), while other studies suggest age of ToM development can vary between similar industrialised countries such as Japan and Korea (Oh & Lewis, 2008) and the United Kingdom and Canada (Wellman et al., 2001).

However, confirming the relationship between ToM and culture in an adult population is limited by small sample sizes and predominantly Anglo-European individual studies (e.g., Liu, Wellman, Tardiff, & Sabbagh, 2008; Shahaeian, Peterson, Slaughter, & Wellman, 2011).

Cross cultural studies on ToM have shown culture-specific effects, particularly ToM tasks that assess facial affect recognition (e.g., Adams et al., 2009; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005). These studies have shown that the ability to infer emotions from the faces of others is superior in same-race faces compared to inferring emotions from other-race faces, where this ability is established during childhood (Sangrigoli et al., 2005). For example, evidence shows that the revised version of the "Reading the Mind in the Eyes" (RME) (Baron-Cohen et al., 2001) suffers from race effects (Paladino et al., 2002). Furthermore, there exists an intercultural advantage in ToM interpretation on the test. For instance, it is easier to decode the mental state of those of the same culture compared to those from other cultures. These findings are supported by both behavioural as well as neural evidence in a sample of white American and native Japanese participants (Adams et al., 2009).In the Adams et al. (2009) study, participants completed the RME in which they had to infer emotional states from expression shown in the eyes. It was found that participants performed better when identifying the emotion of same race individuals and that brain regions, particularly the fusiform area show a superior response to same-race compared to other-race faces (Golby, Gabrielli, Chiao, & Eberhardt, 2001; Herrmann et al., 2006). Furthermore, research has also shown that same race faces are perceived more holistically and better remembered compared to those of other races (Meissner & Brigham, 2001; Michel, Rossion, Han, Chung, & Caldara, 2006; Tanaka, Kiefer, & Bukach, 2004). It was found that the recruitment of neural activity differed in relation to cultural group membership as brain regions, particularly, the fusiform region show a superior response to same-race compared to other-race faces (Golby et al., 2001; Adams et al., 2009). Consequently, these findings have significant implications on the validity and reliability of ToM tests, particularly those tests that require inferring mental states from facial features when they are used in different race contexts.

## Normative data in South Africa

Lately there have been efforts to expand cross-cultural normative databases for psychometric tests in non-Western countries (Boone, Victor, Wen, Razani, & Pontón, 2007; Uzzel, Ponton, & Ardila, 2007). However, despite the evident demand for and advantages of such locally normed tests, there is still a significant lack of normative data on psychological tests in African countries (Ruffieux et al., 2009). Normative data is crucial for the progression and development of neuropsychological research in these contexts. Furthermore, this data is necessary to identify the multiple local risk factors, such as malnutrition or disease, which might have an impact on the neuropsychological development of the people in sub-Saharan Africa. With so few normative data available it is impractical to use these westernized psychometric tests in local practice. It is therefore important to consider the socio-demographic factors that affect neuropsychological competence, ability, or capacity and to adjust these tests accordingly (Rosselli & Ardila, 2003).

South Africa is marked by cultural diversity, where neuropsychological studies have found that differences in factors such as socio-economic status, education and language have a significant influence on test performance (Roos et al., 2010; Skuy, Schutte, Fridjhon, & O'Carrol, 2001). For instance, Skuy et al. (2001) investigated the validity and reliability of using published neuropsychological test norms in South African samples. They found that urban African high school students performed significantly poorer on a neuropsychological test battery compared to an American control group. However, since the tests were constructed in America and only available in English, these tests may not be suitable for the majority of South Africans, where English is not the first language for many. Since South Africa is a linguistically diverse country it is therefore important to control for language. These findings underscore the need for using tests that are normed for use with the population concerned when obtaining neuropsychological data.

## The Reality Unknown and Reading the Mind in Eyes tests

It is suggested that there are two key component processes of ToM, namely social-perceptual processes and socio-cognitive processes. Social-perceptual processes are involved in processing the nonverbal stimuli that make it possible to infer mental states from, for example, the eyes of others. Socio-cognitive processes allows for more abstract reasoning, such as the ability to recognize the false beliefs of others (Sabbagh, 2004). In this study we therefore considered both components of ToM by applying a social-perceptual test known as the Reading the Mind in the Eyes test (RME) as well as a socio-cognitive test known as the Reality-Unknown False Belief task [(RU) see full discussion of these tests under the methods section]. We collected normative data on these two tests. Most ToM tasks have no control over other additional cognitive processes that may not be core to ToM, but are recruited in the performance of ToM tasks. These additional processes may include language, executive function and working memory. Thus, it is imperative that ToM tests control for these. The RU controls the influence for most these additional factors, and therefore provides a more accurate measure for assessing ToM performance.

**Rationale for Research**

The study is relevant because it provides normative data on ToM tests for possible clinical use in the South African context. The literature reviewed shows that there is a lack of available normative data for socio-cognitive tests in general, and especially on ToM. Furthermore, most ToM studies have been done on child populations. However, it has been found that it is also important to consider ToM in an adult population, since ToM can be impaired in neurological damage in adults, which are associated with a wide range of social and socio-cognitive deficits (McDonald & Flanagan, 2004). This neurological damage is often found in individuals who have sustained traumatic brain injury or have brain degenerative diseases or infections, which are all common in a South African context (Bruns & Hauser, 2003).Since most neurocognitive tests are designed in western contexts it is crucial that comparative data is available that makes these tests applicable to a South African context. The collection of normative data is therefore relevant and useful for use in socio-cognitive research in a clinical population.

**Specific Aims/Hypotheses**

This study aims to collect normative data suitable for a South African context. In particular we aim to determine:

- Whether, first language English-speaking students will perform significantly better than students with other linguistic backgrounds on the RME, as first language English speakers would be more proficient and familiar with the English terms used in the test. This provides a justification for grouping together all non-Whites as well as all participants that did not have English as their first language into one category

- Whether, Whites will perform better than non-Whites in line with results from face perception studies (Adams et al., 2009; Golby et al., 2001; Herrmann et al., 2006).
- Whether, females will perform significantly better than males on the RME in line with results from another study (Baron-Cohen et al., 2001).
- Whether, the western cut off norms for the RME test will be higher than that of the South African population.
- Participants will perform at ceiling on the RU test due to the simplified nature of the test and its limited semantic loading.

## Methods

### Research Design and Setting

The study focused on three main variables: gender, race and language. The language variable consisted of two levels; those who are first language English speakers, and those who have a first language other than English. The gender variable had two levels; those who are male and those who are female. The race variable included those who are non-White and those who are White. In South Africa, the term “black” refers to all individuals who are not White (Rushton & Skuy, 2000). This includes Coloureds, Indians, Asian and Black Africans. Therefore, this study consisted of individuals who belong to either the White or non-White (“black”) category. This research is in the exploratory phase. Data collection took place in the Department of Psychology at the University of Cape Town or at the participant’s home. All participants were tested in a quiet room in a single one hour session.

### Participants

Normal adults (*N* = 56, 11 male, 45 female) aged 18 to 32 (mean = 21.14, *SD*= 2. 81) who were all university students were recruited (see Table 1 for demographics of participants). Participants were recruited using the University of Cape Town’s Student Research Participant Programme (SRPP) (*N*=43) along with other individuals (*N*=13) who met the inclusion criteria. The participant noted this information before completing the test. The participants were from different education levels (44 Undergraduates, 12 Postgraduates).

**Table 1**

*Race and Language Breakdown of Participants*

| | White | Non-White | | | | |
|---|---|---|---|---|---|---|
| **Language** | | Black African | Coloured | Indian | Asian | **Total** |
| English | 22 | 6 | 5 | 3 | | **36** |
| Afrikaans | 9 | | | | | **9** |
| Xhosa | | 3 | | | | **3** |
| Zulu | | 4 | | | | **4** |
| Tsonga | | 1 | | | | **1** |
| SeSotho | | 1 | | | | **1** |
| Korean | | | | | 1 | **1** |
| Malayalan | | | | 1 | | **1** |
| **Total** | **31** | **15** | **5** | **4** | **1** | |

This study followed the ethical guidelines for research done on human participants as outlined by the Health Profession Council of South Africa (HPCSA) and the University of Cape Town (UCT) Codes for Research. Participants provided informed consent by signing a form, which explained the test they were to participate in, which ensured anonymity and confidentiality of all results and also informed them that the participation in the study was voluntary and they could withdraw at any point (see Appendix A).

### Inclusion and Exclusion Criteria

All participants had to be able to speak and understand English fairly well as the material for both tests required an understanding of English terms. Participants with a history of psychiatric or neurological impairment were excluded from the study.

## Measures

**Reality-unknown false belief task.** The non-verbal video-based false belief task used in this study is a variation of that developed by Call and Tomasello (1999). It was adapted by Apperly, Samson, & Humphreys (2009) for use on patients with acquired brain damage.

The RU consists of 4 blocks, each of which contains 15 items. In total, there are 60 items which consist of 12 false belief, 12 memory and 12 inhibition control video clips, as well as 24 filler-trials. The test is divided into 4 control factors: two Anti-Strategy Filler

Trials, the Confirmation Filler Trial and True Belief Filler Trial, Memory and Response Inhibition Trials, and 1 test factor, False Belief Trial.

Participants watched short video clips where a woman gives a visual hint to where the hidden object is, by placing a placing a pink card onto one of the two containers. The rules of the test were explained to each participant before staring the test session. The test started with a practise trial to ensure that the participant understood the testing procedure.

In the false belief trial (see Figure 1 below), the participant observes a man letting a woman look inside two containers, but the participant does not see in which container the object is located. The participant then sees the woman leave the room, and in her absence the man then switches the location of the two containers. Thus, the woman will have a false belief about where the object is placed. When the woman comes back she gives a hint to the participant by pointing where she (incorrectly) thinks the object is placed. The video clip is then paused and the participant is then prompted to indicate the container containing the object. To correctly locate the object, the participant had to understand that the woman has a false belief about the location of the object and thus, she pointed to the incorrect location. The participants then decided where they thought the object was placed, and then were given feedback by watching the end of the video clip where the man opens both containers and shows the contents to the camera. The False Belief Trials involved the participant processing the order of the happenings in the video clip; specifically the woman gives a hint after the containers were swapped.

| 1. The woman looks in the containers. | 2. She leaves the room, the man swops the containers. | 3. She indicates where she thinks the object is with a pink card. | 4. The participant must indicate where they think the object's located |
|---|---|---|---|

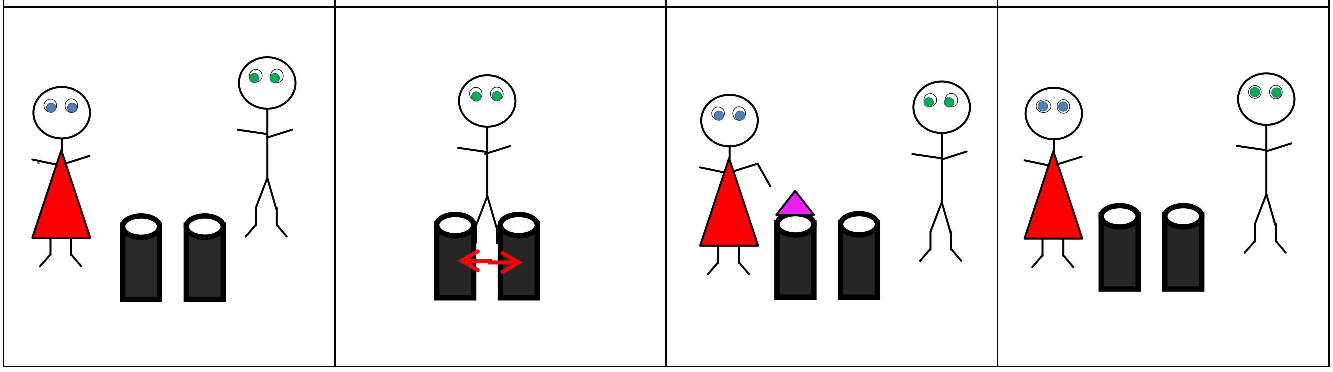


**Figure 1**. False Belief task

To control for this processing of order of events, the working memory controls trials (see Figure 2 below) reversed the order of the hint-giving by woman and container-swapping by the man. Thus, the woman indicates a container before she leaves the room, which allows for the participant to infer where the object is placed. Whilst the woman is absent from the room, the man swaps the location of the two containers without showing the contents of the containers. The woman returns, but does not do anything and simply remains seated. Therefore, the participant had to use the clue that the containers had been swapped to update his/her memory of the container where the object was located, and had to retain this information until a response was required. The False Belief trials also involved the participant to separate his/her attention from the container where the woman had just indicated, and instead point to the other container.

| 1. The woman looks in the containers. | 2. She indicates where she thinks the object is with a pink card. | 3. She leaves the room, the man swops the containers. | 4. The participant must indicate where they think the object's located |
|---|---|---|---|

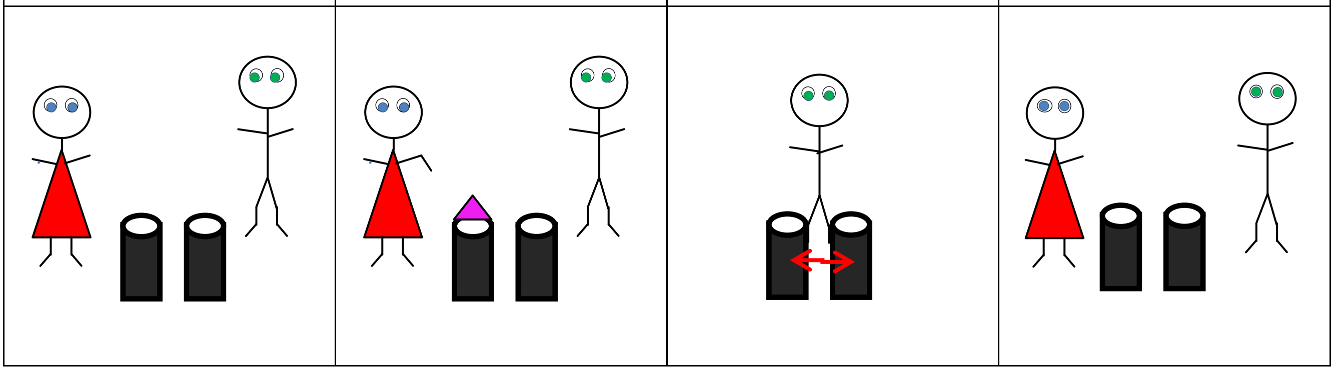


**Figure 2.** Memory Control

A participant who lacked the inhibition control (see Figure 3 below) to separate his/her attention from the incorrect location would get these trials incorrect, regardless of whether they could infer the false belief. In the Inhibition Control Trials, the woman leaves the room and in her absence, the man visibly moves the object from one container to the other. When the woman returns (unsuspectingly) she points to the container that the participant now knows is empty. The participant was then requested to indicate the container that contained the object. Like in the case of the False Belief Trials, to get the correct location of the object the participant had to separate his/her attention from the container, which was pointed at by the woman. However, in these trials, the participant did not need false belief reasoning.

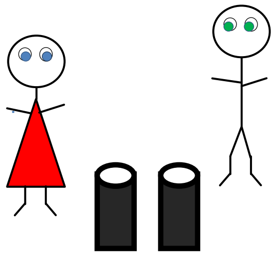
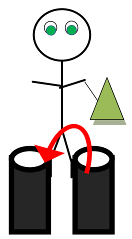
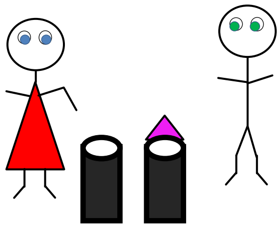
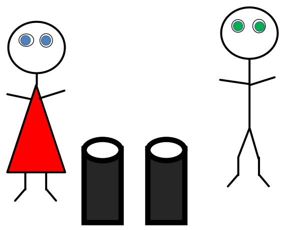
| 1. The woman looks in the containers. | 2. She leaves the room; the man visibly removes the object. | 3. She indicates where she thinks the object is with a pink card. | 4. The participant must indicate where they think the object's located |
|---|---|---|---|
|  |  |  |  |

**Figure 3.** Inhibition Control

On the Confirmation Filler Trials (see Figure 4 below) the woman indicates a container before she leaves the room. In her absence, the man then opens the container to show the object, which is a noticeable reminder that the woman has indicated the object's location in good faith. The man then moves object to the other container. The woman then returns. The participant was then requested to give a response.

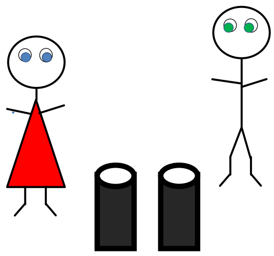
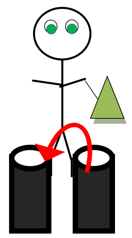
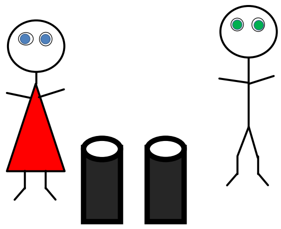
| 1. The woman looks in the containers. | 2. She leaves the room; the man visibly removes the object. | 4. The participant must indicate where they think the object's located |
|---|---|---|
|  |  |  |

Figure 4. Confirmation Filler Trial

True Belief Filler Trials (see Figure 5 below) were created to protect against the participants passing False Belief Trials by adopting an artificial strategy to solve the task i.e. pointing to the opposite container that was indicated by the woman in the video. The woman leaves the room, but in her absence the man does not swap the containers. This means that the woman does not have a false belief about the location of the object and points to the correct container where the object is placed. For the participant to answer correctly, the participant had to indicate the same container that the woman pointed at. Even though it was possible that the participant simply inferred the belief of the woman, this was not a reliable indicator of belief-reasoning as the participant can also get this trial correct by pointing to wherever the woman indicates, without making inferences about the woman's belief. A crucial point in this task is that the correct responses to the False Belief Fillers needed the participant to indicate the same location to the one that the woman indicated, whereas correct answers to the true false belief trials needed the participant to point to the opposite location to where the woman indicated. Therefore, if participants performed well on the False Belief Fillers, then it would be ensured that the good performance on the False Belief Trials showed genuine belief reasoning.

| 1. The woman looks in the containers. | 2. She leaves the room, the man lifts the containers. | 3. She indicates where she thinks the object is with a pink card. | 4. The participant must indicate where they think the object's located |
|---|---|---|---|

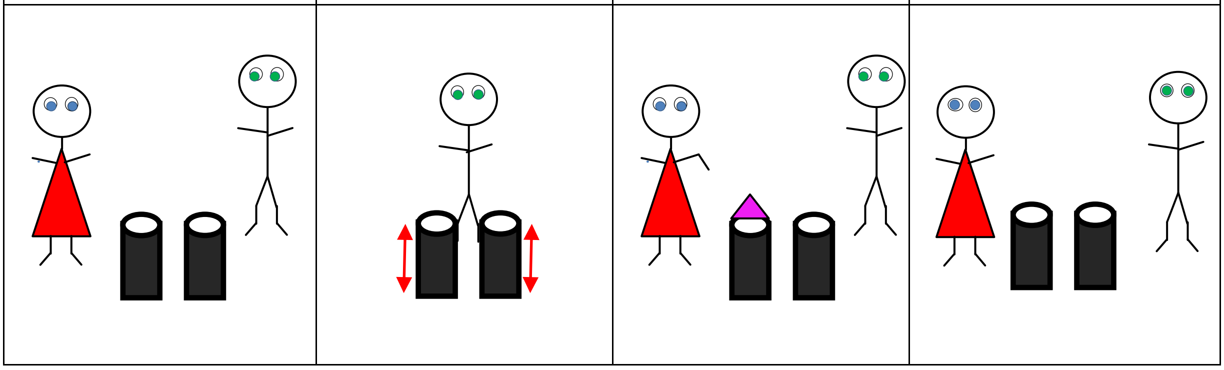


**Figure 5.** True Belief Filler Trial

**Reading the mind in the eyes test.** The revised version of RME (Baron-Cohen et al., 2001) is a refined measurement of adult ToM ability. Although the RME is considered an advanced ToM test, it focuses and assesses only one aspect of ToM, which is the attribution of appropriate mental states from facial cues. The RME assesses the ability to infer another person's thinking or feeling from the expression around on the eyes. The expression of the eye region has been shown to be informative and essential for effective social interactions (Adams et al., 2009). Each pair of eyes depicts a specific emotion and the participant must decide between four words which one best describes what the person in the photo is feeling. Problems with the original version of the test (Baron-Cohen et al., 1997) were that participants had to choose from only two words. Consequently, the test was less robust to identify individual differences and normal adults usually scored close to the ceiling of the test

(Baron-Cohen et al., 2001). The revised version was therefore modified to include three decoy items in each trial, as well as increasing the number of items in the test (Baron-Cohen et al., 2001).

**Procedure**

The four blocks of the RU test were administered in one session. The duration of this session was approximately 60 minutes. The participants were informed that the test was non-verbal and were read the following instructions: "*You will see a man and a woman on a video. The man will hide a little green object in one of two boxes. The woman's job is to help you find in which box the green object is*". The participants were then required to watch a short practise trial in order to ensure that they understood the task. Two practice items were given. If the participant responded correctly on both, the test was started; if not the practice trials were presented again, until the participant responded to the two items correctly. Participants were required to watch another series of short video clips for each of the blocks. On each clip the participant observed the sequence of events, and the video clip was paused. Participants then indicated in which container they thought the object was. If they were unable to respond, the video clip was replayed until they were able to do so.

***Scoring.*** After each video clip, the researcher scored the participant's answer on a scoring sheet (see Appendix B). The participant received a score out of 12 for each of the factors: False Belief Trial, Memory Control Trials and Response Inhibition Trial and the Confirmation and True Belief Filler Trials.

The RME was administered on the same day. Each of the participants was handed an instruction sheet to read through. The instruction sheet read: *"For each set of eyes, choose which word best describes what the person in the picture is thinking or feeling. You may feel that more than one word is applicable but please choose just one word, the word which you consider to be most suitable. Before making your choice, make sure that you have read all 4 words. You should try to do the task as quickly as possible but you will not be timed. If you really don't know what a word means you can look it up in the definition hand* out".

Thereafter, the participants were asked to read through the glossary of the words included in the test. They were encouraged to read through the meanings of the words they were uncertain of. Furthermore, the participants were informed that they could refer back to the glossary at any time during the test. A practice trial of the test was also conducted to familiarize the participants with the test. Subsequently the test was administered.

***Scoring.*** After each photograph, the researcher scored the participant's response on a scoring sheet (see Appendix C). The test consists of 36 photographs of different pairs of eyes so the participant received a score out of 36.

**Data Analysis**

The main aim of this study was to provide norms for use with clinical samples. To that end cut off scores were established by taking scores two standard deviations below the control's mean as showing probable impaired ToM for participants of same age.

The three main independent variables were identified namely gender, race and first language as predictors of ToM performance. Firstly, the relationship between gender and ToM in the RME test was investigated using an independent sample *t* test. The variables race and first language in the RME test were combined in a two-way Analysis of Variance (ANOVA). Both factors were used in one statistical test as research shows that language and race are related and thus may have a combined influence on ToM performance (Adams et al., 2009). Ideally, all three variables (gender, language and race) should have been used in a three-way ANOVA but this was not possible due to a small sample an unequal distribution of participants across groups. Assumptions of normality were violated for some scores and this also reduced statistical power. In addition, RME performance across items was represented in a graph to identify which items were unsuitable or difficult for this sample (see Figure 1).

For the RU task, a multiple regression was conducted to investigate whether the control factors : Memory Control, Response Inhibition, Confirmation Filler Trial and True Belief Filler Trial had an influence on the dependent variable, that is false belief performance. It was found that the Memory Control, Response Inhibition, True Belief Filler Trial did have a significant influence on ToM performance. Subsequently a one-way Analysis of Covariance (ANCOVA) was conducted to partial out the influence of these factors on the false belief performance (Field, 2009). One of the assumptions of ANCOVA is homogeneity of regression of slopes of the covariate and the dependant variable. Testing this assumptions involved working out whether or not there is an interaction between the covariate and the dependant variable, for each covariate and then for each independent variable. There were was no interaction between any of the covariates for either two independent variables and thus we could proceed with the ANCOVA analysis. A correlation was also performed to investigate the relationship in ToM performance on the RU and the RME. We used an alpha level of 0.05 as a threshold for all the statistical analyses in this study. The statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS) version 17.0 (SPSS Inc., 2008).

**Table 2**

*Scores for Reading the Mind in the Eyes test and the Reality-Unknown False Belief task*

| Group | | | | *RME (36)* | | | *RU Total (60)* | | | *RU False belief (12)* | | | *RU Confirmation Filler Trial (12)* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Race** | **Gender** | **First Language** | *N* | *M* | *SD* | ***Cut-off*** | *M* | *SD* | ***Cut-off*** | *M* | *SD* | ***Cut-off*** | *M* | *SD* | ***Cut-off*** |
| White | Female | English | 18 | 28.83 | 3.37 | ≤ **22.09** | 55.17 | 5.26 | ≤ **44.68** | 9.28 | 3.05 | ≤ **3.18** | 12.00 | .00 | ≤ **12.00** |
| | | Non-English | 6* | 29.67 | 4.46 | ≤ **20.75** | 58.17 | 2.14 | ≤ **53.89** | 11.00 | 1.27 | ≤ **8.46** | 11.83 | .41 | ≤ **11.01** |
| | Male | English | 4* | 30.00 | 1.83 | ≤ **26.34** | 55.75 | 5.97 | ≤ **43.91** | 10.25 | 1.71 | ≤ **6.83** | 12.00 | .00 | ≤ **12.00** |
| | | Non-English | 3* | 27.67 | 3.79 | ≤ **20.09** | 57.33 | 4.62 | ≤ **48.09** | 10.67 | 2.31 | ≤ **6.05** | 12.00 | .00 | ≤ **12.00** |
| Non-White | Female | English | 12 | 27.00 | 3.69 | ≤ **19.62** | 50.67 | 5.03 | ≤ **40.61** | 7.67 | 2.90 | ≤ **1.87** | 11.75 | .62 | ≤ **10.51** |
| | | Non-English | 8* | 24.63 | 5.10 | ≤ **14.43** | 50.75 | 6.92 | ≤ **36.91** | 7.75 | 3.77 | ≤ **.21** | 11.38 | 1.06 | ≤ **9.26** |
| | Male | English | 3* | 25.33 | 3.06 | ≤ **22.21** | 58.33 | 2.08 | ≤ **54.17** | 11.33 | .58 | ≤ **10.17** | 12.00 | .00 | ≤ **12.00** |
| | | Non-English | 2* | 25.00 | 7.07 | ≤ **10.86** | 54.50 | 7.78 | ≤ **38.94** | 9.50 | 3.54 | ≤ **2.42** | 12.00 | .00 | ≤ **12.00** |

| Group | | | | *RU Memory (12)* | | | *RU Response Inhibition (12)* | | | *RU True Belief Filler Trial (12)* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Race** | **Gender** | **First Language** | *N* | *M* | *SD* | ***Cut-off*** | *M* | *SD* | ***Cut-off*** | *M* | *SD* | ***Cut-off*** |
| White | Female | English | 18 | 11.44 | 1.10 | ≤ **9.24** | 12.00 | .00 | ≤ **12.00** | 11.00 | 1.33 | ≤ **8.34** |
| | | Non-English | 6* | 12.00 | 0.00 | ≤ **12.00** | 11.83 | .41 | ≤ **11.01** | 11.50 | .84 | ≤ **9.82** |
| | Male | English | 4* | 11.50 | 1.00 | ≤ **9.50** | 12.00 | .00 | ≤ **12.00** | 10.00 | 3.37 | ≤ **3.26** |
| | | Non-English | 3* | 11.67 | .58 | ≤ **10.51** | 12.00 | .00 | ≤ **12.00** | 11.00 | 1.73 | ≤ **7.54** |
| Non-White | Female | English | 12 | 11.08 | .90 | ≤ **9.28** | 11.92 | .29 | ≤ **11.34** | 9.05 | 1.98 | ≤ **5.09** |
| | | Non-English | 8* | 11.00 | 1.85 | ≤ **7.30** | 11.63 | .52 | ≤ **10.59** | 9.25 | 2.49 | ≤ **4.27** |
| | Male | English | 3* | 12.00 | .00 | ≤ **12.00** | 12.00 | .00 | ≤ **12.00** | 11.00 | 1.73 | ≤ **7.54** |
| | | Non-English | 2* | 11.50 | .71 | ≤ **10.08** | 12.00 | .00 | ≤ **12.00** | 9.50 | 3.54 | ≤ **2.42** |

*Results could not be interpreted due to small sample sizes and lack of statistical power

**Reading the Mind in the Eyes Test**

**Gender.**

A one-tailed *t* test for independent samples indicated that the RME test performance for females was (*M* = 27.68, *SD* = 3.59) and for males (*M* = 28.17, *SD* = 3.59). This difference was not statistically significant, *t* (54) = - .366, *p* = .358; however, it represents a moderate effect size *r* = .45. The assumption of homogeneity of variance was met since Levene's Test of Equality of Error Variance was not statistically significant, *F* (54) = 1.436, *p* = .236. This shows that there was no significant gender difference in performance on the RME test.

**Race and Language.**

A two-way ANOVA compared the RME scores of first language English speakers and non-first language English speakers as well as White and non-White participants (see Table 2). The assumption of homogeneity of variance was met since Levene's Test of Equality of Error Variance was not statistically significant, *F* (3, 52) = 2.298, *p* = .088.

There was a statistically significant main effect of race, *F* (1, 52) = 7.947, *p* = .007. White participants (*M* = 29.03, *SD* = 3.38) scored significantly higher than non-White participants (*M* = 26.24, SD = 4.30).

There was no statistically significant main effect for first language, *F* (1, 52) = 1.468, *p* = .231. Scores for participants whose first language was English (*M* = 28.32, *SD* = 3.37) were not significantly different of participants whose first language was not English (*M* = 26.74, *SD* = 5.03).

The interaction effect between race and first language was also not statistically significant, *F* (1, 52) = 1.367, *p* = .248, indicating that race and first language did not have a combined effect on the RME test performance.

**Table 2**

*Analysis of Variance of RME for Race and First Language*

| *Source* | *SS* | *df* | *MS* | *F* | *p** |
|---|---|---|---|---|---|
| Race | 144.319 | 1 | 144.319 | 7.947 | .007 |
| First Language | 21.109 | 1 | 21.109 | 1.468 | .231 |
| First Language*Race | 19.665 | 1 | 19.665 | 1.367 | .248 |
| Error | 747.988 | 52 | 14.384 | | |
| Total | 895.429 | 55 | | | |

*α=.05

Although the interaction effect between race and language was not statistically significant, the profile plots (see Figure 6 and Figure 7) suggested that there might possibly be underlying interactional effects. Due to the small sample size and lack of statistical power it could have made it harder to detect a significant interaction between first language and race. Pairwise comparisons were therefore conducted to investigate these potential interactions (see Table 3). It was found that there was a statistically significant race difference for participants with a non-English first language preference (*p* = .017). Non-English first language speaking White participants (*M* = 29.00, *SD* = 4.12) performed significantly better on the RME test than non-English first languages non-Whites (*M* = 24.70, *SD* = 5.08). There were no other statistically significant interaction effects.

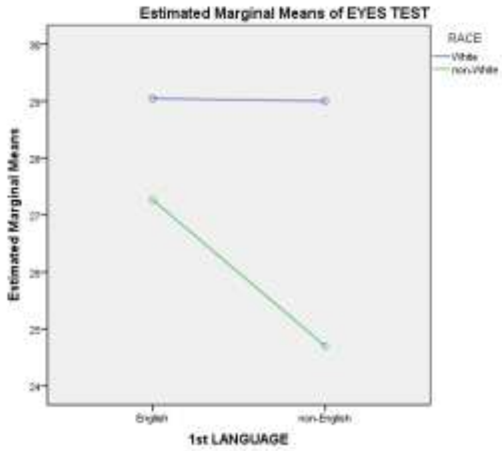



**Figure 6.**
Profile Plot for First Language by Race

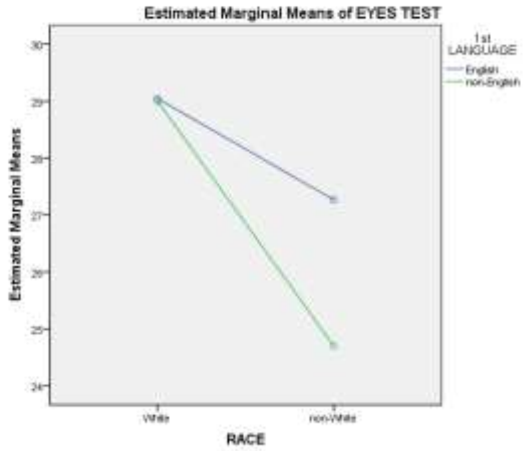



**Figure 7.**
Profile Plot for Race by First Language

**Table 3**

*Pairwise Comparisons of Race and First Language*

| | | *M* | *SD* | *p** |
|---|---|---|---|---|
| White | English | 29.05 | 3.13 | .976 |
| | Non-English | 29.00 | 4.12 | |
| Non-White | English | 27.27 | 3.52 | .103 |
| | Non-English | 24.70 | 5.08 | |
| English | White | 29.05 | 3.13 | .167 |
| | Non-White | 27.27 | 3.52 | |
| Non-English | White | 29.00 | 4.12 | .017 |
| | Non-White | 24.70 | 5.08 | |

*α=.05

Figure 8 represents the percentage of participants that got each item on the test correct. This graph showed that there were two items (item 17 and 23) that the participants performed poorly on. For the item 23, with the correct answer "defiant", only 57.4% responded correctly to this item. For the item 17, with the correct answer "doubtful", only 50 % of the participants answered this item correctly.

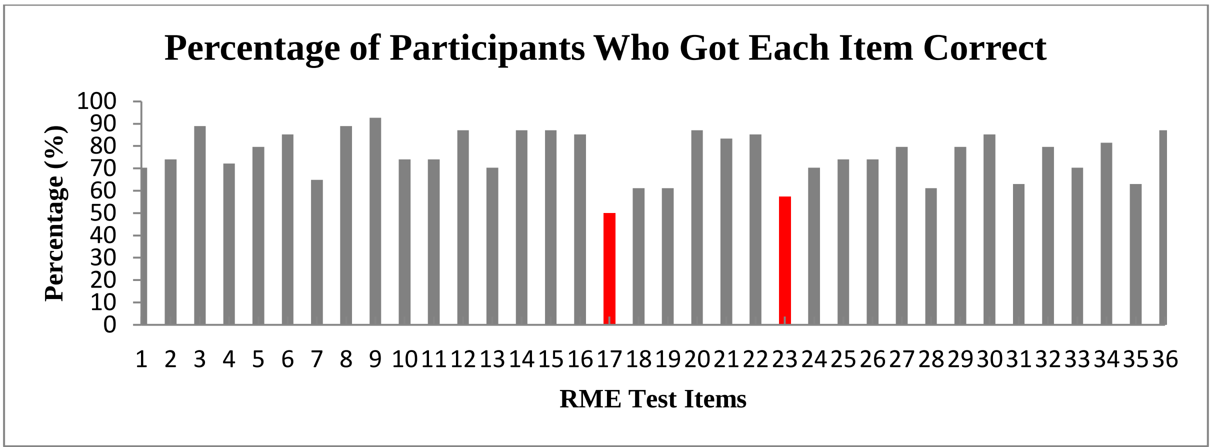



**Figure 8.** Percentage of participants who got each item correct.

## Reality-Unknown False belief Task

### Descriptive Statistics.

Descriptive statistics were conducted to screen to identify weather there were outliers. The data for the four factors (Memory Control, Response Inhibition Control, Confirmation and True Belief Filler Trials) were slightly negatively skewed. A log transformation was ensued on the two factors namely Confirmation Filler Trial and Memory Control that was severely negatively skewed in an effort to make the data normally distributed. However, the data remained skewed .The data for the dependent variable; false belief performance was normally distributed.

### Multiple Regression.

Two multiple regression analyses were used to test whether the control factors of the RU i.e. Confirmation Filler Trial, Memory Control, Response Inhibition and a True Belief Filler Trial significantly influenced the test factor for false belief performance. The results of the first regression indicated the control factors explained 49.6 % of the variance ($R$ = .705, $R^2$ = .496, $F$ (4, 51) = 12.560, $p$<.001) (see Table 4). It was found that the True Belief Filler Trial contributed significantly to the model, $\beta$ = .629, $t$ = 5.436, $p$ < .001.

**Table 4**

*Multiple Regression on Control Factors Confirmation Filler Trial, Memory Control, Response Inhibition & True Belief Filler Trial*

| | *B* | *SE B* | *Β* | *t* | *p** | *CI* +95% | *CI* - 95% |
|---|---|---|---|---|---|---|---|
| Confirmation Filler | -.37 | .57 | -.07 | -.64 | .528 | -1.52 | .79 |
| Memory Control | .14 | .00 | .05 | .47 | .643 | -.46 | .74 |
| Response Inhibition | 1.31 | 1.13 | .13 | 1.16 | .254 | -.97 | 3.58 |
| True Belief Filler | .91 | .14 | .63 | 5.44 | .000 | .58 | 1.25 |

$R^2$=.496, *α=.05

The second multiple regression analysis excluded the True Belief Filler Trial, which was used to identify whether the participants used artificial strategies to solve the task. The performance on the three control factors, Confirmation Filler Trial, Memory Control and Response Inhibition, explained 20.4% of the variance ($R$ = .452, $R^2$ = .204, $F$ (3, 52) =4.456, $p$=.007) (see Table 5). It was found that Memory Control (β = .282, $t$ = 2.258, $p$ = .028) and Response Inhibition (β = .331, $t$ = 2.560, $p$ = .013) also significantly influenced the false belief performance.

**Table 5**

*Multiple Regression on Control Factors Confirmation Filler Trial, Memory Control, Response Inhibition excluding True Belief Filler Trial*

| | *B* | *SE B* | *β* | *t* | *p** | *CI* +95% | *CI* -95% |
|---|---|---|---|---|---|---|---|
| Confirmation Filler | -.56 | .71 | -.10 | -.79 | .432 | -.20 | .87 |
| Memory Control | .77 | .34 | .28 | 2.26 | .028 | .09 | 1.46 |
| Response Inhibition | 3.39 | 1.33 | .33 | 2.56 | .013 | .73 | 6.06 |

$R^2$=.204, *α=.05

**ANCOVA.**

One one-way analyses of covariance (ANCOVA) and a two-way ANCOVA were performed on all the factors (The Confirmation Filler trial, the Memory Control and Response Inhibition Control) except the True Belief Filler Trial. The True Belief Filler Trial was excluded to investigate whether the significance on this trial was masking the possible significance of the other factors. For the first ANCOVA the independent variable, gender, included had two levels, males and females. The dependent variable was the participants' false belief performance. The covariates were the RU control factors which included three levels: Memory Control, Response Inhibition and the True Belief Filler Trial. A preliminary analysis evaluating the homogeneity-of-regression (slopes) assumption indicated that the relationship between the covariates and the dependent variable did not differ significantly as a function of the independent variable, $F(1, 52) = 1.135$, $p = .292$. The assumption of homogeneity of variance was not met since Levene's Test of Equality of Error Variance was statistically significant, $F\ (1, 54) = 5.902$, $p = .018$. The ANCOVA was significant, $F\ (1, 51) = 4.107$, $p = .048$ (see Table 6). This shows that males performed better than females on the RU.

**Table 6**

*Analysis of Co-Variance for Control Factors by False Belief Performance for Gender*

| Source | SS | *df* | *MS* | *F* | *p** | partial $\eta^2$ |
|---|---|---|---|---|---|---|
| Memory Control | .24 | 1 | .24 | .05 | .816 | .00 |
| Response Inhibition | 2.22 | 1 | 2.22 | .50 | .482 | .01 |
| True Belief Filler | 149.04 | 1 | 149.04 | 33.74 | .000 | .40 |
| Gender | 18.14 | 1 | 18.14 | 4.11 | .048 | .08 |
| Error | 225.31 | 51 | 4.42 | | | |
| Total | 479.55 | 55 | | | | |

*α=.05

For the second ANCOVA the independent variables were race which included two levels: White and non-White, and first language, which also included two levels: English and non-English. The dependent variable was the participants' false belief performance, the false belief test and the covariates were the RU control factors, which included three levels: Memory Control, Response Inhibition and True Belief Filler Trial. The assumption of homogeneity of variance was met since Levene's Test of Equality of Error Variance was statistically significant, $F\ (3, 52) = 2.523$, $p = .068$. Preliminary analyses evaluating the homogeneity-of-regression (slopes) assumption were done. These analyses indicated that the relationship between each of the covariates and the dependent variable did not differ significantly as a function of the independent variable, which means the assumption was upheld for all the covariates (see Table 7).

**Table 7**

*Assumption of homogeneity-of-regression (slopes)*

| | *MS* | *F* | *p* |
|---|---|---|---|
| Memory Control *Gender | 11.97 | 1.56 | .217 |
| Response Inhibition*Gender | ** | ** | ** |
| True Belief Filler*Gender | 9.81 | 2.34 | .132 |
| Memory Control*Race | .02 | .00 | .958 |
| Response Inhibition*Race | 12.25 | 1.62 | .209 |
| True Belief Filler*Race | .27 | .06 | .813 |
| Memory Control*First Language | 12.33 | 1.54 | .220 |
| Response Inhibition*First Language | 1.60 | .20 | .654 |
| True Belief Filler*First Language | 2.97 | .63 | .430 |

** Not available because *SD* = 0

The ANCOVA was not statistically significant, *F* (1, 49) = .485, *p* = .490 (see Table 8). There was no significant difference in performance between Whites and non-Whites and English and non-English participants.

**Table 8**

*Analysis of Co-Variance for Control Factors by False Belief Performance for Race and First Language*

| Source | SS | *df* | *MS* | *F* | *p** | partial $\eta^2$ |
|---|---|---|---|---|---|---|
| Memory Control | .59 | 1 | .59 | .12 | .729 | .00 |
| Response Inhibition | 7.13 | 1 | 7.13 | 1.48 | .230 | .03 |
| True Belief Filler | 122.58 | 1 | 122.58 | 25.43 | .000 | .34 |
| First Language | 4.46 | 1 | 4.46 | .94 | .336 | .00 |
| Race | .27 | 1 | .27 | .06 | .816 | .02 |
| First Language*Race | 2.34 | 1 | 2.34 | .49 | .490 | .01 |
| Error | 236.20 | 49 | 4.82 | | | |
| Total | 479.55 | 55 | | | | |

*α=.05

**Correlation between RU and RME**

Pearson’s correlations were conducted for the variables RME test scores and RU false belief task performance. Inspection of the intercorrelation matrix (see Table 9) revealed statistically significant correlations between both variables. RME test scores was positively correlated with false belief performance, $r = .331, p = .013$, although this correlation was small.

**Table 9**

*Intercorrelation matrix of False Belief and RME scores*

| Variable | False belief | RME |
|---|---|---|
| False Belief | - | 0.331 |
| RME | 0.331 | - |

*α=.05

## Discussion

To date there are no normative data collected for ToM tests in a South African context. The aim of this study was to collect normative data on two ToM tests, the Reading the Mind in the Eyes test and the Reality-Unknown False Belief task, for use in a South African context. However, due to small sample sizes and thus a lack of statistical power, the norm scores attained may not be fully representative of the targeted population. Statistical analyses were performed on both tests. We found that within a South African tertiary student population, Whites performed better on the RME task compared to non-Whites. The pairwise comparisons also showed a significant difference in performance in the groups whose first language was not English, where Whites performed better than non-Whites. We also found that for the RU task, participants did not perform at ceiling as expected. However, the marathon nature in which the test was administered may have confounded our findings. We also found a significant gender difference for RU task, where males performed better than females.

### Norms Scores

The results obtained in this study showed that the South African participants did not perform at ceiling for most of the factors in the RU task. Since other studies have found that normal participants in a western population perform at ceiling, the results obtained were unexpected (Apperly et al., 2005). However, due to the small and unequal distribution across groups, it was decided that all groups with sample sizes less than ten would not be interpreted to validate the results obtained (see Table 2). With this criterion in mind the norms for the White English females and the non-White English females could be interpreted. The cut-off scores were similar for these two groups on both tests. There were slight differences in cut-off scores for the two groups on the True Belief Filler Trial, where the baseline cut-off score for Whites English females (≤ 8.34) is higher than for non-White English females (≤ 5.09).

## Reading the Mind in the Eyes Test

### Overall scores on the RME.

This study showed that normal participants performed below ceiling, which is consistent with a previous study (Baron-Cohen et al., 2001). For item 17, when the result was compared against a previous study the student population scored similar on this item compared to our sample (Baron-Cohen et al., 2001) (see Figure 6). This suggests that this item was particularly difficult item for both student populations. For item 23, when this result was compared against a student sample in another study, our sample performed much lower in comparison. This suggests that this item may be unsuitable for use in a South African student population.

### Race.

In this study a statistically significant main effect for race was found, where Whites performed better than non-Whites on the RME. This study replicated earlier findings which show that non-White South Africans' performance on western neuropsychological tests, e.g. "Raven's Standard Progressive Matrices", are poorer compared to their American respective norm groups (Rushton, Skuy, & Fridjhon, 2002). A possible reason for Whites performing significantly better than non-Whites could be because the images used in the present study consisted of only whites faces. Studies have shown that it is easier to interpret the mental state of those of the same race compared to those from other cultures or races (Adams et al., 2009). Neuro-imaging studies have also shown that individuals have superior brain region activation for same-race faces (Herrmann et al., 2007). This suggests that Whites have race-advantage of performing better on such a test compared to non-Whites, since all the images in the test are only of white faces. Thus, there is high cultural loading on the RME, which suggests that the validity of the RME test in collecting ToM data in a South African context is limited.

### Gender.

There was no statistically significant gender difference in ToM performance on the RME test. This is in contrast to the significant sex difference that was found in the original as well as the revised version of the RME in an American sample (Baron-Cohen et al., 2001).

The sample in the present study was small and unequal so this might have lowered the chances of detecting a gender difference.

**First language.**

Although the RME is a verbal-based test there was no statistically significant main effect for first-language, contrary to what we had hypothesized. Since the participants were able to refer back to the glossary at any point during testing this might have significantly influencing the main effect. This enabled participants regardless of their first language preferences to have an equal chance of getting the items correct. In addition, 47 out of the 56 participants attend UCT which is an English University, and therefore are more likely to be proficient in English.

**Reality Unknown False Belief Task**

The RU is a relatively novel test and has only been used in a few studies Apperly et al., 2006; Apperly et al., 2009; Grant, Apperly, & Oliver, 2007). Normative data in a western population on the RU task found that normal participants performed at ceiling (Apperly et al., 2005). However, in this study the participants did not perform at ceiling. The control factors memory and response inhibition, which are both executive function processes also had a significant influence on false belief. This is consistent with other studies that have acknowledged the relationship between ToM and other cognitive functions (Carlton, Moses & Breton, 2002; Pellicano, 2007; Perner, Kloo, & Gornik, 2007; Sabbagh et al., 2006).

However, the control factor that was found to have the greatest significant influence on false belief performance was the True Belief Filler Trial (see method section for sequence of events). Since, these filler trials guard against participants' using artificial strategies to solve the task, it showed that in this study, participants in this study did not pay attention, were fatigued and bored. Thus, the participants may have adopted a superficial way of answering the test items i.e. the participants could have simply pointed to the opposite container to what the women indicated. This was explained by the fact they did not do well on this control task. This may suggest that the sample used in this study may have been problematic, as we used normal participants who should have performed at ceiling.

An ANCOVA indicated that there was a statistically significant effect for gender on the RU. However, this statistically significant effect may be due to experimental error, as our sample size was small and lacked statistical power. The comparison of the performance on the RME and the RU found that there was a small correlation between the two. However, it is important to note that the RME is standardized; whereas, as aforementioned, the validity of the RU scores is debateable. In addition, although both the RU and the RME is a test for ToM, the RME was testing another aspect of ToM, namely facial affect recognition.

## Limitations and Directions for Future Research

As an honours project, the present study was restricted by time constraints. There are several methodological limitations of this study that should be acknowledged. In the present study data was drawn from a small and unequal sample and the results attained should therefore be treated with caution. Firstly, the sample is not representative of a South African student population since most participants were predominantly White, female undergraduate psychology students. Hence, the sample is possibly biased because there were few male as well as non-White participants included in the study. This sample was also unrepresentative of the general South African population since the sample was drawn from a university population. Students at tertiary institutes tend to have a higher educational level than most 18-32 year olds, which may have had an influence on test performance. Therefore, the norms in this study only apply to educated young adults, which indicate that more data is needed on participants from different demographic profiles.

In this study the small and unrepresentative sample led to a loss of statistical power and thus limited the reliable interpretation of the results. Preferably a follow-up study with a larger sample size should be ensued to confirm these results. The impact that the participants' socio-economic statuses, quality of secondary education, level of tertiary education and second language have on their false belief performance were not explored. Therefore, if there was any influence of these factors on false belief, then they need to be included as independent variables in future studies.

The neuropsychological measures used in this study have not been subjected to the validity tests that are necessary to estimate their cross-cultural relevance. Furthermore, even though one of the tests was nonverbal, it was not completely free of culture bias. The participants were divided only in two categories i.e. White and non-White and English and non-English. Future research should accommodate for more categories so that difference in performances across different races and first languages can be more accurately identified. This is particularly relevant in a South African context, where there is a diverse population in terms of race and language background.

The nature of the administration used in this study could account for the participants performing below ceiling on the RU. Future studies should introduce more breaks between the blocks to prevent participants from using superficial strategies for answering the items in the test. This study also did not follow a particular sequence of testing the RU and the RME. This could have been an artefact and affected the performance on the two tests, especially because the RU was a long study. Thus, in future studies the tests should be administered adhering to same test sequence for all participants.

## Conclusion

ToM is an effective aspect of successful social interaction. Evidently, there is a necessity for reliable and valid ToM tests in clinical as well as non-clinical populations. The purpose of this study was to collect normative data on two established ToM tests within a South African context. However, most standardized ToM tests were developed for a western population where normal adults performed at ceiling. Yet, when these tests were applied in a South African context, the same results were not obtained. Due to the marathon manner in which the RU task was administered the participants in this sample may have performed below ceiling due to boredom or lack of concentration. It was found that there was a statistically significant race difference where Whites performed better than non-Whites on the RME task. Due to the race effect on the RME, the faces in the test should include different race faces and not only White faces to minimise the cultural bias. The RU is a more applicable and valid measure for use in a South African context as it is free from culture bias, it controls for factors that have an influence on ToM performance and also has limited semantic loading.

## References


Abu-Akel, A., & Abushua'leh, K. (2004). Theory of mind in violent and nonviolent patients with paranoid schizophrenia. *Schizophrenia Research*, *69*, 45–53.

Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., … & Nomura, M. (2009). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, *22*, 97–108.

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*, 231–239.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*, 165–178.

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693-716.

Apperly, I. A., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. (2006). Intact first-and second-order false belief reasoning in a patient with severely impaired grammar. *Social Neuroscience*, *1*, 334-348.

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.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2005). Domain-specificity and theory of mind: evaluating neuropsychological evidence. *Trends in Cognitive Sciences*, *9*, 572–577.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental psychology*, *4*, 190-201.

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*,*103*, 37–49.

Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., & Reiss, A. L. (2004). White matter structure in autism: preliminary evidence from diffusion tensor imaging. *Biological Psychiatry*, *55*, 323–326.

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., Ring, H., Moriarty, J., Schmitz, B., Costa, D. &, Ell, P. (1994). Recognition of mental state terms: Clinical findings in children with autism and a functional neuroimaging study of normal adults. *Brain Journal of Psychiatry, 165,* 640–9.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry, 38,* 813–822.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience*, *11*, 1891–1898.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-Functioning Autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*, 241-251.

Baune, B.T., McAfoose, J., Leach, G., Quirk, F., & Mitchell, D. (2009). Impact of psychiatric and medical comorbidity on cognitive function in depression. *Psychiatry and Clinical Neurosciences, 63*, 392-400.

Bibby, H., &, McDonald, S. (2005). Theory of mind after traumatic brain injury. *Neuropsychologia*, *43*, 99-114.

Boone, K. B., Victor, T. L., Wen, J., Razani, J., & Pontón, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, *22*, 355–365.

Brekke, J. S., Hoe, M., Long, J., & Green, M. F. (2007). How neurocognition and social cognition influence functional change during community-based psychosocial rehabilitation for individuals with schizophrenia. *Schizophrenia bulletin*, *33*, 1247-1256

Brekke, J. S., Long, J. D., Nesbitt, N., & Sobel, E. (1997). The impact of service characteristics on functional outcomes from community support programs for persons with schizophrenia: A growth curve analysis. *Journal of Consulting and Clinical Psychology*, *65*, 464-475.

Bruns, J., & Hauser, W.A. (2003). The epidiomology of traumatic brain injury. *Epilepsia, 44,* 2-10.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70,* 381–395.

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., … & Tapanya, S.

(2005). Synchrony in the Onset of Mental-State Reasoning. *Psychological Science*, *16*, 378–384.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*, 73–92.

Channon, S., & Crawford, S. (2000). The effects of anterior lesions on performance on a story comprehension test: Left anterior impairment on a theory of mind-type task. *Neuropsychologia*, *38*, 1006–1017.

Combs, D. R., Adams, S. D., Penn, D. L., Roberts, D., Tiegreen, J., & Stem, P. (2007). Social Cognition and Interaction Training (SCIT) for inpatients with schizophrenia spectrum disorders: preliminary findings. *Schizophrenia Research, 91*, 112–116.

Corcoran, R., & Frith, C. D. (2003). Autobiographical memory and theory of mind: Evidence of a relationship in schizophrenia. *Psychological Medicine, 33*, 897–905.

Couture, S. M., Penn, D. L., & Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: a review. *Schizophrenia Bulletin*, *32*, 44-63.

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist, 23*, 1173–1195.

Crawford, J.R., Garthwaite, P.H., & Betkowska, K. (2009). Bayes' theorem and diagnostic tests in neuropsychology: Interval estimates for post-test probabilities. *The Clinical Neuropsychologist*, *23*, 624–644.

Cummings, J. L. (1997). The Neuropsychiatric Inventory: Assessing psychopathology in dementia patients. *Neurology, 48*, 10-16.

Field, A. P. (2009). *Discovering Statistics Using SPSS*. SAGE Publications Ltd.

Figueras-Costa, B., & Harris, P. (2001). Theory of mind development in deaf children: A nonverbal test of false-belief understanding. *Journal of Deaf Studies and Deaf Education, 6*, 92-108.

Fletcher ,P.C., Happé. F., Frith, .U. (1995).Other minds in the brain: A functional imaging study of ‘‘theory of mind’’ in story comprehension. *Cognition*, *57*,109–28.

Folstein, M. F., Folstein, S. E., & McHugh, P. R., others. (1975). Mini-mental state. *A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res, 12*, 189–198.

Gallagher, H.L., Happé, 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.

Golby, A.J., Gabrieli, J.D.E., Chiao, T.Y, & Eberhardt, J.L. (2001). Differential responses in the fusi-form region to same-race and other-race faces. *Nature Neuroscience, 4*, 845-850.

Golden, C. J., MacInnes, W. D., Ariel, R. N., Ruedrich, S. L., Chu, C. C., Coffman, J. A., … & Graber, B. (1982). Cross-validation of the ability of the Luria-Nebraska Neuropsychological Battery to differentiate chronic schizophrenics with and without ventricular enlargement. *Journal of Consulting and Clinical Psychology*, *50*, 87-95.

Grant, C. M., Apperly, I., & Oliver, C. (2007). Is theory of mind understanding impaired in males with fragile X syndrome? *Journal of Abnormal Child Psychology*, *35*, 17–28.

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*, 346–359.

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.

Happé, F., Malhi, G. S., & Checkley, S. (2001). Acquired mind-blindness following frontal lobe surgery? A single case study of impaired "theory of mind" in a patient treated with stereotactic anterior capsulotomy. *Neuropsychologia*, *39*, 83–90.

Herrmann, M. J., Schreppel, T., Jaager, D., Koehler, S., Ehlis, A. C., & Fallgatter, A. J. (2007). The other-race effect for face perception: An event-related potential study. *Journal of Neural Transmission, 114*, 951-957.

Hobart, M. P., Goldberg, R., Bartko, J. J., & Gold, J. M. (1999). Repeatable battery for the Assessment of Neuropsychological Status as a screening test in schizophrenia, II: Convergent/discriminant validity and diagnostic group comparisons. *American Journal of Psychiatry*, *156*, 1951-1957.

Homer, B. D., Solomon, T. M., Moeller, R. W., Mascia, A., DeRaleau, L., & Halkitis, P. N. (2008). Methamphetamine abuse and impairment of social functioning: A review of the underlying neurophysiological causes and behavioural implications. *Psychological Bulletin*, *134*, 301-310.

Kee, K. S., Horan, W. P., Salovey, P., Kern, R. S., Sergi, M. J., Fiske, A. P., Lee, J., et al., others. (2009). Emotional intelligence in schizophrenia. *Schizophrenia Research*, *107*, 61–68.

Kendall, E., & Terry, D. (1996). Psychosocial adjustment following closed head injury: A model for understanding individual differences and predicting outcome- Neuropsychological Rehabilitation. *Neuropsychological Rehabilitation, 6*, 101-132.

Lee, K. H., Farrow, T. F. D., Spence, S. A., & Woodruff, P. W. R. (2004). Social cognition, brain networks and schizophrenia. *Psychological Medicine*, *34*, 391–400.

Levin, H.S. (1995). Neurobehavioural outcome of closed head injury: Implications for clinical trials. *Journal of Neurotrauma*, *12*, 601-610.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, *44*, 523-532.

McDonald, S., & Flanagan, S. (2004). Social perception deficits after traumatic brain injury: interaction between emotion recognition, mentalizing ability, and social communication. *Neuropsychology, 18*, 572–579.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3-35.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*, 608.-615

Milders, M., Fuchs, S., & Crawford, J. R. (2003). Neuropsychological impairments and changes in emotional and social behaviour following severe traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *25*, 157–172.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development*, *78*, 622–646.

Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development*, *79*, 80–99.

Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica*, *124*, 106–128.

O'Shea, E. (2003). Costs and consequences for the carers of people with dementia in Ireland. *Dementia*, *2*, 201-212.

Paladino, M.P., Leyens, J.P., Rodriguez, R., Rodriguez, A., Gaunt, R., Demoulin, S. (2002). Differential association of uniquely non-human and human emotions with the in-group and the out-group. *Group Process Intergroup Relations, 5*, 105-117.

Penn, D. L., Corrigan, P. W., Bentall, R. P., Racenstein, J., & Newman, L. (1997). Social cognition in schizophrenia. *Psychological Bulletin*, *121*, 114-132.

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology*, *43*, 974-990.

Perner, J., Kloo, D., & Gornik, E. (2007). Episodic memory development: Theory of mind is part of re-experiencing experienced events. *Infant and Child Development*, *16,* 471–490.

Pijnenborg, G. H., Withaar, F. K., Evans, J. J., Van Den Bosch, R. J., Timmerman, M. E., & Brouwer, W. H. (2009). The predictive value of measures of social cognition for community functioning in schizophrenia: implications for neuropsychological assessment. *Journal of the International Neuropsychological Society, 15*, 239–247.

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. *Journal of Clinical and Experimental Neuropsychology*, *20*, 310–319.

Rieffe, C., Terwogt, M. M., & Cowan, R. (2005). Children's understanding of mental states as causes of emotions. *Infant and Child Development*, *14*, 259–272.

Roos, A., Calata, D., Jonkers, L., Maritz, S.J., Kidd, M., Barnels, W.M.U., & Hugo, F.J. (2010). Normative data for the Tygerberg Cognitive Battery and Mini-Mental Status Exam in a South African population. *Comprehensive Psychiatry*, *51*, 207-216.

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: a critical review. *Brain and Cognition*, *52*, 326–333.

Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). Theory of mind impairments and their relationship to executive functioning following frontal lobe excisions. *Brain*, *124*, 600-616.

Ruffieux, N., Njamnshi A.K., Mayer, E., Sztajzel, R., Eta, S.C., Doh, R.F., … & Kengne, A.M., (2009). Neuropsychology in Cameroon: First normative data for cognitive tests among school-aged children. *Child Neuropsychology*, 1-19.

Rushton, J. P., & Skuy, M. (2000). Performance on Raven's Matrices by African and White university students in South Africa. *Intelligence*, *28*, 251–265.

Rushton, J. P., Skuy, M., & Fridjhon, P. (2002). Jensen effects among African, Indian, and white engineering students in South Africa on Raven's standard progressive matrices. *Intelligence*, *30*, 409–423.

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition*, *55*, 209–219.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind. *Psychological Science, 17*, 74-81.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & De Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16*, 440-458.

Seltzer, B., Vasterling, J. J., Yoder, J. A., & Thompson, K. A. (1997). Awareness of deficit in Alzheimer's disease: relation to caregiver burden. *The Gerontologist*, *37*, 20-39.

Sergi, M.J., Rossovsky, Y., Neuchterlain, K.H., & Green, M.F. (2006). Social Perception as a mediator of the influence of early visual processing on functional status in schizophrenia. *American Journal of Psychiatry, 163*, 448-454.

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*, 1239-1247.

Silliman, E. R., Diehl, S. F., Bahr, R. H., Hnath-Chisolm, T., Zenko, C. B., & Friedman, S. A. (2003). A new look at performance on theory-of-mind tasks by adolescents with autism spectrum disorder. *Language, Speech, and Hearing Services in Schools*, *34*, 236-248.

Skuy, M., Schutte, E., Fridjhon, P., & O'Carroll, S. (2001). Suitability of published neuropsychological test norms for urban African secondary school students in South Africa. *Personality and Individual Differences*, *30*, 1413–1425.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*, 587-592.

SPSS Inc. (2008). SPSS Statistics (Version 17.0) [Computer software]. Chicago, IL: Author.

Sprong, M., Schothorst, P., Vos, E., Hox, J., & Van Engeland, H. (2007). Theory of mind in schizophrenia: meta-analysis. *The British Journal of Psychiatry, 191*, 5-13.

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*, *41*, 209–220.

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

Stuss, D.T., Gallup, G.G., Alexander, M.P. (2001). The frontal lobes are necessary for theory of mind. *Brain, 124*, 279–86.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants.

*Psychological Science, 18,* 580-586.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93,* 1–9.

Turkstra, L. S. (2008). Conversation-based assessment of social cognition in adults with traumatic brain injury. *Brain Injury*, *22,* 397–409.

Uzzell, B. P., Pontón, M. O., & Ardila, A. (2007). *International handbook of cross-cultural neuropsychology*. Routledge.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*, 655–684.

Welz, T., Hosegood, V., Jaffar, S., Bätzing-Feigenbaum, J., Herbst, K., & Newell, M. L. (2007). Continued very high prevalence of HIV infection in rural KwaZulu-Natal, South Africa: a population-based longitudinal study. *Aids, 21,* 1467-1472.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 41-68.

Zucker, N. L., Losh, M., Bulik, C. M., LaBar, K. S., Piven, J., & Pelphrey, K. A. (2007). Anorexia nervosa and autism spectrum disorders: Guided investigation of social cognitive endophenotypes. *Psychological Bulletin, 133*, 976-1006.

## Appendix A

**Consent Form**

CONSENT TO PARTICIPATE IN RESEARCH

We are inviting you to be in our research study because we would like to learn more about the way people make sense of and predict their own behaviour as well as that of others.

If you agree to be in this study we will ask you to come to our lab to do some activities with us. For example, we may ask you to watch a video and answer questions.

These exercises and activities will not hurt you, but some of them may be long and you may feel tired at times. If you do, you can stop and rest at any time.

Signing this paper means that you want to be in the study. If you don't want to be in the study, don't sign the paper. The amount and quality of care you receive won't change if you don't sign this paper, and it won't change if you change your mind later and want to stop.

You can ask any questions that you have about the study. If you have a question later that you didn't think of now, you can call us on 0846545784/0795435708 or ask us next time.

Signature of Participant ____________________ Date _________

Signature of Investigator ____________________ Date ________

# Appendix B

## Reality-Unknown False Belief task Scoring Sheet

NAME:
DATE:

*Instructions:* *"You will see a man and a woman on a video.*
*The man will hide a green little object in one of two boxes.*
*The woman's job is to help you finding in which box the green object is."*
Give 2 practice items, if the participant responds correctly on both, start the test;
otherwise present the practice trials again until the participant gives 2 correct responses.

**FB-block 1**

| | Trial ID | Categ | Resp | Participant's response |
|---|---|---|---|---|
| 1 | B21 | B | R | |
| 2 | D11 | D | L | |
| 3 | C25 | C | R | |
| 4 | E12 | E | L | |
| 5 | A23 | A | R | |
| 6 | E11 | E | L | |
| 7 | C14 | C | L | |
| 8 | A24 | A | R | |
| 9 | B17 | B | L | |
| 10 | D26 | D | R | |
| 11 | E24 | E | R | |
| 12 | B11 | B | L | |
| 13 | D24 | D | R | |
| 14 | C16 | C | L | |
| 15 | A15 | A | L | |

**FB-block 2**

| | Trial ID | Categ | Resp | Participant's response |
|---|---|---|---|---|
| 1 | B12 | B | L | |
| 2 | C13 | C | L | |
| 3 | B23 | B | R | |
| 4 | A13 | A | L | |
| 5 | D23 | D | R | |
| 6 | B14 | B | L | |
| 7 | D14 | D | L | |
| 8 | A27 | A | R | |
| 9 | E26 | E | R | |
| 10 | C11 | C | L | |
| 11 | E14 | E | L | |
| 12 | C21 | C | R | |
| 13 | E13 | E | L | |
| 14 | D21 | D | R | |
| 15 | A22 | A | R | |

NAME:
DATE:

*Instructions:* *"You will see a man and a woman on a video.*
*The man will hide a green little object in one of two boxes.*
*The woman's job is to help you finding in which box the green object is."*
Give 2 practice items, if the participant responds correctly on both, start the test;
otherwise present the practice trials again until the participant gives 2 correct responses.

**FB-block 3**

| | Trial ID | Categ | Resp | Participant's response |
|---|---|---|---|---|
| 1 | A26 | A | R | |
| 2 | C24 | C | R | |
| 3 | D15 | D | L | |
| 4 | B22 | B | R | |
| 5 | E15 | E | L | |
| 6 | C23 | C | R | |
| 7 | A14 | A | L | |
| 8 | E27 | E | R | |
| 9 | C15 | C | L | |
| 10 | A17 | A | L | |
| 11 | B25 | B | R | |
| 12 | D25 | D | R | |
| 13 | B13 | B | L | |
| 14 | D19 | D | L | |
| 15 | E22 | E | R | |

**FB-block 4**

| | Trial ID | Categ | Resp | Participant's response |
|---|---|---|---|---|
| 1 | B15 | B | L | |
| 2 | D13 | D | L | |
| 3 | B26 | B | R | |
| 4 | E16 | E | L | |
| 5 | A16 | A | L | |
| 6 | A25 | A | R | |
| 7 | B24 | B | R | |
| 8 | A11 | A | L | |
| 9 | D22 | D | R | |
| 10 | E21 | E | R | |
| 11 | C12 | C | L | |
| 12 | E25 | E | R | |
| 13 | D16 | D | L | |
| 14 | C22 | C | R | |
| 15 | C26 | C | R | |

## Appendix C

## Reading the Mind in the Eyes test Scoring Sheet

| ***List of Target Mental State Terms for Each Item (in Italic) and Their Distractors*** | | | | | |
|---|---|---|---|---|---|
| ***Item*** | ***Word*** | | | | ***Score*** |
| Pla | jealous | ***panicked*** | arrogant | hateful | |
| 1 | ***playful*** | comforting | irritated | bored | |
| 2 | terrified | ***upset*** | arrogant | annoyed | |
| 3 | joking | flustered | ***desire*** | convinced | |
| 4 | joking | ***insisting*** | amused | relaxed | |
| 5 | irritated | sarcastic | ***worried*** | friendly | |
| 6 | aghast | ***fantasizing*** | impatient | alarmed | |
| 7 | apologetic | friendly | ***uneasy*** | dispirited | |
| 8 | ***despondent*** | relieved | shy | excited | |
| 9 | annoyed | hostile | horrified | ***preoccupied*** | |
| 10 | ***cautious*** | insisting | bored | aghast | |
| 11 | terrified | amused | ***regretful*** | flirtatious | |
| 12 | indifferent | embarrassed | ***sceptical*** | dispirited | |
| 13 | decisive | ***anticipating*** | threatening | shy | |
| 14 | irritated | disappointed | depressed | ***accusing*** | |
| 15 | ***contemplative*** | flustered | encouraging | amused | |
| 16 | irritated | ***thoughtful*** | encouraging | sympathetic | |
| 17 | ***doubtful*** | affectionate | playful | aghast | |
| 18 | ***decisive*** | amused | aghast | bored | |
| 19 | arrogant | grateful | sarcastic | ***tentative*** | |
| 20 | dominant | ***friendly*** | guilty | horrified | |
| 21 | embarrassed | ***fantasizing*** | confused | panicked | |
| 22 | ***preoccupied*** | grateful | insisting | imploring | |
| 23 | contented | apologetic | ***defiant*** | curious | |
| 24 | ***Pensive*** | irritated | excited | hostile | |
| 25 | panicked | incredulous | despondent | ***interested*** | |
| 26 | alarmed | shy | ***hostile*** | anxious | |
| 27 | joking | ***cautious*** | arrogant | reassuring | |
| 28 | ***Interested*** | joking | affectionate | contented | |
| 29 | impatient | aghast | irritated | ***reflective*** | |
| 30 | grateful | ***flirtatious*** | hostile | disappointed | |
| 31 | ashamed | ***confident*** | joking | dispirited | |
| 32 | ***serious*** | ashamed | bewildered | alarmed | |
| 33 | embarrassed | guilty | fantasizing | ***concerned*** | |
| 34 | aghast | baffled | ***distrustful*** | terrified | |
| 35 | puzzled | ***nervous*** | insisting | contemplative | |
| 36 | ashamed | nervous | ***suspicious*** | indecisive | |
| | | | | **Total score** | |