DOES TRAINING IMPROVE PERFORMANCE ON A PERSPECTIVE-TAKING TASK?

A thesis
submitted in partial fulfilment
of the requirements for the degree
of
Master of Applied Psychology (Behaviour Analysis)
at
The University of Waikato
by

Laura Baker

The University of Waikato
2009

Abstract

This study investigated the effect of training on perspective-taking with 22 normally developed adults. The perspective-taking task, similar to that used by McHugh, Barnes-Holmes, Barnes-Holmes, Whelan, and Stewart (2007), required the participants read two related statements per trial presented on a computer screen. They pressed one of two keys to indicate if they thought the second statement was true or false. The statements differed along three dimensions, perspective (Self, Other, or Photo), belief (true- or false- belief), and correct response (true or false). Latency to respond, timed from the end of the statement presentation, and accuracy were recorded. A reaction time task, that requiring participants to indicate if a statement (“This is…(colour name)”) about a coloured square was true or false, was included to assess the effect of task repetition on response latencies. There were four blocks of reaction time trials alternating with three blocks of perspective-taking trials (Pre-test, Training, and Post-test). During the Training phase there was feedback on the accuracy of each response. Feedback was not given in the Pre-test, Post-test, or reaction time trials. Extended training on the perspective-taking task reduced latencies on this task over and above the decreases seen in the latencies on the reaction time tests, and this reduction generalized to a novel stimulus set. The Self and Other questions resulted in longer latencies than the Photo questions (both before and after the removal of reaction time) as predicted by Relational Frame Theory. The longer latencies were associated with greater relational complexity and partially replicated the results of McHugh et al. (2007a). These results suggest that training with multiple exemplars can be used to decrease response latencies, and so to improving performance, on a perspective-taking task.

Acknowledgements

I would like to sincerely thank my supervisor Mary Foster, and Jennifer Kinloch for your continued guidance and support throughout the year, for keeping me focused and on track, for promptly replying to my many questions, and for understanding my frustrations (especially with R). Your time and effort has been greatly appreciated.

To Andrew Malcolm for making the computer programme which he appropriately named ‘patience’, something I’ve needed a lot of over the past year.

To my friends, family, and especially my partner Emerson for your constant support, understanding, and encouragement that got me through the tough times. You’ve been there when I’ve needed a good laugh, and when I needed to be persuaded to get back to my work.

Table of Contents

Page

List of Tables

Table Page

Table Page

List of Figures

List of Appendices

Recent interest in the development of perspective-taking has emerged from the apparent link between perspective-taking deficits and autistic spectrum disorder (McHugh, Barnes-Holmes, & Barnes-Holmes, 2004). It has been well established that perspective-taking skills develop as part of a child's normal development, and that perspective-taking skills increase as a function of age (McHugh et al., 2004). It is also apparent that autistic populations do not follow this same pattern (Sodian & Firth, 1992). The investigation of populations both with normal perspective-taking skills and with perspective-taking deficits may be helpful in understanding its development.

Perspective-taking has been defined as "understanding that another person's beliefs about events may be different from reality and that those beliefs will guide future behaviour" (LeBlanc, Coates, Daneshvar, Chalop-Christy, Morris, & Lancaster, 2003, p.253). Perspective-taking skills are an important part of development as they provide the necessary tools for an individual to understand and predict social behaviour and to understand social interactions of moral significance such as deception (McHugh, Barnes-Holmes, Barnes-Holmes, & Stewart, 2006).

Populations with Deficits

Perspective-taking deficits are more prevalent in autistic populations than other non-autistic developmentally disabled children of comparable mental ages whose performance is similar to normally developing children (Sodian & Frith, 1992). Sodian and Frith (1992) compared the performance of three groups of children, normally developing, learning disabled, and autistic who were matched on mental age. The children were tested on perspective-taking tasks involving deception, sabotage, and false-belief. The autistic group performed significantly worse than the learning disabled and normal group on the false-belief tasks. The autistic group also performed extremely poorly on the deception tasks compared to the sabotage tasks. This was not found with either the learning disabled, or the normal group. The authors suggest that these results demonstrate that the autistic children not only fail to infer other people's false-beliefs, but also fail to manipulate other people's beliefs (deception tasks). Interestingly the autistic group performed well on the sabotage tasks, as did the other two groups. Sodian and Frith (1992) also suggest that deficits in the autistic group are likely to be specific to the area referred to as representation of mental states as opposed to a more general impairment in social interactive skills.

Russell, Mauthner, Sharpe and Tidswell (1991) found similar results to Sodian and Firth (1992) when they investigated deception in normally developing children, children with Down's syndrome, and children with autism. The simple scenario used to test deception involved pointing to one of two boxes, one of which contained sweets, to deceive the experimenter so that the child got the sweets and the experimenter did not. To correctly deceive the experimenter, the child would point to the box without any sweets so that the experimenter would take the empty box and the child would have the box containing sweets. Most of the 4year old children in the normal and Down's syndrome groups successfully completed this task, whereas most of the autistic group (nine out of 11 children) failed to deceive the experimenter along with most 3year olds in the normal and Down's syndrome groups.

There has been little research investigating the perspective-taking abilities of homogeneous non-autistic developmentally disabled groups. However, Sodian and Frith (1992) say that anecdotal evidence from parents and teachers clearly suggests that the ability to lie and deceive is an area of deficit in autistic populations. Autism is characterized by deficits in the ability to form reciprocal social relations. Rehfeldt, Dillen, Ziomek, and Kowalchuk (2007) suggested that these deficits and the inability to perform well on perspective-taking tasks may be linked. Being able to view a situation from another's perspective would contribute to successful social relations such as reciprocal conversation, cooperative play, and the ability to sympathize and empathize.

To data, previous attempts to train autistic individuals to perspective-take have involved the use of video modelling, however, these studies have shown mixed results and the research lacks reliable generalization and maintenance data (e.g., LeBlanc et al., 2003; Charlop-Christy & Daneshvar, 2003). LeBlanc et al. (2003) used video modelling and reinforcement to train three autistic children on perspective-taking tasks. The authors conclude that "video modelling with reinforcement was an effective teaching procedure for these perspective-taking tasks" (p.257), however closer inspection may suggest otherwise. Video modelling involved the children watching a video of an adult correctly completing the task with emphasis focusing on relevant visual cues. The video was paused following the adult's correct response to the questions and then the child was immediately presented with the perspective-taking questions. All three children failed the baseline task and then passed the task and other task variations after completion of training. These results appear to show mastery, however at closer inspection they may not be as conclusive as suggested. One child showed mixed results during the training sessions, and did not have any follow up data. Another required a booster session before being able to pass the follow-up task, and the third child did not have any follow-up data.

In a similar study Charlop-Christy and Daneshvar (2003), also using video modelling to train perspective-taking to children with autism, concluded that "the present results contribute to a growing literature that states that perspective-taking can be taught to children with autism" (p16). Of the three children in the study, only one appeared to master the task. The other two children required a number of training sessions on some tasks, and one of these children received special training but was still unable to pass the task. There was no follow-up data to show how participants would respond following a break in training, therefore we cannot know whether the skills taught were maintained over time.

The results of these two studies show that some autistic children are able to pass perspective-taking tasks following video modelling training, while others show mixed results, or fail. It is not clear what criterion was used to conclude that the interventions were successful. The generalization data of LeBlanc et al. (2003) and Charlop-Christy and Daneshvar (2003) show mixed results of the acquired perspective-taking skills to novel stimulus variations, with some participants able to generalize to all novel variations while others only appeared to generalize to a few novel variations. Follow-up data is lacking, therefore little can be said about the maintenance of this skill over time. This then poses the question, does a trained skill need to be maintained for the training to be deemed successful?

With all this taken into account it would be more accurate to conclude that video modelling shows promising results in training perspective-taking skills to some children with autism, however, further research is required to investigate the longitudinal maintenance effects, and to investigate why some children respond to the video modelling training with more success than others.

The results of using video modelling for training children with autism on perspective-taking tasks are interesting given that video modelling has been successfully combined with other strategies (e.g., reinforcement) to teach conversational and self-care skills (Charlop-Christy, Le, & Freeman, 2001), purchasing skills (Alcantara, 1994), and academic skills (Charlop-Christy et al., 2001) to this population. Recent research suggests that individuals with ASD have more difficulty imitating than normally developing children and children with other developmental disorders (Vivanti, Nadig, Oxonoff, & Rogers, 2008). This inability to learn by imitation may have an influence on an autistic child learning perspective-taking from watching a video and copying the response of an adult. Together, these results suggest that perspective-taking skills appear to be difficult to teach to autistic populations.

Theories

There are many theories in the literature that attempt to provide an account of perspective-taking. Those from the cognitive field tend to involve the description of processes that occur somewhere within the mind that lead to overt observable behaviours. For example, Leekam and Perner (1991) suggest that autistic children lack a Theory of Mind (ToM) because of an inability to metarepresent. Having a ToM refers to a person having mastered all the mental processes required to be able to successfully engage in perspective-taking. ToM will be discussed later in more detail. Leekam and Perner (1991) interpret Leslie's sense of the term metarepresentation as referring to the ability to decouple. This is the ability to copy a primary representation into a metarepresentational context "so that instead of directly representing objects or states in the world, the metarepresentation is detached or screened off from its causal tie to reality" (p.204). Leekam and Perner (1991) argue that if autistic children have a metarepresentational deficit in this sense then they should have difficulties with both mental representations of false-belief and with external representations such as photographs, but report that their study did not show this to be true. Leekam and Perner (1991) then conclude that autistic children do not have a metarepresentational deficit in Leslie's sense of the term. They also suggest that their data leave open the possibility that children with autism have a metarepresentational failure in Pylyshyn's sense of the term which refers to the ability to represent the representational relationship itself. Leslie and Thaiss (1992) argued that Leekam and Perner's (1991) misinterpreted Leslie's sense of metarepresentation and that Leekam and Perner's data actually support the metarepresentational account of perspective-taking.

The above example illustrates the ambiguity and confusion that arises from talking about mental processes that cannot be empirically proven. It also shows the discussions and debates between different authors in the way they interpret the same mental process. For these reasons, cognitive theories tend to be incomplete in their account of perspective-taking as they do not provide an empirical explanation as to how and why these mental processes develop.

ToM, as mentioned earlier, is an account of perspective-taking that has recently received attention in both the cognitive and behavioural fields. The ToM approach views a person's knowledge of themselves and others across five levels, from simple visual perspective-taking to understanding true- and false-beliefs (McHugh, Barnes-Holmes, Barnes-Holmes, Stewart & Dymond, 2007b). The first three levels describe early perspective-taking abilities that develop a basic understanding of the perceptions or perspectives of another (McHugh et al., 2006). Level 1 (simple visual perspective-taking) is the simplest form of perspective-taking and involves the ability to understand that different people can see different things. Level 2 (complex visual perspective-taking) builds on the abilities learnt in level 1 and involves the fact that people can see the same things differently. Level 3 (the principle of seeing leads to knowing) involves the ability to relate visual stimuli (seeing) to facts about the world (knowing); that a person knows because they have seen (McHugh et al., 2004).

Levels 4 and 5 describe more complex attributions of true- and false-belief and appropriate predictions of actions based on this information (McHugh et al., 2006). Level 4 (true-belief) involves the ability to make predictions about another's behaviour on the basis of true-belief. Level 5 (false-belief) involves the ability to make predictions about another's behaviour on the basis of false-belief and be aware that previous beliefs may have been false (McHugh et al., 2006).

According to ToM perspective-taking skills emerge over the course of typical child development and children excel on perspective-taking tasks once they are approx 5 years old (Rehfeldt et al., 2007). When all five levels of perspective taking are mastered, a person is said to be able to ‘mentally represent’ another person’s beliefs and have a ToM (McHugh, Barnes-Holmes, Barnes-Holmes, Whelan & Stewart, 2007a). Autistic populations often fail false-belief tasks (level 5 of ToM) therefore are said to lack a ToM.

Relational Complexity (RC) theorists suggest that performance on ToM tasks is predicted by the number of variables that are to be related (Haldord, Wilson, & Phillips, 1998). Andrews, Halford, Bunch, Bowden, and Jones (2003) demonstrated that performance on false-belief tasks was predicted from non-ToM tasks that involved relating three variables. Young children have a considerable amount of knowledge about mental states but the complexity of integrating the information required to develop ToM is a developmental factor (Andrews et al., 2003). According to McHugh et al. (2007a) these results support RC as a better explanation of false-belief outcomes than ToM. RC theorists also suggest that the development of children’s relational skills is experience driven (Halford et al., 1998). However, the details of the critical experience or learning history required to develop ToM/perspective-taking remains unclear (McHugh et al., 2007a).

ToM provides a description of the steps involved in developing perspective-taking but McHugh et al. (2006) argue that ToM fails to identify the precise behavioural processes involved in perspective-taking. According to Howlin, Baron-Cohen and Hadwin (1999), perspective-taking is a skill that can be taught across increasing levels of informational states. Again it is unclear what specific learning is required in order to successfully reach these developmental milestones proposed by ToM.

Recently behavioural psychologists have investigated perspective-taking, traditionally the domain of cognitive psychology, under the rubric of Relational Frame Theory (RFT) (McHugh et al., 2006). RFT is a functional analytic approach that accounts for the development of language and higher cognition as generalized operant behaviour (McHugh et al., 2007b). The basis of RFT lies in the idea that deriving stimulus relations is learned behaviour (Hayes, Barnes-Holmes, & Roche, 2001). Development occurs as a result of generalized patterns of arbitrarily applicable relational responding.

Sidman (1971) conducted the first behaviour analytic experiment demonstrating stimulus equivalence with a learning disabled subject. In this experiment the subject who could verbalise pictures was then taught to match spoken words to printed works. Then without training he was able to spontaneously match printed words to pictures and verbalise printed words. These untrained relations are referred to as derived stimulus relations.

The emergence of untrained behaviour occurs in many different situations. Horne and Lowe (1998) report that stimulus equivalence is a term that originated from equivalence classes in mathematics, that includes the relations of reflexivity, symmetry, and transitivity, and that describes the relation between pairs of elements in a set as equivalent relations. To illustrate stimulus equivalence in its simplest form with a three stimulus equivalence class, a participant is trained on two conditional discriminations, to match stimulus A with stimulus B, and then trained to match stimulus B with stimulus C. A, B and C, are simply names applied to stimuli for ease of talking about relations. For example, stimulus A may be a picture of a car, stimulus B the written word ‘car’, and stimulus C the spoken word ‘car’. The stimuli do not have to be different forms of the same response class, for example, A B and C could be three different nonsense syllables, or Chinese characters. For stimulus equivalence to have occurred the participant will demonstrate seven emerged relations that show transitivity, the ability to match stimulus A with stimulus C even though the relationship has not actually been trained; reflexivity, the ability to match stimulus A to A, B to B and C to C; symmetry, the ability to match B to A, and C to B; and transitivity symmetry, the ability to match stimulus C to A (Tierney & Bracken, 1998).

Relational frames do not have to be equivalent as above, but can involve relations with formal properties or nonarbitrary relations such as more than, less than, same and opposite.

McHugh et al. (2007a) say that the class or type of relational responding identified in perspective-taking is known as deictic frames or relations (I-You, Here-There, and Now-Then). This frame of deictic relations is unlike other relations as it does not appear to have formal or nonarbitrary counterparts (McHugh et al., 2007b). In order to accurately understand the perspective of oneself or another accurately, an individual must consider all three deictic relations. For example, consider the statement “If you put the pencils in the smarties box and I am there”, “You would think the smarties box contained pencils”. In this true statement the individual must consider the deictic relations; it is ‘I’ that put the pencils in the smarties box ‘here’ and ‘now’. Although each situation requires that each deictic frame be considered, not all situations are equally complex. There is an important distinction between the perspective of self (I) and other (you) in terms of relational complexity (McHugh et al., 2007a). In situations where an individual takes the perspective of self they must consider the deictic relation of I-you, whereas taking the perspective of another involves the combination of the deictic relation of I-you and an if-then relation. The individual must derive ‘if I were you, then I would …” in order to take the perspective of the other person.

According to McHugh et al. (2004) deictic relations emerge as part of our learning history through the responding to questions such as “What was I doing there?” and “What are you doing now?” Instances of perspective-taking do not have to involve the specific words, I, you, now, then, here, or there. Instead they will often be substituted for other relevant words that participate in the frame such as individuals, places or time, for example “It’s one o’clock and I am home (here and now), but Molly (you) is still at playschool (there and now).” (McHugh et al., 2004, p.166). It is not the exact words that are important, but the generalized relational properties between the deictic relations that remains constant. These properties appear to be learned through talking about ones own perspective in relation to others. For example, ‘I’ is always talked about from the perspective of ‘here’ (Hayes et al., 2001).

Recent research supports the Behavioural Analytic account of perspective-taking and unlike the Cognitive or ToM theories, the Behavioural Analytic account has empirical evidence supporting this theory’s development of perspective-taking skills. The two major theories in the literature, ToM and RFT, have the same goal of being able to teach perspective-taking skills to those in which they are absent. However, they differ in the suggestions for the most effective method to establish these skills. According to ToM, attempts are made to teach children to understand informational states defined by ToM, whereas according to RFT the most effective method of teaching would be to target the relational frames directly (McHugh et al., 2004).

Recent Research

The literature surrounding RFT is well established in the areas of language and higher cognition, and the support for its use in understanding perspective-taking appears promising.

McHugh et al. (2004) established that perspective -taking from a RFT perspective follows a developmental trend. In this study five different age groups (adulthood 18-30years; adolescence 12-14years; late childhood 9-11years; middle childhood 6-8years; and early childhood 3-5years) were given the same protocol that assessed their abilities to respond to perspective-taking tasks. Findings showed that task accuracy increased as a function of age. Studies 2 and 3 ruled out the possibility that low rates of accuracy recorded by the youngest groups were a function of the length of statements, and that the high rates of accuracy recorded for adult participants were a function of cueing. Overall these results support the RFT account of perspective -taking as generalized operant behaviour.

RFT states that generalized operant behaviour is learned. One indicator of learning is a decrease in response latencies on the task being learned. O'Hora, Roche, Barnes-Holmes, and Smeets (2002) have shown that training with multiple exemplars reduces response latencies for normally developing college students in relational frames of more-than, less-than, and same-opposite. Participants response latencies decreased with extended training (i.e., the participants got faster at responding after training), and were able to generalize to novel stimuli. Further results, which were also consistent with those of Steele and Hayes (1991), showed that response latencies increase as a function of task complexity.

Relatively few recent studies in the area of RFT have used response latency or response time as a response measure, and much of the recent research in the area of perspective taking has used accuracy as a response measure (e.g., Rehfeldt et al., 2007; McHugh et al., 2007b; McHugh et al., 2006). Spencer and Chase (1996) have argued that response latencies are a more sensitive measure than response accuracy because changes in trials measured by response latencies can still be observed when response accuracy has stabilized.

McHugh et al. (2007a) conducted one of the few recent studies in the area of perspective taking that used response latency as a response measure. In this study they investigated the role of relational complexity within the frame of deictic relations through the use of false-belief tasks. The understanding of false-belief tasks has been likened to perspective-taking because in order to complete the task accurately participants must be able to derive the perspective of another individual. Twenty participants completed a block of trials differentiated along three primary dimensions (1) perspective-taking (Self, Other, and Photograph); (2) belief (true-belief or false-belief); (3) statement type (whether the response was true or false). The findings showed that the main effect for perspective was significant; response latencies for photograph were shorter than for Self, and Self response latencies were shorter than for Other. These results were consistent with predictions made by RFT.

McHugh et al. (2007a) based their task on that of Sabbagh and Taylor (2000) who presented normally developed college students with a short narrative that described a character's belief about the location of an object that had been relocated in his or her absence. There were two conditions, one which involving a photograph being taken, and the other was simply the belief. McHugh et al. (2007a) used these two conditions and also included the self condition as a measure of control. Sabbagh and Taylor (2000) reported that the belief and photograph conditions were equivalent in terms of reading difficulty and comprehension as there was no significant difference in accuracy or length of time to read each statement between the two conditions.

To date, no research has investigated the effects of training using multiple exemplars and feedback on perspective-taking tasks. The nature of the deictic relations involved in perspective-taking is such that they are unlike other relational frames as they do not have non-arbitrary properties (McHugh et al., 2007a). There are no physical or non-arbitrary properties on which to make specific rules that can guide an individual when responding to perspective-taking situations. For an individual to accurately respond to a perspective-taking task they must understand and be able to extract the relevant information about the relations involved in the task. In normal development it is thought that children develop perspective-taking skills by receiving feedback for responding to questions such as "What am I doing now?" or "What were you doing then?" (McHugh et al., 2007a). If, during normal development, children learn perspective-taking through a reinforced learning history then this would suggest that generating a learning history for individuals with perspective-taking deficits may increase their perspective-taking skills. O'Hora et al. (2002) has shown that the use of multiple exemplars during training has successfully increased performance on relational frame tasks that involve non-arbitrary properties. To date there has been no research investigating if this same method of training is effective at improving performance on perspective-taking tasks.

Taking into account the above information, the aim of the current research was to replicate the findings of McHugh et al. (2007a) that relational complexity increases response latencies on perspective-taking tasks, and to investigate the effects of training on response latencies. To date, no research in this area has investigated the effects of repeated trials on response latency. It could be that the reaction time would get shorter over time regardless of any learning about the types of relation, as it is possible that the participants would get faster at simply responding to the task. Thus, the current research included a simple reaction time test repeated throughout the experiment to investigate the effect of repeated trials on performance on this task. The aim was to examine changes in time taken to complete this simple task (that required similar mechanical response to the perspective-taking task). These were to be compared with any changes in the time taken to complete the perspective-taking task. It was thought that the reaction time task would take less time than the perspective-taking task as it would be easier. However, it was expected that extended training would reduce response latencies on both tasks, so the response latencies on the reaction time task would serve as a measure of the time it takes just to respond, as it does not include the time taken to react to the various perspective-taking questions.

RFT predicts a number of outcomes (1) that response latencies will be longer as a function of relational complexity; (2) that extended training on perspective-taking tasks will decrease response latencies; and (3) that generalization to novel stimuli will occur. It is possible that (4) response latencies on the reaction time tests will decrease over time, but that these decreases in the time it takes to do this task will not account for all the changes in reaction time found for the perspective-taking task, although there is no data available on which to base this.

Method

Ethics and Participant Recruitment

Ethics approval was gained from The University of Waikato Psychology Department Research and Ethics Committee. Participants were recruited through notices (Appendix A) presented to Waikato University students enrolled in courses PSYC102 and PSYC103 during laboratories, tutorials, Moodle, and through notices displayed on notice boards around the university campus. Participants enrolled in first year psychology courses (PSYC102 or PSYC103) received 1% course credit for each hour of participation.

Participants

There were 22 participants that took part in this research. All were students enrolled in PSYC103-08B at the University of Waikato and received 2% course credit for their participation.

Setting

The experiment was carried out in a small windowless room (approx 3m by 6m), containing a desktop computer, desk, and computer chair. Each participant was alone in the room for the duration of the experiment.

Apparatus

The experiment was run from a server on the University of Waikato's network system. The automated programme was presented to the participants on a 17in monitor, and controlled by the participant with a standard keyboard and mouse.

Stimuli

The stimuli presented during the Reaction Time Tests consisted of three pairs of colours, red/blue, yellow/green, and pink/orange and these were termed Set 1. Each trial consisted of a statement “This is … (name of colour)” followed by a coloured square.

The perspective-taking stimulus sets were based on those used by McHugh et al. (2007a). Two stimulus sets, Set 2 cookie jar/doll/cookie, and Set 4 smarties box/pencil/smarties, were taken exactly from McHugh et al. (2007a). The remaining three stimulus sets were derived based on properties required for the task, and were Set 3 toy box, pizza, toy, Set 5 dog kennel, tv, dog, and Set 6 money box, paper clip, money. The stimuli used for the perspective-taking tasks differentiated along three primary dimensions: (1) perspective-taking (Self, Other, and Photograph); (2) belief (true-belief or false-belief); (3) statement type (whether the response was true or false). Each trial consisted of the presentation of two statements, the second statement was about the first statement.

A full list of trials for all Sets can be found in Appendix D.

Procedure

This research consisted of a within subjects design. All participants completed the same task.

Upon arrival to the experimental room, participants received an information sheet (Appendix B) outlining the requirements of the experiment. The participants were verbally given the opportunity to ask questions, and were given a consent form to sign (Appendix C). The experiment proper began at the completion of the above administration forms.

The experiment consisted of seven phases with a total of 216 trials taking approximately one and a half hours to complete and the phases were presented in the following order.

Reaction Time Test

↓

Perspective-taking Pre-test

↓

Reaction Time Test

↓

Perspective-taking Training

↓

Reaction Time Test

↓

Perspective-taking Post-test

↓

Reaction Time Test

Participants were seated at the computer that was displaying the information screen for the first phase of the experiment in the centre of a white background with black Arial font size 12.

> "In this phase of the experiment you will be presented with a statement and a coloured square. Your task is to read the statement that will be presented word by word on the screen, look at the coloured square below the statement, and then choose your answer. If you think the statement is true, press the 'Z' key on the left of the keyboard. If you think the statement is false, press the 'M' key on the right of the keyboard. The screen will then go grey. When you are ready for the next trial, press the space bar to continue. Please answer each question as accurately and as fast as you can. Once you have made a response you will not be able to change you answer. To begin this task, press the space bar"

The Reaction Time Test consisted of 12 trials each presented once in a random order. Trials were counter balanced so that there were three pairs of colours each with a true and false correct response. For example the colour pair of red/blue consisted of four trials, "This is red" and red square; "This is red" and a blue square; "This is blue" and a blue square; "This is blue" and a red square.

The other pairs of colours used were yellow/green and pink/orange (Set 1, Appendix D).

The screen for each trial consisted of a full sized white background with the statement appearing word by word every 0.512s so that the final statement was in the centre horizontally in the top part of the screen in black Arial font size 16. The coloured square followed 0.512s after the last word of the statement and appeared in the bottom middle of the screen measuring approx 4x4cm on a 17in monitor with resolution 1024x768.

The statement and coloured square remained on the screen until the participant made a response by press the 'Z' or 'M' key. Immediately following a response a grey intermediate screen would appear with the words 'Press the space bar to continue in the centre of the screen in black Arial font size 16. The next trial began immediately after the participant pressed the space bar. At the completion of the 12 trials the following screen appeared in the centre on a white background with black Arial font size 12.

> "This stage of the experiment is now complete. Press the space bar twice to continue onto the next phase of the experiment"

The second phase of the experiment was the perspective-taking Pre-test. Participants were presented with the following information screen on a white background with black Arial font size 12 in the centre of the screen.

> “In this phase of the experiment you will be presented with TWO statements. The second statement will be about the original statement. Your task is to read the statements which will be presented word by word on the screen, and press the ‘Z’ key on the left if you think that the second statement is true, or press the ‘M’ key on the right if you think the second statement is false. The screen will then go grey. When you are ready for the next trial, press the space bar to continue. Please answer each question as accurately and as fast as you can. Once you have made a response you will not be able to change you answer. To begin this task, press the space bar”

The perspective-taking Pre-test consisted of a total of 48 trials made up from two stimulus sets each containing 24 trials that were each presented once in a random order without replacement followed by a second presentation of each trial in a random order without replacement. The stimulus sets used were cookie jar/doll/cookies (Set 2), and toy box/pizza/toy (Set 3) (Appendix D).

Each test screen consisted of a full sized white background with the two statements appearing word by word every 0.512s so that the final two statements appeared in the centre horizontally in the top part of the screen in black Arial font size 16. The second statement immediately followed the first statement and there were two blank lines separating the two statements.

The statements remained on the screen until the participant made a response by pressing the ‘Z’ or ‘M’ key. Immediately following a response a grey intermediate screen would appear with the words ‘Press the space bar to continue in the centre of the screen in black Arial font size 16. The next trial began immediately after the participant pressed the space bar. At the completion of the 48 trials the following screen appeared in the centre on a white background with black Arial font size 12.

> “This stage of the experiment is now complete. If you require a short break please see the experimenter. Press the space bar twice to continue onto the next phase”

The third phase of the experiment was an exact replication of the Reaction Time Test described above.

The fourth phase of the experiment was the perspective-taking Training. Participants were presented with the following information screen on a white background with black Arial font size 12 in the centre of the screen.

> “In this phase of the experiment you will be presented with two statements similar to those presented in the second phase. Again the second statement will be about the original statement, and your task is to decide if the second statement is true or false by pressing the ‘Z’ on the left or ‘M’ on the right respectively. You will then receive

> feedback on your response. The screen will then go grey. When you are ready for the next trial, press the space bar to continue. Please answer each question as accurately and as fast as you can. Once you have made a response you will not be able to change you answer. To begin this task, press the space bar"

The perspective-taking Training was the same as the perspective-taking Pre-test with only two differences. The first was the stimulus sets presented. These were smarties box/pencil/smarties (Set 4), and dog kennel/TV/dog (Set 5) (Appendix D). The second difference was that this phase of the experiment provided the participants with feedback on their responses. Following a response on either the 'Z' or 'M' key the screen would immediately present the feedback screen for the duration of 2s. For a correct response this was a white background with the word 'Correct' presented in the centre of the screen in black Arial font size 72, and be accompanied by a tone of approximately 1000Hz. For an incorrect response this was a white background with the word 'Wrong' presented in the centre of the screen in black Arial font size 72 alone without a tone. Following the feedback, the intermediate screen appeared as in the perspective-taking Pre-test phase. At the completion of the 48 trials the participants were presented with the same completion statement as at the end of the perspective-taking Pre-test phase.

The fifth phase of the experiment was an exact replication of the Reaction Time Test described above.

The sixth phase of the experiment was the perspective-taking Post-test. This was a replication of the perspective-taking Pre-test with the only difference being the stimuli presented. The stimulus sets were cookie jar/doll/cookies (Set 2) (same as Pre-test), smarties box/pencil/smarties (Set 4) (same as Training), and a novel set money box/paper clip/money (Set 6) (Appendix D). Therefore this phase consisted of 72 trials made up from three stimulus sets each containing 24 trials that were each presented once in a random order without replacement followed by a second presentation of each trial in a random order without replacement. At the completion of the 72nd trial the participants were presented with the same completion statement as at the end of the Pre-test and Training phases.

The final phase of the experiment was an exact replication of the Reaction Time Test described above. Upon completion participants were presented with the following statement.

> "This part of the experiment is now complete. Thank you for your participation. Please report to the experimenter."

Results

Accuracy

There were twenty-two participants in this study. Participant 22 did not gain a full set of data as a result of a computer error during the Post-test phase of the experiment so their data are not presented. Table 1 shows the percentage correct over all phases of the experiment for the remaining 21 participants. Nineteen out of the 21 participants responded with a mean accuracy of 92.6%-100% across all phases. Participant 3 responded with an overall accuracy of 77.3%, and Participant 21 responded with an overall accuracy of 60.2%. Raw data for all participants can be found in Appendix K1-K22.

Investigation of Participant 3's responses in Appendix K3 revealed that the majority of errors (41 out of 49 incorrect responses) were made on three types of question, 14 Self/False/False questions, 13 Other/False/True questions, and 14 Other/False/False questions. Further analysis of mean response latencies derived from Participant 3's raw data in Appendix K3 showed that mean latencies of the incorrect responses were similar to the total means for each of the perspective-taking phases (the mean of all the response latencies in the Pre-test phase was 1.23s and the mean of the incorrect response latencies in the Pre-test phase was 1.41s; the mean of all the response latencies in the Training phase was 0.63s and the mean of the incorrect response latencies in the Training phase was 0.72s; and the mean of all the response latencies in the Post-test phase was 0.71s and the mean of the incorrect response latencies in the Post-test phase was 0.75s). Participant 3's data was kept in the group analysis, as the incorrect responses did not appear to have an effect on the overall response latencies.

Table 1

Percentage of Correct Trials Across All Phases of the Experiment for Each Participant

Participant	Reaction Time 1	Pre-test	Reaction Time 2	Training	Reaction Time 3	Post-test	Reaction Time 4	**Mean**
1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	**100.0**
2	91.7	95.8	91.7	100.0	100.0	100.0	100.0	**98.1**
3	100.0	72.9	100.0	72.9	100.0	68.1	100.0	**77.3**
4	100.0	93.8	100.0	85.4	91.7	95.8	100.0	**93.5**
5	100.0	100.0	100.0	95.8	100.0	98.6	100.0	**98.6**
6	91.7	95.8	91.7	97.9	91.7	100.0	91.7	**96.8**
7	100.0	87.5	100.0	97.9	100.0	97.2	100.0	**95.8**
8	100.0	95.8	100.0	97.9	100.0	98.6	91.7	**97.7**
9	100.0	91.7	100.0	93.8	91.7	97.2	91.7	**94.9**
10	91.7	91.7	100.0	97.9	100.0	95.8	100.0	**95.8**
11	100.0	100.0	91.7	97.9	91.7	98.8	91.7	**97.7**
12	91.7	100.0	75.0	93.8	91.7	97.2	91.7	**94.9**
13	91.7	100.0	100.0	100.0	100.0	97.2	100.0	**98.6**
14	100.0	97.9	100.0	93.8	100.0	95.8	100.0	**96.8**
15	100.0	95.8	100.0	100.0	100.0	95.8	100.0	**97.7**
16	100.0	100.0	100.0	100.0	91.7	98.8	100.0	**99.1**
17	83.3	91.7	75.0	91.7	100.0	97.2	91.7	**92.6**
18	100.0	89.6	83.3	95.8	100.0	100.0	91.7	**95.4**
19	91.7	81.3	91.7	97.9	91.7	97.2	100.0	**93.1**
20	91.7	100.0	100.0	95.8	91.7	100.0	100.0	**98.1**
21	91.7	47.9	83.3	62.5	91.7	45.8	100.0	**60.2**
Mean	**96.0**	**91.9**	**94.4**	**93.7**	**96.4**	**94.1**	**97.2**	**93.9**

Participant 21 responded with low accuracy throughout the experiment gaining only 52.0% correct responses over the perspective-taking phases (Table 1). Investigation of Participant 21's responses revealed 86 errors over the course of the experiment (Appendix K21). Eighty two errors occurred during the perspective-taking phases and this participant showed poor performance on all types of questions. Participant 21's data were excluded from further analysis as their high number of errors indicated that they did not show mastery of the task, as would be expected when using response latency as a measure. Thus there were 20 participants' data for analysis.

Response Latencies

The response latencies for each trial across the seven phases of the experiment for each of the remaining 20 participants are shown in Figure 1. The different symbols indicate the different phases; filled symbols indicate reaction time phases and open symbols indicate perspective-taking phases. The y axis is taken to 10s and 14 participants had all response latencies below this. Participants 5, 8, 10, 13, 15, and 19 each had response latencies above 10s which are indicated in Figure 1 as **x** on the x axis.

Twelve participants' response latencies showed a general downward trend across the three perspective taking phases, Pre-test, Training, and Post-test. This was evident for Participants 1, 2, 4, 6, 7, 10, 11, 12, 13, 16, 17, and 20. Four participants' (3, 9, 15, and 19) response latencies showed trends with the latencies in the Training phase being shorter than those in the Pre-test and Post-test phases. Four participants' (5, 8, 14, and 18) response latencies appeared to be fairly consistent over the three perspective taking phases.

The reaction time phases response latencies for the majority of participants (excluding Participants 8, 12, 14, and 18) were more scattered or longer for Reaction Time Test 1 than for Reaction Time Test 2, 3, and 4. Participants 8, 12, 14, and 18 response latencies were fairly consistent across all four reaction time tests.

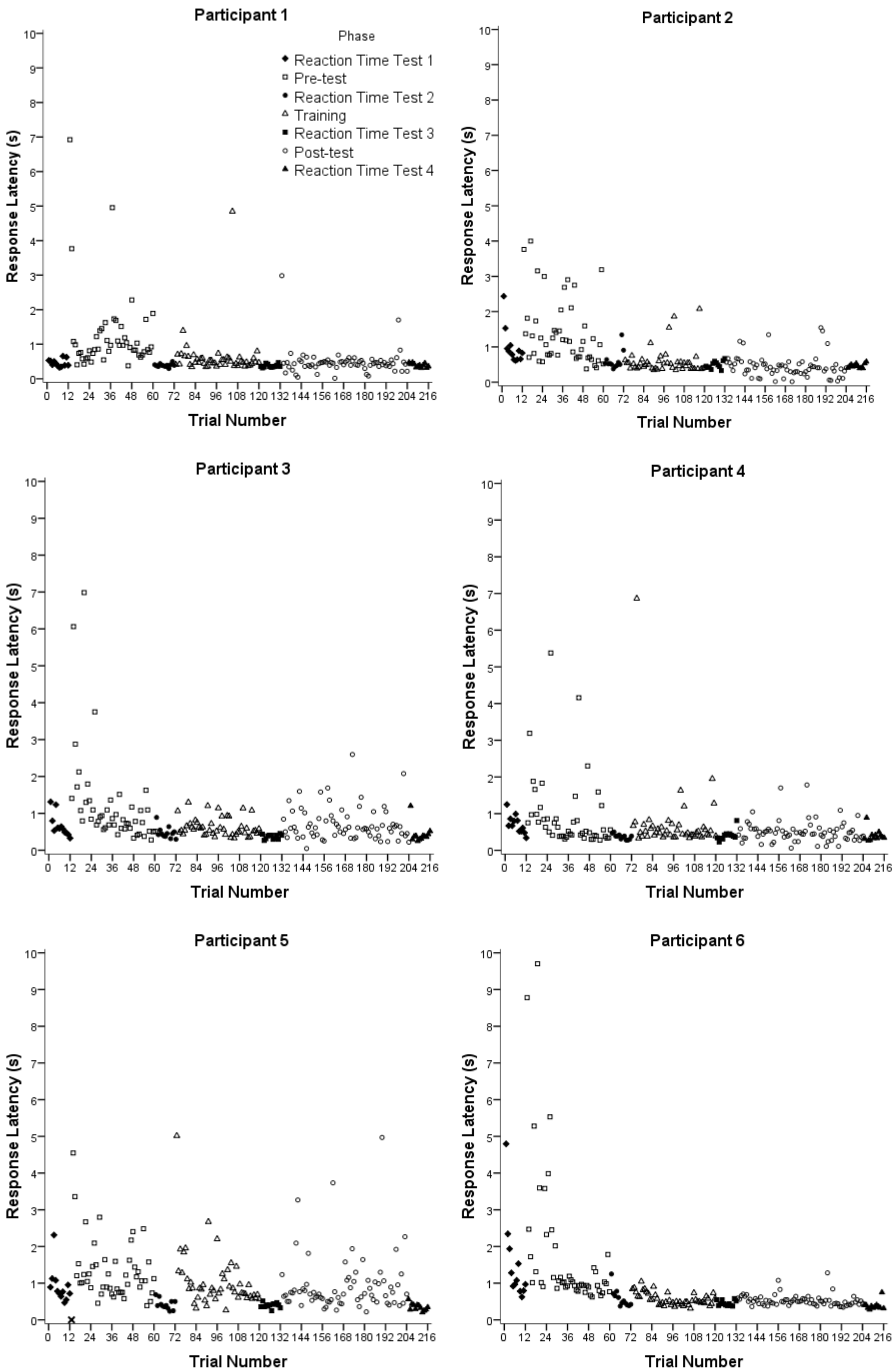

Figure 1 continued on the next page

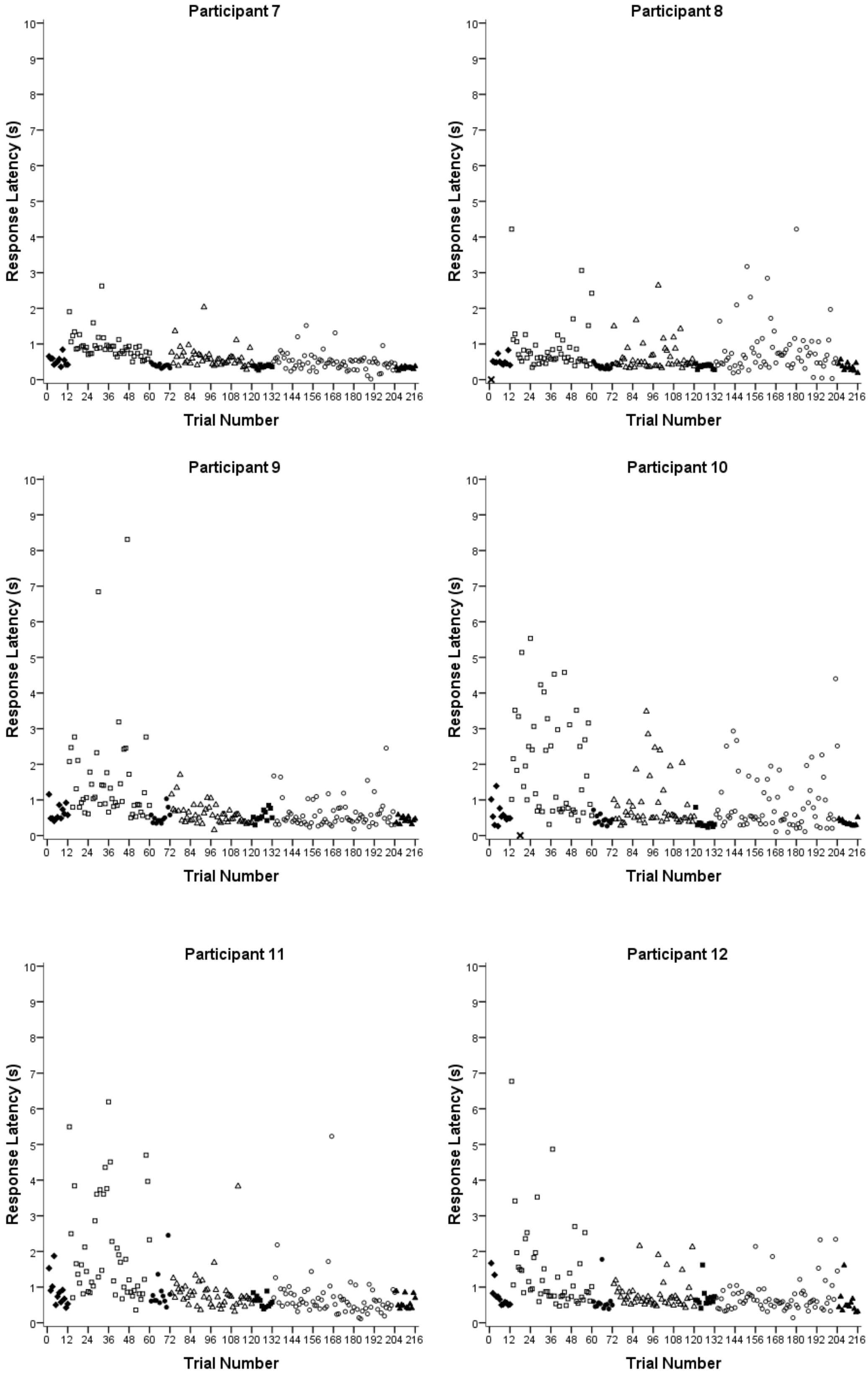

Figure 1 continued on the next page

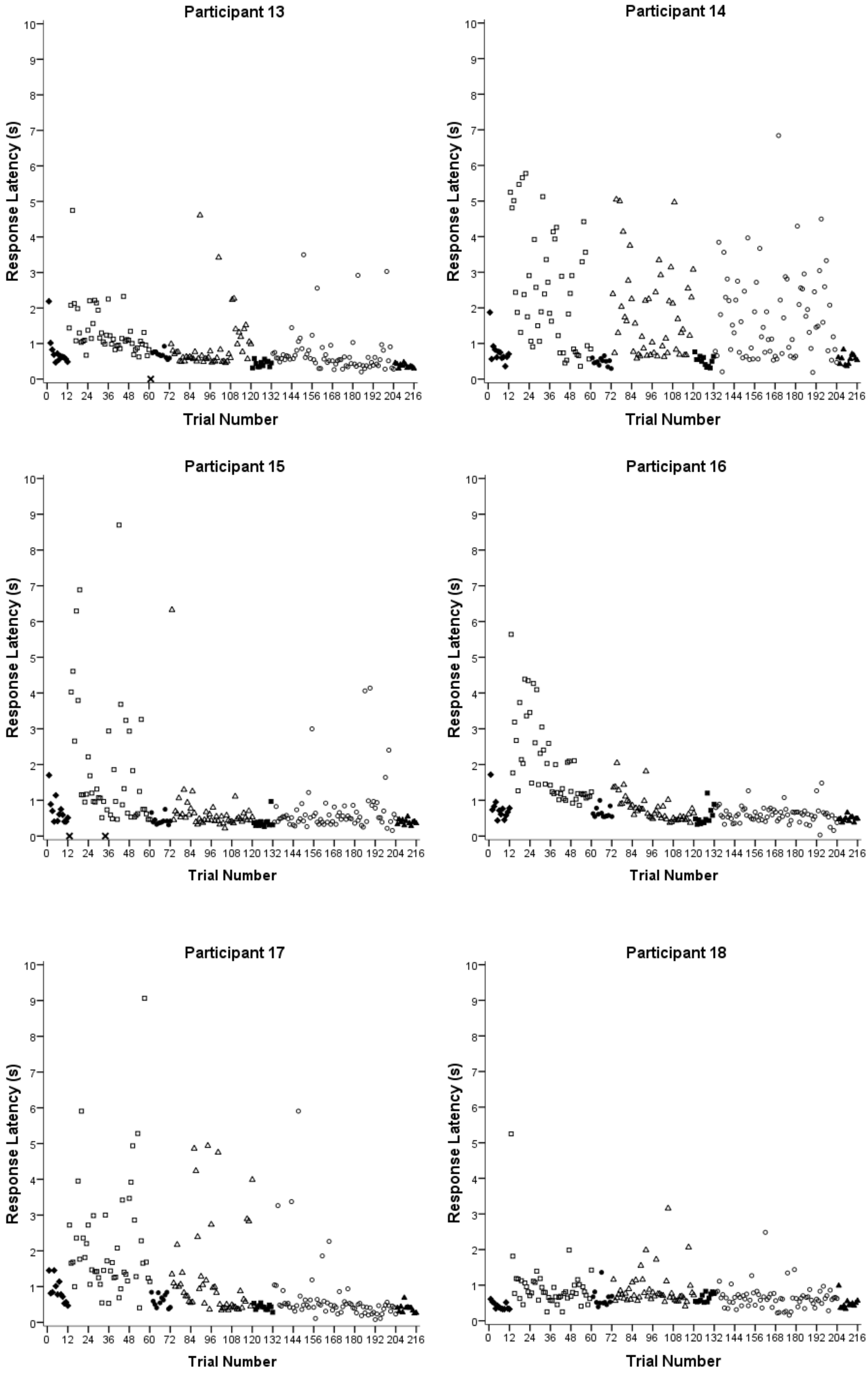

Figure 1 continued on the next page

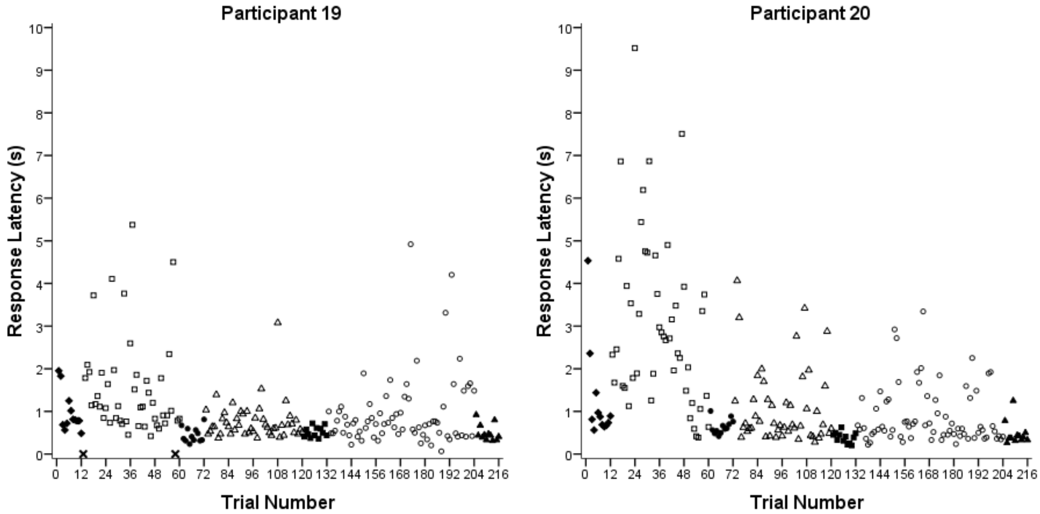

Figure 1: Response Latencies for all Trials across all Phases (and Response Latencies exceeding 10s indicated by **x** on the x axis) of the Experiment for all Participants

Means and Standard Deviations of Individual Data

Examination of individual response latencies (Appendix K1-K20) and Figure 1 showed that Participant 13 had a clear outlier of 56.45s; this was adjusted to the next longest reaction time response latency of 2.18s prior to any further calculations. The means and standard deviations of the response latencies across the perspective-taking phases for all participants are shown in Table 2. The mean response latencies across all participants and perspective-taking phases ranged from 0.44s to 3.05s. The mean response latencies for the Pre-test were longer than those for the Training and Post-test phases for all participants except Participant 8 whose longest mean response latency was for the Post-test phase. Participants 3, 8, 10, 15, and 19 had longer mean response latencies for the Post-test than for Training. Participant 9 had equal mean response latencies for the

Training and Post-test phases. All other participants had longer mean response latencies for Training than for the Post-test.

Table 2

Means and Standard Deviations (in brackets) of the Response Latencies (in Seconds) for the Perspective-taking Phases

Participant	Pre-test	Training	Post-test
1	1.25 (1.16)	0.62 (0.65)	0.48 (0.37)
2	1.38 (0.92)	0.59 (0.36)	0.44 (0.31)
3	1.23 (1.29)	0.63 (0.25)	0.71 (0.45)
4	0.95 (1.01)	0.72 (0.96)	0.51 (0.30)
5	1.55 (1.65)	1.04 (0.77)	0.95 (0.79)
6	1.81 (1.93)	0.57 (0.17)	0.53 (0.15)
7	0.93 (0.36)	0.62 (0.29)	0.50 (0.25)
8	0.89 (0.71)	0.64 (0.44)	0.90 (1.27)
9	1.55 (1.45)	0.60 (0.29)	0.60 (0.39)
10	2.25 (2.14)	0.87 (0.73)	0.89 (0.81)
11	1.99 (1.45)	0.82 (0.52)	0.70 (0.65)
12	1.43 (1.20)	0.81 (0.41)	0.73 (0.44)
13	1.33 (0.69)	0.93 (0.78)	0.72 (0.61)
14	2.45 (1.64)	1.84 (1.25)	1.74 (1.23)
15	2.65 (4.35)	0.69 (0.86)	0.71 (0.72)
16	2.04 (1.15)	0.77 (0.37)	0.60 (0.23)
17	2.17 (1.59)	1.31 (1.30)	0.66 (0.85)
18	0.94 (0.72)	0.88 (0.49)	0.65 (0.33)
19	1.96 (2.36)	0.76 (0.43)	0.95 (0.83)
20	3.05 (2.01)	1.09 (0.89)	0.89 (0.68)
Mean	**1.69 (1.49)**	**0.84 (0.61)**	**0.74 (0.58)**

The means and standard deviations of the response latencies across the reaction time phases for all participants are shown in Table 3. All participants except Participants 13 and 18 had longer mean response latencies for Reaction Time Test 1 than for Reaction Time Tests 2, 3, and 4. Participant 13's longest mean response latency was for Reaction Time Test 2, and Participant 18's longest mean response latency was for Reaction Time Test 3. Mean response latencies across all participants and reaction time tasks ranged from 0.33s to 1.50s. The standard deviations for the reaction time phases (Table 3) were generally much smaller for all participants than those for the perspective-taking phases (Table 2).

Table 3

Means and Standard Deviations (in brackets) of the Response Latencies (in Seconds) for the Reaction Time Phases

Participant	Reaction Time 1	Reaction Time 2	Reaction Time 3	Reaction Time 4
1	0.46 (0.11)	0.39 (0.05)	0.37 (0.05)	0.39 (0.06)
2	0.99 (0.52)	0.61 (0.27)	0.49 (0.10)	0.46 (0.06)
3	0.67 (0.31)	0.48 (0.16)	0.37 (0.06)	0.43 (0.25)
4	0.71 (0.25)	0.37 (0.07)	0.41 (0.14)	0.40 (0.17)
5	0.92 (0.48)	0.42 (0.14)	0.39 (0.07)	0.35 (0.10)
6	1.50 (1.16)	0.59 (0.25)	0.45 (0.07)	0.39 (0.12)
7	0.53 (0.13)	0.39 (0.05)	0.37 (0.04)	0.33 (0.04)
8	1.47 (3.28)	0.37 (0.07)	0.37 (0.06)	0.36 (0.11)
9	0.64 (0.23)	0.53 (0.20)	0.56 (0.16)	0.44 (0.09)
10	0.61 (0.31)	0.44 (0.13)	0.34 (0.15)	0.36 (0.08)
11	0.88 (0.43)	0.87 (0.55)	0.61 (0.16)	0.57 (0.19)
12	0.76 (0.37)	0.61 (0.37)	0.71 (0.31)	0.56 (0.35)
13	0.78 (0.47)	0.82 (0.44)	0.43 (0.09)	0.37 (0.06)
14	0.77 (0.38)	0.48 (0.10)	0.51 (0.14)	0.55 (0.14)
15	0.71 (0.38)	0.44 (0.13)	0.39 (0.19)	0.39 (0.07)
16	0.77 (0.33)	0.65 (0.15)	0.56 (0.26)	0.48 (0.08)
17	0.88 (0.33)	0.62 (0.17)	0.43 (0.08)	0.38 (0.12)
18	0.41 (0.10)	0.63 (0.26)	0.65 (0.12)	0.49 (0.17)
19	0.97 (0.47)	0.45 (0.18)	0.52 (0.12)	0.50 (0.20)
20	1.26 (1.14)	0.64 (0.17)	0.39 (0.13)	0.48 (0.28)
Mean	**0.84 (0.56)**	**0.54 (0.20)**	**0.47 (0.13)**	**0.43 (0.14)**

Question Types (Self, Other, and Photo)

The means and standard deviations of the response latencies for the three types of questions (Self, Other, and Photo) across all phases are shown in Table 4. The range of mean response latencies for the Self, Other, and Photo questions, were 0.66s-2.06s, 0.64s-2.12s, and 0.63s-1.73s respectively. Participants 3, 7, 8, 9, 11, 12, 13, 14, 17, 18, 19, and 20 had their shortest mean response latency for the Photo questions. Participants 10 and 15 had their shortest mean response latency for the Self questions. Participants 1, 2, 4, 5, and 6 had their shortest mean response latency for the Other questions and Participant 16 had their shortest mean response latency for both the Other and Photo questions.

Table 4

Means and Standard Deviations (in brackets) of the Response Latencies (in Seconds) for the Question Type (Self, Other, and Photo) Across All the Perspective-taking Phases

Participant	Self	Other	Photo
1	0.82 (1.00)	0.65 (0.43)	0.75 (0.90)
2	0.86 (0.76)	0.68 (0.54)	0.72 (0.76)
3	0.87 (0.55)	0.99 (1.18)	0.64 (0.41)
4	0.77 (1.03)	0.64 (0.48)	0.67 (0.76)
5	1.15 (1.00)	1.05 (0.65)	1.24 (1.54)
6	0.91 (1.26)	0.80 (0.83)	1.01 (1.39)
7	0.66 (0.43)	0.68 (0.35)	0.63 (0.23)
8	1.01 (1.43)	0.79 (0.69)	0.67 (0.36)
9	0.91 (1.16)	1.00 (1.01)	0.71 (0.48)
10	1.16 (1.08)	1.30 (1.20)	1.36 (1.94)
11	1.11 (0.97)	1.13 (1.02)	1.08 (1.24)
12	0.93 (0.61)	1.10 (1.08)	0.82 (0.57)
13	1.00 (0.82)	1.02 (0.79)	0.84 (0.54)
14	2.06 (1.31)	2.12 (1.58)	1.73 (1.25)
15	1.16 (1.56)	1.17 (2.12)	1.45 (3.58)
16	1.12 (1.04)	1.03 (0.91)	1.03 (0.78)
17	1.38 (1.68)	1.26 (1.34)	1.19 (1.05)
18	0.79 (0.66)	0.84 (0.52)	0.76 (0.37)
19	1.20 (1.03)	1.31 (1.86)	1.03 (1.41)
20	1.56 (1.54)	1.75 (1.63)	1.38 (1.52)
Means	**1.07 (1.05)**	**1.06 (1.01)**	**0.99 (1.05)**

Sets

There were six different sets of questions used in the three perspective-taking phases. Two sets of questions were presented twice. Set 2 was used in the Pre-test and Post-test, and Set 4 was used in Training and the Post-test. Set 6 was a novel set presented in the Post-test and was used to test for generalisation. The means and standard deviations of the response latencies for each set for each phase are shown in Table 5.

Table 5

Means and Standard Deviations (in brackets) of the Response Latencies for Each Set in the Perspective-taking Phases (Pre-test, Training, and Post-test)

Participant	Pre-test Set 2	Set 3	Training Set 4	Set 5	Post-test Set 2	Set 4	Set 6
1	1.19(0.91)	1.31(1.39)	0.56(0.15)	0.68(0.91)	0.46(0.13)	0.51(0.27)	0.47(0.58)
2	1.52(0.98)	1.25(0.86)	0.61(0.39)	0.56(0.33)	0.37(0.17)	0.61(0.35)	0.35(0.33)
3	1.06(0.74)	1.41(1.67)	0.66(0.27)	0.60(0.22)	0.72(0.38)	0.79(0.51)	0.63(0.47)
4	0.99(0.74)	0.91(1.23)	0.60(0.37)	0.84(1.31)	0.49(0.24)	0.56(0.28)	0.48(0.37)
5	1.76(2.15)	1.34(0.91)	1.18(1.00)	0.90(0.41)	0.93(0.69)	0.91(0.47)	1.02(1.10)
6	1.47(1.16)	2.14(2.46)	0.59(0.19)	0.55(0.15)	0.50(0.15)	0.59(0.19)	0.49(0.08)
7	0.89(0.25)	0.98(0.44)	0.63(0.34)	0.62(0.24)	0.46(0.21)	0.60(0.29)	0.43(0.20)
8	0.87(0.65)	0.91(0.77)	0.76(0.56)	0.53(0.21)	0.71(0.41)	1.13(2.00)	0.87(0.85)
9	1.18(0.61)	1.91(1.91)	0.70(0.31)	0.49(0.23)	0.64(0.37)	0.66(0.48)	0.50(0.31)
10	2.58(2.68)	1.93(1.41)	0.99(0.86)	0.74(0.57)	0.91(0.90)	1.04(0.80)	0.72(0.73)
11	2.28(1.73)	1.71(1.07)	0.94(0.68)	0.69(0.24)	0.58(0.26)	0.72(0.42)	0.81(1.01)
12	1.41(1.31)	1.44(1.11)	0.87(0.49)	0.74(0.31)	0.76(0.50)	0.82(0.40)	0.60(0.39)
13	1.42(0.84)	1.24(0.50)	0.85(0.62)	1.00(0.92)	0.67(0.55)	0.87(0.85)	0.61(0.32)
14	2.34(1.69)	2.56(1.61)	1.86(1.23)	1.81(1.29)	1.70(1.50)	1.82(0.99)	1.72(1.21)
15	3.84(5.88)	1.45(0.99)	0.66(0.27)	0.71(1.20	0.68(0.74)	0.87(0.89)	0.59(0.48)
16	2.11(1.22)	1.96(1.10)	0.78(0.45)	0.76(0.27)	0.58(0.20)	0.62(0.15)	0.60(0.32)
17	2.38(1.88)	1.95(1.23)	1.72(1.47)	0.89(0.96)	0.63(0.72)	0.71(0.64)	0.65(1.14)
18	0.91(0.98)	0.98(0.32)	1.05(0.62)	0.71(0.20)	0.52(0.19)	0.69(0.17)	0.73(0.50)
19	2.16(2.12)	1.76(2.60)	0.80(0.30)	0.72(0.54)	0.89(0.52)	1.13(0.94)	0.81(0.96)
20	3.44(1.80)	2.66(2.16)	1.32(1.10)	0.86(0.55)	0.89(0.62)	0.97(0.73)	0.80(0.70)
Mean	**1.79(1.52)**	**1.59(1.29)**	**0.91(0.58)**	**0.77(0.55)**	**0.70(0.47)**	**0.83(0.59)**	**0.69(0.60)**

All 20 participants had longer mean response latencies for the first presentation of Set 2 during the Pre-test phase than for the second presentation during the Post-test phase. Participants 3, 8, 10, 13, 15, and 19 had longer mean response latencies for the second presentation of Set 4 during Post-test phase than for the first presentation during the Training phase and Participants 2 and 6 had equal mean response latencies for each presentation of Set 4. All other participants had longer mean response latencies for the first presentation of Set 4 during the Training phase than for the second presentation during the Post-test phase. The mean response latencies for the novel Set 6 were not consistently different across participants from those of the other sets for the Post-test phase.

Group Analysis - ANOVAs and *t*-tests

Examination of individual data (Figure 1 & Appendix K1-K20) showed several outliers. As a result group analyses for the perspective-taking tasks were originally conducted with five different data sets, Original - the original response latencies for all participants, except Participant 13's outlier of 56.45s which was adjusted; Max 10 - all outliers above 10s were adjusted to the next longest response latency for either the perspective-taking phases or reaction time phases; Max 5 – all outliers above 5s were given a response latency of 5s. Max 2.5 – all outliers above 2.5s were given a response latency of 2.5s. McHugh et al. (2007a) excluded all errors or outliers above 2.5s and the McH data set dealt with the data in the same manner. Analyses for the Max 10 and Max 5 data sets are not presented, as most of the results were similar to those of the Max 2.5 data sets, however there analyses can be found in Appendices E-J. Analyses for the remaining three data sets (Original, Max 2.5, and McH) are presented. There were only two participants (Participant 8 and 13) with response latencies in the reaction time tests above 2.5s so analysis was only conducted on the Original data set for the reaction time tasks.

Perspective-taking Phases ANOVAs and *t*-tests

The group mean response latencies and standard errors for each data set for each of the perspective-taking phases are shown in Figure 2. The Pre-test group mean response latencies are longer for all the data sets when compared to those for the Training and Post-test phases. Within each of the Training and Post-test phases the group mean response latencies and standard errors for the Original and Max 2.5 data sets are similar, whereas the group mean response latencies for the McH data set are clearly shorter.

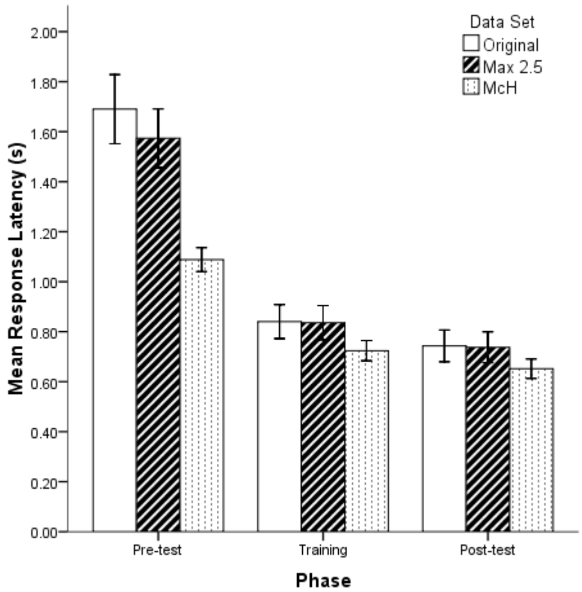

Figure 2: Group Means and Standard Errors of the Response Latencies for all the Trials from the Perspective-taking Phases (Pre-test, Training, and Post-test) for Each Data Set (Original, Max 2.5, and McH)

Repeated measures ANOVAs and *t*-tests were conducted on the response latencies for each data set for the perspective-taking phases. The ANOVAs, means and standard deviations, and *t*-test results are shown in Tables 6 and 7 respectively. The main effect of the perspective-taking phases was significant

($p<0.05$) for all the data sets and the effect sizes (partial ή²) were large according to Cohen's (1988) criteria (Table 6).

Table 6

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for all the Trials from the Perspective-taking Phases (Pre-test, Training, and Post-test) ANOVAs for the Original, Max 2.5, and McH Data Sets

Data Set	*df*	F	Partial ή²	Pre-test	Training	Post-test
Original	2,38	51.783*	0.732	1.69 (0.62)	0.84 (0.30)	0.74 (0.28)
Max 2.5	2,38	71.231*	0.789	1.33 (0.33)	0.79 (0.24)	0.71 (0.24)
McH	2,38	61.304*	0.763	1.09 (0.22)	0.72 (0.18)	0.65 (0.17)

* significant at $p<0.05$

Table 7

The df, t value, and Effect Size (Cohen's d) of the Response Latencies for all the Trials from the Perspective-taking Phases (Pre-test, Training, and Post-test) t-tests for the Original, Max 2.5, and McH Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Pre-test and Training	Original	19	7.084*	1.584
	Max 2.5	19	8.820*	1.972
	McH	19	8.112*	1.814
Pre-test and Post-test	Original	19	7.645*	1.710
	Max 10	19	8.026*	1.795
	McH	19	8.221*	1.836
Training and Post-test	Original	19	2.309*	0.516
	Max 2.5	19	2.559*	0.572
	McH	19	3.262*	0.729

* significant at $p<0.05$

Paired samples *t*-tests (Table 7) indicated that response latencies for the Pre-test phase were significantly longer than those for the Training and Post-test phases ($p<0.05$) for all the data sets and the effect sizes (Cohen's *d*) were large according to Cohen's (1992) criteria. Response latencies for the Training phase were significantly longer than those for the Post-test phase ($p<0.05$) for all the data sets and the effect sizes (Cohen's *d*) were medium.

Reaction Time ANOVAs and *t*-tests

The group mean response latencies and standard errors for the Original data set for each of the reaction time phases are shown in Figure 3. The differences between the group mean response latencies of the Reaction Time Tests successively decrease between each presentation of the Reaction Time Test, and the standard errors also decrease. Reaction Time Test 1 clearly has a longer group mean response latency and a larger standard error than Reaction Time Tests 2, 3, and 4. The group mean response latency for Reaction Time Test 2, 3, and 4 decrease across presentations. The standard error bars do not overlap over all phases. Thus, the group mean response latencies became faster and showed less variation over time.

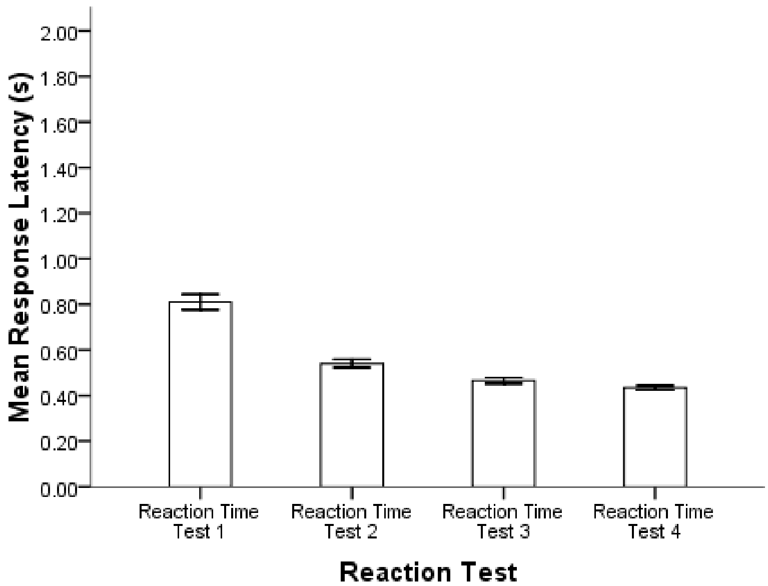

Figure 3: Group Means and Standard Errors of the Response Latencies for all the Trials from the Reaction Time Phases for the Original Data Set

Repeated measures ANOVA and *t*-tests were conducted on the response latencies for the reaction time tasks. The ANOVA, means and standard deviations, and *t*-test results for the Original data set are shown in Table 8 and 9 respectively. The main effect of reaction time was significant ($p<0.05$) and the effect size (partial ή²) was large (Table 8).

Table 8

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for the Trials from the Reaction Time Phases ANOVA for the Original Data Set

df	F	Partial ή²	Reaction Task 1	Reaction Task 2	Reaction Task 3	Reaction Task 4
3,57	23.803*	0.556	0.84 (0.30)	0.54 (0.14)	0.47 (0.10)	0.43 (0.07)

* significant at $p<0.05$

Table 9

The df, the t value and Effect Size (Cohen's d) of the Response Latencies for all the Trials from the Reaction Time Phases t-tests for the Original Data Set

	df	*t*	Cohen's *d*
Reaction Time 1 & 2	19	4.183*	0.935
Reaction Time 1 & 3	19	5.036*	1.126
Reaction Time 1 & 4	19	5.815*	1.300
Reaction Time 2 & 3	19	2.679*	0.599
Reaction Time 2 & 4	19	3.772*	0.843
Reaction Time 3 & 4	19	2.145*	0.480

* significant at $p<0.05$

Paired samples *t*-tests (Table 9) indicated that the response latencies were significantly longer for the Reaction Time Test 1 than those for Reaction Time Tests 2, 3, and 4 ($p<0.05$), and the effect sizes (Cohen's *d*) were large. Response latencies for Reaction Time Test 2 were significantly longer than those for Reaction Time Tests 3 and 4 ($p<0.05$), and the effect sizes (Cohen's *d*) were medium and large respectively. Response latencies for Reaction Time Test 3 were significantly longer than those for Reaction Time Test 4 ($p<0.05$), and the effect size (Cohen's *d*) was small.

Question Type (Self, Other, and Photo) ANOVAs and *t*-test

The group mean response latencies and standard errors for each data set for each of the question types (Self, Other, and Photo) from the perspective-taking phases are shown in Figure 4. As expected with the various treatments of the outliers the group mean response latencies within each type of question decrease across the three data sets. The group mean response latencies for the Self and Other questions are similar for the Original and Max 2.5 data sets, and are slightly longer than those for the Photo questions. The standard errors overlap between the three types of questions for these two data sets. Whereas the group mean response latencies for the McH data set are fairly similar across all types of question, and the standard errors overlap across question type (Self, Other, and Photo) but not across the Original and Max 2.5 data sets.

Repeated measures ANOVAs and *t*-tests were conducted on the response latencies of the different types questions (Self, Other, and Photo) for all three data sets. The ANOVAs, means and standard deviations, and *t*-test results for each data set are shown in Table 10 and 11 respectively. The main effect of question type (Self, Other, Photo) was significant for the Original and Max 2.5 data sets ($p<0.05$) and the effect sizes (partial $ή^2$) were medium and large respectively (Table 10). However, the main effect of question type was not significant for the McH data set and the effect size (partial $ή^2$) was medium (Table 10).

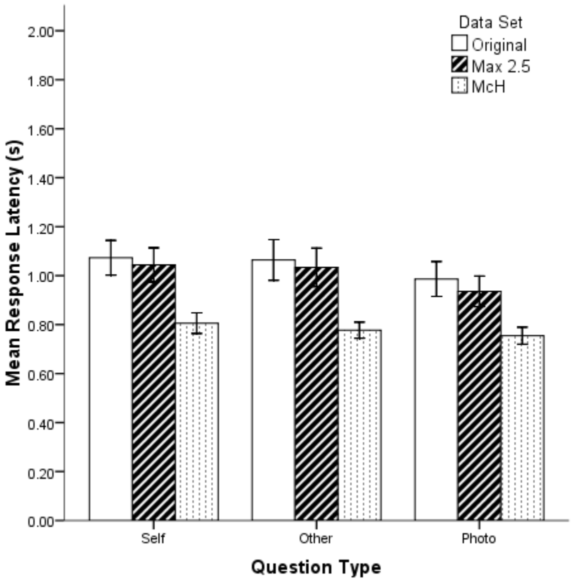

Figure 4: Group Means and Standard Errors of the Response Latencies for the Question Type (Self, Other, and Photo) for all the Trials from the Perspective-taking Phases for all the Data Sets (Original, Max 2.5, and McH)

Table 10

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Type (Self, Other, and Photo) for all the Trials from the Perspective-taking Phases ANOVAs Results for the Original, Max 2.5, and McH Data Sets

Data Set	*df*	F	Partial ή²	Self	Other	Photo
Original	2,38	3.324*	0.149	1.07 (0.32)	1.06 (0.37)	0.99 (0.32)
Max 2.5	2,38	11.152*	0.370	0.94 (0.24)	0.93 (0.25)	0.86 (0.22)
McH	2,38	2.198	0.109	0.81 (0.18)	0.78 (0.14)	0.76 (0.16)

* significant at $p<0.05$

Paired samples *t*-tests (Table 11) indicated that the response latencies for the Self questions were not significantly longer than those for the Other questions (p>0.05) for all the data sets and the effect sizes (Cohen's *d*) were all small. Response latencies for the Self questions were significantly longer than those for the Photo questions (p<0.05) for all the data sets and the effects sizes (Cohen's *d*) were medium for the Original and McH data sets, and large for the Max 2.5 data set. Response latencies for the Other questions were not significantly longer than those for the Photo questions for the Original data and McH data (p>0.05) and the effects sizes (Cohen's *d*) were both small, but the response latencies were significantly longer for the Max 2.5 data set (p<0.05) and the effect size (Cohen's *d*) was medium.

Table 11

The df, the t value, and Effect Size (Cohen's d) for the Question Type (Self, Other, and Photo) of the Response Latencies for all the Trials from the Perspective-taking Phases t-test Results for the Original, Max 2.5, and McH Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Original	19	0.319	0.009
	Max 2.5	19	0.569	0.127
	McH	19	1.487	0.341
Self & Photo	Original	19	2.389*	0.534
	Max 2.5	19	4.347*	0.972
	McH	19	2.208*	0.507
Other & Photo	Original	19	1.737	0.388
	Max 2.5	19	3.356*	0.750
	McH	19	0.619	0.142

* significant at $p<0.05$

Question Type (Self, Other, and Photo) ANOVAs and *t*-tests for each of the Perspective-taking Phases (Pre-test, Training, and Post-test)

The group mean response latencies and standard errors of the response latencies for the three data sets for the question type (Self, Other, and Photo) for the Pre-test, Training, and Post-test phases are shown in Figures 5, 6, and 7 respectively. The various treatments of the outliers in each data set result in decreases of the group mean response latencies across the three data sets (Original, Max 2.5, and McH) for each type of question (Self, Other, or Photo) for the Pre-test (Figure 5), Training (Figure 6), and Post-test (Figure 7) phases.

Figure 5 shows that for the Pre-test phase the group mean response latencies and the standard errors within each set are very similar across the three types of question (Self, Other, and Photo).

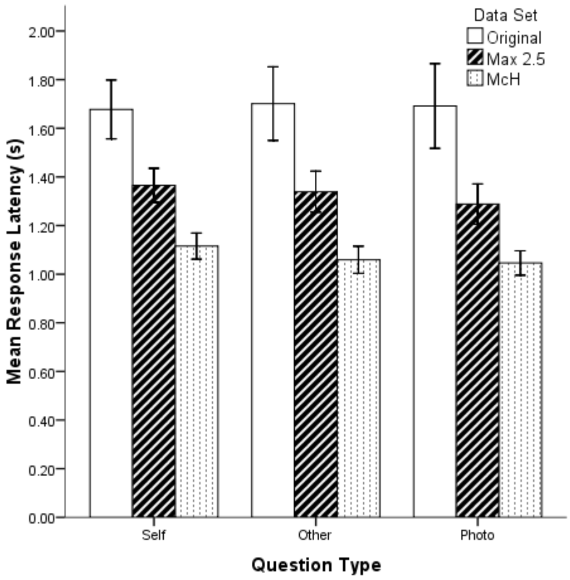

Figure 5: Group Means and Standard Errors of the Response Latencies for the Question Type (Self, Other, or Photo) for the Trials from the Pre-test Phase for all the Data Sets (Original, Max 2.5, and McH)

Figure 6 shows that for the Training phase the group mean response latencies for the Original and Max 2.5 data sets are similar within each data set for the Self and Other questions, and are longer than those for the Photo questions. The group mean response latencies for the McH data set show that the Self questions are slightly longer than those of the Photo questions, and the group mean response latencies of the Photo questions are slightly longer than those for the Other questions. Figure 6 also shows that the standard errors overlap across each type of question (Self, Other, and Photo) within each data set.

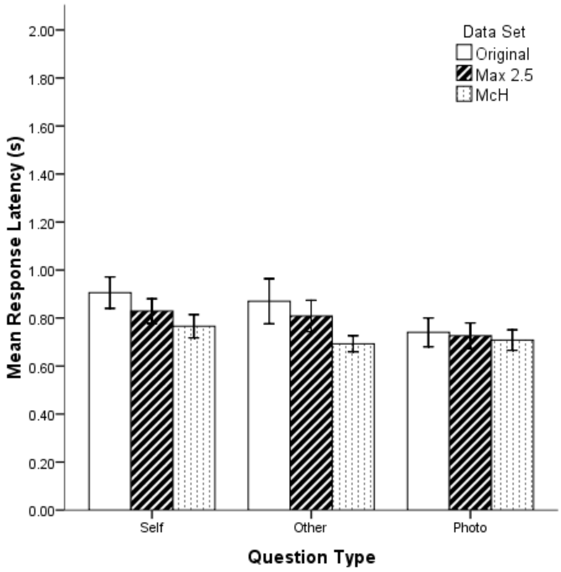

Figure 6: Group Means and Standard Errors of the Response Latencies for the Question Type (Self, Other, or Photo) for the Trials from the Training Phase for all the Data Sets (Original, Max 2.5, and McH)

Figure 7 shows that for the Post-test phase the group mean response latencies within each data set are similar for the Self and Other questions, and slightly longer than those for the Photo questions. The standard errors for each type of question (Self, Other, and Photo) overlap within each of the three data sets.

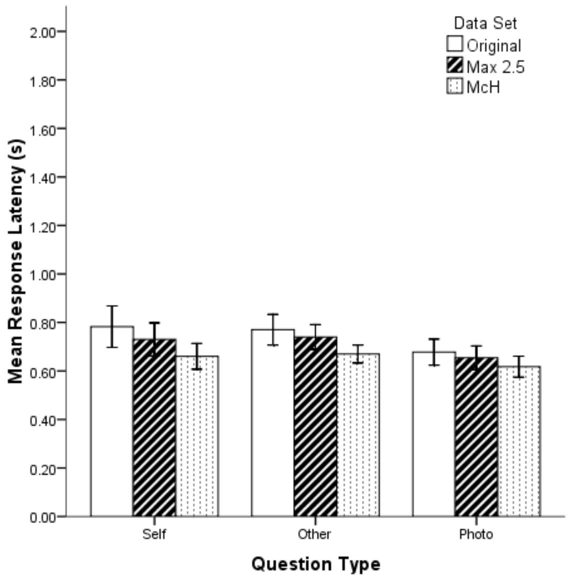

Figure 7: Group Means and Standard Errors of the Response Latencies for the Question Type (Self, Other, or Photo) for the Trials from the Post-test Phase for all the Data Sets (Original, Max 2.5, and McH)

Repeated measures ANOVAs were conducted on the response latencies across question type (Self, Other, and Photo) for each of the Pre-test, Training, and Post-test phases. The ANOVAs, means, and standard deviations for each data set for each of the perspective-taking phases (Pre-test, Training, and Post-test) are shown in Table 12. The main effect of question type (Self, Other, Photo) was not significant for the Pre-test phase for all data sets ($p>0.05$) and the effect sizes (partial $ή^2$) were small for all data sets. The main effect of question type (Self, Other, Photo) was significant for the Training and Post-test phases for the Original and Max 2.5 data sets ($p<0.05$) and the effect sizes (partial $ή^2$) were large for the Training phase and medium for the Post-test phase, but not significant for the McH data set ($p>0.05$) and the effect sizes (partial $ή^2$) were small.

Table 12

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from each of the Pre-test, Training, and Post-test Phases ANOVAs Results for the Original, Max 2.5 and McH Data Sets

Phase	*df*	F	Partial ή²	Self	Other	Photo
Original Data Set						
Pre-test	2,38	0.027	0.001	1.68 (0.54)	1.70 (0.68)	1.69 (0.78)
Training	2,38	5.374*	0.220	0.91 (0.29)	0.87 (0.42)	0.74 (0.28)
Post-test	2,38	3.411*	0.152	0.78 (0.38)	0.77 (0.28)	0.68 (0.24)
Max 2.5 Data Set						
Pre-test	2,38	1.480	0.072	1.37 (0.31)	1.34 (0.38)	1.29 (0.37)
Training	2,38	4.984*	0.208	0.83 (0.23)	0.81 (0.29)	0.73 (0.24)
Post-test	2,38	4.305*	0.185	0.73 (0.30)	0.74 (0.23)	0.65 (0.22)
McH Data Set						
Pre-test	2,34	0.761	0.043	1.11 (0.24)	1.06 (0.24)	1.06 (0.22)
Training	2,34	1.068	0.059	0.73 (0.14)	0.69 (0.14)	0.69 (0.15)
Post-test	2,36	1.181	0.062	0.66 (0.23)	0.67 (0.16)	0.62 (0.20)

* significant at p<0.05

Paired samples *t*-tests for the Pre-test phase (Table 13) indicated that there were no significant differences in response latencies between the Self and Other questions, the Self and Photo questions or the Other and Photo questions (p>0.05) for all the data sets and the effect sizes (Cohen's *d*) were all small.

Table 13

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from the Pre-test Phase t-test Results for the Original, Max 2.5, and McH Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Original	19	-0.327	-0.073
	Max 2.5	19	0.688	0.154
	McH	17	1.398	0.330
Self & Photo	Original	19	-0.132	-0.029
	Max 2.5	19	1.637	0.366
	McH	17	1.105	0.260
Other & Photo	Original	19	0.077	0.017
	Max 2.5	19	1.004	0.225
	McH	17	-0.029	-0.000

* significant at p<0.05

Paired samples *t*-tests for the Training phase (Table 14) indicated that the response latencies for the Self questions were not significantly longer than those for the Other questions for all data sets (p>0.05) and the effect sizes (Cohen's *d*) were small. The response latencies for the Self and Other questions were significantly longer than those for the Photo questions for the Original and Max 2.5 data sets (p<0.05) and the effect sizes (Cohen's *d*) were medium to large, but not significantly longer for the McH data set (p>0.05) and the effect sizes were medium and small.

Table 14

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from the Training Phase t-test Results for the Original, Max 2.5, and McH Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Original	19	0.580	0.130
	Max 2.5	19	0.570	0.127
	McH	17	1.474	0.347
Self & Photo	Original	19	3.620*	0.809
	Max 2.5	19	3.120*	0.698
	McH	18	1.508	0.346
Other & Photo	Original	19	2.542*	0.568
	Max 2.5	19	2.387*	0.534
	McH	17	0.066	0.022

* significant at $p<0.05$

Paired samples *t*-tests for the Post-test phase (Table 15) indicated that the response latencies for the Self and Other questions were not significantly different for all data sets (p>0.05) and the effect sizes (Cohen's *d*) were small. The response latencies for the Self and Other questions were significantly longer than for the Photo questions for the Original and Max 2.5 data sets (p<0.05) and the effect sizes (Cohen's *d*) were small to medium, but the response latencies were not significantly longer for the McH data set (p>0.05) and the effect sizes were small.

Table 15

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from the Post-test Phase t-test Results for the Original, Max 2.5, and McH Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Original	19	0.261	0.058
	Max 2.5	19	-0.273	-0.061
	McH	18	-0.265	-0.061
Self & Photo	Original	19	2.107*	0.471
	Max 2.5	19	2.373*	0.531
	McH	18	1.452	0.333
Other & Photo	Original	19	2.949*	0.659
	Max 2.5	19	3.244*	0.725
	McH	18	1.417	0.325

* significant at $p<0.05$

Question Type (Self, Other, and Photo) ANOVAs and *t*-test for the Combined Pre-test and Training Data

The data from the Pre-test and Training phases were combined and analysed to match the current study to that of McHugh et al. (2007a) on the number of trials presented. The group mean response latencies and standard errors of the response latencies for the three data sets for the question types (Self, Other, and Photo) for the Combined Pre-test and Training phases are shown in Figure 8. The various treatments of the outliers in each data set result in decreases of the group mean response latencies across the three data sets (Original, Max 2.5, and McH) for each type of question (Self, Other, or Photo). The group mean response latencies of the Self, Other, and Photo questions are similar within each data set and the standard errors overlap.

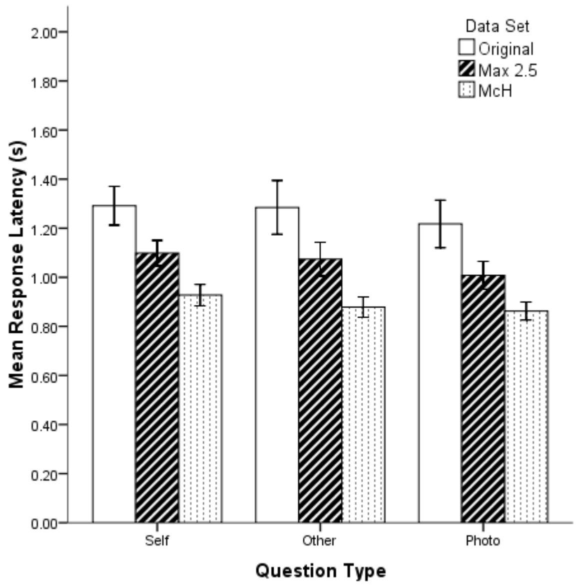

Figure 8: Group Means and Standard Errors of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from the Combined Pre-test and Training Phases for all Data Sets (Original, Max 2.5, and McH)

Repeated measures ANOVAs were conducted on the response latencies across question type (Self, Other, and Photo) for the combined Pre-test and Training phases. The ANOVAs, means, and standard deviations for each data set for the combined Pre-test and Training phases are shown in Table 16. The main effect of question type (Self, Other, Photo) was not significant for the Original and McH data sets ($p>0.05$) and the effect sizes (partial $ή^2$) were small and medium respectively, but was significant for the Max 2.5 data set ($p<0.05$) and the effect size (partial $ή^2$) was large.

Table 16

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Type (Self, Other, or Photo) for all Trials from the Combined Pre-test and Training Phases, ANOVAs Results for the Original, Max 2.5, and McH Data Sets

Data Set	*df*	F	Partial ή²	Self	Other	Photo
Original	2,38	1.117	0.056	1.29 (0.35)	1.29 (0.49)	1.22 (0.43)
Max 2.5	2,38	6.434*	0.253	1.10 (0.23)	1.07 (0.31)	1.01 (0.25)
McH	2,38	2.875	0.138	0.93 (0.19)	0.88 (0.18)	0.87 (0.16)

* significant at $p<0.05$

Paired samples *t*-tests (Table 17) indicated that the response latencies for the Self questions were not significantly longer than those for the Other questions for the Original and Max 2.5 data sets (p>0.05) and the effect sizes (Cohen's *d*) were small, but were significantly longer for the McH data set (p<0.05) and the effect size (Cohen's *d*) was medium. Response latencies for the Self questions were not significantly longer than those for the Photo questions for the Original data set (p>0.05) and the effect size (Cohen's *d*) was small, but were significantly longer for the Max 2.5 and McH data sets (p<0.05) and the effect sizes (Cohen's *d*) were large and medium respectively. Response latencies for the Other questions were not significantly longer than those for the Photo questions for the Original and McH data sets (p>0.05) and the effect sizes (Cohen's *d*) were small, but were significantly longer for the Max 2.5 data set (p<0.05) and the effect size (Cohen's *d*) was medium.

Table 17

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Question Type (Self, Other, or Photo) for all Trials from the Combined Pre-test and Training Phases, t-test Results for the Original, Max 2.5, and McH Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Original	19	0.114	0.026
	Max 2.5	19	0.920	0.206
	McH	18	2.432*	0.558
Self & Photo	Original	19	1.570	0.351
	Max 2.5	19	3.827*	0.856
	McH	18	2.287*	0.525
Other & Photo	Original	19	0.986	0.220
	Max 2.5	19	2.304*	0.515
	McH	18	0.158	0.036

* significant at $p<0.05$

Belief Question Types (True or False) *t*-tests

The group mean response latencies and standard errors for the three data sets (Original, Max 2.5, and McH) for the question belief (true or false) for the perspective-taking phases are shown in Figure 9. The various treatments of the outliers for each data set showed that the group mean response latencies for the type of belief (true or false) decrease across the three data sets. The group mean response latencies within each data set are similar for the type of belief (true or false) and the standard errors overlap.

Paired samples *t*-tests were conducted on the response latencies of the question belief for the perspective-taking phases. The means, standard deviations, and *t*-tests for each data set are also shown in Table 18. Response latencies for the false-belief questions were not significantly longer than those of the true-belief questions for the Original and McH data sets (p>0.05) and the effect sizes (Cohen's *d*) were small, but were significantly longer for the Max 2.5 data set (p<0.05) and the effect size (Cohen's *d*) was medium.

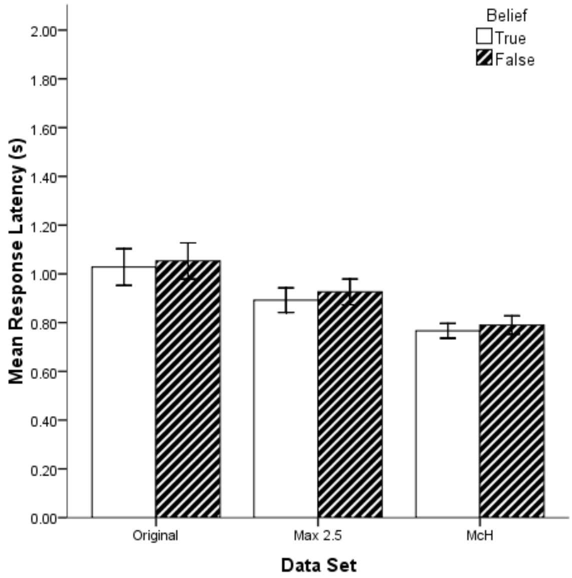

Figure 9: Group Means and Standard Errors of the Response Latencies for the Question Belief (True or False) for all the Trials from the Perspective-taking Phases for all the Data Sets (Original, Max 2.5, and McH)

Table 18

The df, the t value, the Effect Size (Cohen's d), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Belief (True and False) for all the Trials from the Perspective-taking Phases t-tests for the Original, Max 2.5, and McH Data Sets

Data Set	*df*	*t*	Cohen's *d*	True Belief	False Belief
Original	19	-0.663	-0.148	1.03 (0.34)	1.05 (0.33)
Max 2.5	19	-2.243*	-0.502	0.89 (0.22)	0.93 (0.24)
McH	19	-1.954	-0.437	0.77 (0.14)	0.79 (0.17)

* significant at $p<0.05$

The group mean response latencies and standard errors for the three data sets (Original, Max 2.5, and McH) for the question belief (true or false) for the Pre-test, Training, and Post-test phases are shown in Figure 10, 11, and 12 respectively. The various treatments of the outliers in each data set resulted in decreases of the group mean response latencies across the three data sets (Original, Max 2.5, and McH) for the belief of the question (true or false) for the Pre-test phase (Figure 9), but not for the Training (Figure 10), and Post-test (Figure 11) phases. Figures 10 and 12 show that for the Pre-test and Post-test phases respectively the group mean response latencies are similar for the type of belief (true or false) within each data set and the standard errors overlap. Figure 11 shows that for the Training phase the group mean response latencies are longer for the false-belief questions than those for the true-belief questions for each data set and the standard errors overlap.

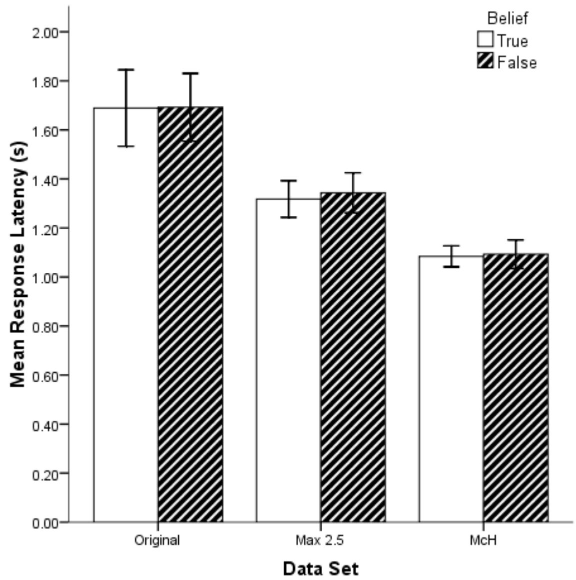

Figure 10: Group Means and Standard Errors of the Response Latencies for the Question Belief (True or False) for the Trials from the Pre-test Phase for all the Data Sets (Original, Max 2.5, and McH)

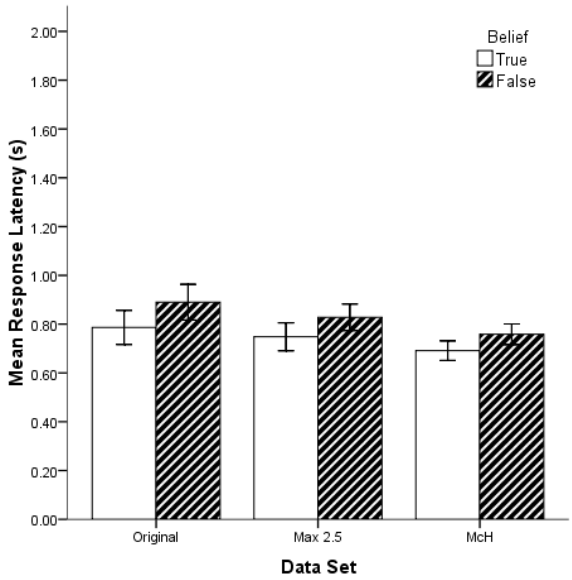

Figure 11: Group Means and Standard Errors of the Response Latencies for the Question Belief (True or False) for the Trials from the Training Phase for all the Data Sets (Original, Max 2.5, and McH)

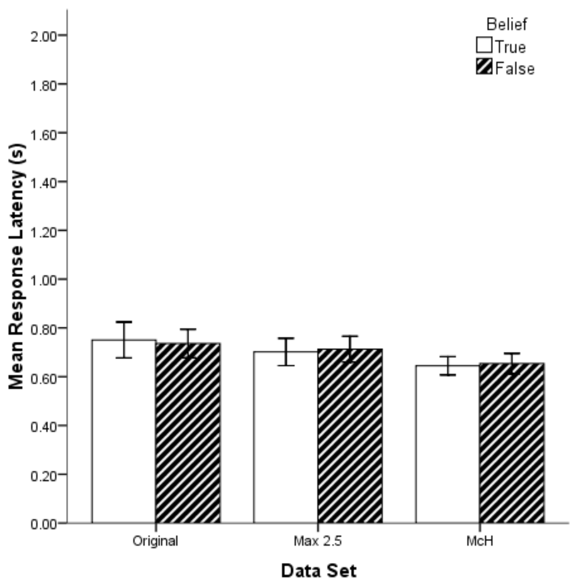

Figure 12: Group Means and Standard Errors of the Response Latencies for the Question Belief (True or False) for the Post-test Phase for all the Data Sets (Original, Max 2.5, and McH)

Paired samples *t*-tests were conducted on the response latencies for the true- and false-belief questions for each of the perspective-taking phases (Pre-test, Training, and Post-test). The means, standard deviations, and *t*-tests for each data set for the Pre-test, Training, and Post-test are shown in Table 19. Response latencies for the false-belief questions were not significantly longer than those for the true-belief questions for the Pre-test and Post-test phases for all the data sets (p>0.05) and the effect sizes (Cohen's *d*) were small, but were significantly longer for the Training phase for all the data sets (p<0.05) and the effect sizes (Cohen's *d*) were all medium.

Table 19

The df, the t value, the Effect Size (Cohen's d), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Belief (True and False) for the Trials from the Pre-test, Training, and Post-test Phases ANOVAs Results for the Original, Max 2.5, and McH Data Sets

Phase	*df*	*t*	Cohen's *d*	True Belief	False Belief
Original Data Set					
Pre-test	19	-0.029	0.006	1.69 (0.70)	1.69 (0.62)
Training	19	-2.320*	-0.519	0.79 (0.31)	0.89 (0.33)
Post-test	19	0.396	0.088	0.75 (0.33)	0.74 (0.26)
Max 2.5 Data Set					
Pre-test	19	-0.570	-0.128	1.32 (0.33)	1.34 (0.36)
Training	19	-2.565*	-0.574	0.75 (0.26)	0.83 (0.24)
Post-test	19	-0.459	-0.103	0.70 (0.25)	0.71 (0.24)
McH Data Set					
Pre-test	19	-0.305	-0.068	1.08 (0.19)	1.09 (0.26)
Training	19	-3.505*	-0.784	0.69 (0.18)	0.76 (0.19)
Post-test	19	-0.503	-0.112	0.64 (0.17)	0.65 (0.19)

* significant at $p<0.05$

Sets

The group mean response latencies and standard errors for the three data sets (Original, Max 2.5, and McH) for each of the different sets of questions from the perspective-taking phases are shown in Figure 13 and the means and standard deviations are shown in Table 20. Set 1 was the Reaction Time Test so was not included in the set analysis. Figure 13 shows that the group mean response latencies for Sets 2 and 3 from the Pre-test were clearly longer than those for the sets from the Training and Post-test phases for the Original and Max 2.5 data sets, whereas the group means for the McH data set, while still longer, were not as long as those from the Original and Max 2.5 data sets. The standard error bars within each Set showed that the Original and Max 2.5 data sets had similar variations in response latencies, indicating that the outliers adjusted from the Max 2.5 data set had little influence on the means and variations when compared to the Original data. As expected, the McH data set had smaller standard errors due to the large number of response latencies above 2.5s and errors that were removed. The standard error bars also showed that there was more variation in the response latencies for Sets 2 and 3 for the Original and Max 2.5 data sets than for all Sets shown in the Training, and Post-test. The group mean response latencies for Set 4 were consistently longer than those for the other Sets from the Training and Post-test phases. In general, the group mean response latencies for the Sets were similar for the Training and Post-test phases. Thus, as a group the participants responded faster and more consistently in the Training and Post-test phases compared to the Pre-test phase.

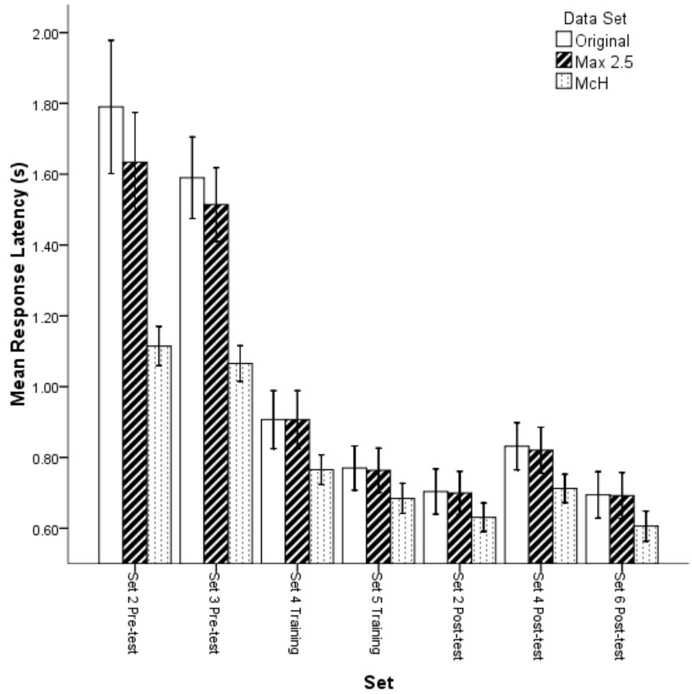

Figure 13: Group Means and Standard Errors of the Response Latencies for the Different Sets of Questions from all the Perspective-taking Phases for Each Data Set (Original, Max 2.5, and McH)

Table 20

Means and Standard Deviations (in brackets) of the Response Latencies for Each Set from the Perspective-taking Phases (Pre-test, Training, and Post-test) for the Original, Max 2.5, and McH Data Sets

Data Set	Pre-test		Training		Post-test		
	Set 2	Set 3	Set 4	Set 5	Set 2	Set 4	Set 6
Original	1.79(0.84)	1.59(0.52)	0.91(0.37)	0.77(0.28)	0.70(0.29)	0.83(0.30)	0.69(0.30)
Max 2.5	1.36(0.38)	1.30(0.31)	0.85(0.27)	0.72(0.23)	0.68(0.24)	0.79(0.25)	0.65(0.24)
McH	1.11(0.25)	1.07(0.23)	0.77(0.19)	0.68(0.19)	0.63(0.18)	0.71(0.18)	0.61(0.19)

Repeated measures ANOVAs and *t*-tests were conducted on the response latencies across the sets. *T*-tests were conducted on the mean response latencies from Set 2 Pre-test and those from Set 2 Post-test to examine the effect of simply doing the task again on response latencies even though this set was not used in training. The mean response latencies from Set 4 Training and those from Set 4 Post-test were compared to examine to effects of training on response latencies. The mean response latencies from Sets 2 and 6 Post-test, and those from Sets 4 and 6 Post-test were compared to examine the effect of generalization on response latencies to a novel stimulus set.

The ANOVAs and the *t*-test results for each data set are shown in Tables 21 and 22 respectively. The main effect of set was significant for all the data sets ($p<0.05$) and the effect sizes (partial $ή^2$) were large (Table 21).

Table 21

The df, the F value, the Effect Size (Partial $ή^2$) of the Response Latencies for the Sets from the Perspective-taking Phases ANOVAs Results for the Original, Max 2.5, and McH Data Sets

Data Set	*df*	F	Partial $ή^2$
Original	6,114	34.586*	0.645
Max 2.5	6,114	58.107*	0.754
McH	6,114	46.222*	0.709

* significant at $p<0.05$

Paired samples *t*-tests (Table 22) indicated that the response latencies for Set 2 Pre-test were significantly longer than those for Set 2 Post-test for all the data sets ($p<0.05$) and the effect sizes (Cohen's *d*) were large. Response latencies for Set 4 Training were not significantly longer than those for Set 4 Post-test for Original and Max 2.5 data sets ($p>0.05$) and the effect sizes (Cohen's *d*) were small, but were significantly longer for the McH data set ($p>0.05$) and the effect size (Cohen's *d*) was medium. Response latencies for Set 2 Pre-test and Set 3 Pre-test were significantly longer than those for the novel Set 6 Post-test for all the data sets ($p<0.05$) and the effect sizes (Cohen's *d*) were large.

Table 22

The df, the F value, the Effect Size (Partial ή²) of the Response Latencies for the Sets from the Perspective-taking Phases t-test Results for the Original, Max 2.5, and McH Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Set 2 pre-test &	Original	19	6.241*	1.396
Set 2 post-test	Max 2.5	19	8.858*	1.981
	McH	19	8.584*	1.919
Set 4 training &	Original	19	1.154	0.258
Set 4 post-test	Max 2.5	19	1.433	0.321
	McH	19	2.269*	0.507
Set 2 pre-test &	Original	19	6.045*	1.352
Set 6 post-test	Max 2.5	19	8.226*	1.839
	McH	19	7.929*	1.773
Set 3 pre-test &	Original	19	8.547*	1.911
Set 6 post-test	Max 2.5	19	9.141*	2.044
	Mch	19	8.100*	1.811

* significant at $p<0.05$

Adjusted Perspective Data

A new data set was generated by subtracting the mean response latency for each reaction time phase from the mean response latency for the perspective-taking phase immediately preceding it for each participant for the Original data set. These are termed the adjusted data and are shown in Table 23.

Repeated measures ANOVAs were conducted with these mean response latencies in order to investigate whether participants' response latencies got shorter throughout the experiment over and above the reaction time changes for the Original data. The ANOVA, means and standard deviations, and *t*-test results for the adjusted data set are shown in Table 24 and 25 respectively. The main effect of perspective phase remained significant (p<0.05) and the effect size (partial ή²) was medium (Table 24). Paired samples *t*-tests (Table 25) indicated that response latencies for the Pre-test phase were significantly longer than those for the Training and Post-test phases (p<0.05) and the effect sizes (Cohen's *d*) were large. However, response latencies for the Training phase were not significantly longer than for the Post-test phase (p>0.05) and the effect size (Cohen's *d*) was small.

Table 23

Calculated Means of Response Latencies of the New Adjusted Perspective Data for Each Participant for the Original Data Set

Participant	Pre-test Mean	Reaction Time Test 2 Mean	Adjusted Pre-test Mean	Training Mean	Reaction Time Test 3 Mean	Adjusted Training Mean	Post-test Mean	Reaction Time Test 4 Mean	Adjusted Post-test Mean
1	1.25	0.39	0.86	0.62	0.37	0.25	0.48	0.39	0.09
2	1.38	0.61	0.78	0.59	0.49	0.10	0.44	0.46	-0.02
3	1.23	0.48	0.76	0.63	0.37	0.26	0.71	0.43	0.28
4	0.95	0.37	0.58	0.72	0.41	0.31	0.51	0.40	0.10
5	1.55	0.42	1.13	1.04	0.39	0.65	0.95	0.35	0.60
6	1.81	0.59	1.22	0.57	0.45	0.11	0.53	0.39	0.14
7	0.93	0.39	0.54	0.62	0.37	0.25	0.50	0.33	0.17
8	0.89	0.37	0.52	0.64	0.37	0.27	0.90	0.36	0.55
9	1.55	0.53	1.02	0.60	0.56	0.04	0.60	0.44	0.16
10	2.25	0.44	1.82	0.87	0.34	0.53	0.89	0.36	0.53
11	1.99	0.87	1.12	0.82	0.61	0.21	0.70	0.57	0.13
12	1.43	0.61	0.82	0.81	0.71	0.10	0.73	0.56	0.16
13	1.33	0.82	0.51	0.93	0.43	0.49	0.72	0.37	0.35
14	2.45	0.48	1.97	1.84	0.51	1.32	1.74	0.55	1.19
15	2.65	0.44	2.21	0.69	0.39	0.29	0.71	0.39	0.32
16	2.04	0.65	1.39	0.77	0.56	0.21	0.60	0.48	0.12
17	2.17	0.62	1.55	1.31	0.43	0.88	0.66	0.38	0.28
18	0.94	0.63	0.31	0.88	0.65	0.23	0.65	0.49	0.16
19	1.96	0.45	1.51	0.76	0.52	0.25	0.95	0.50	0.45
20	3.05	0.64	2.40	1.09	0.39	0.71	0.89	0.48	0.41
Mean	**1.69**	**0.54**	**1.15**	**0.84**	**0.47**	**0.37**	**0.74**	**0.43**	**0.31**

Table 24

The df, the F value, the Effect Size (Partial Eta Squared), Mean and Standard Deviation (in brackets) of the Response Latencies for the Adjusted Perspective-taking Data from the Perspective-taking Phases ANOVA for the Original Data Set

df	F	Partial η^2	Pre-test	Training	Post-test
2,38	44.882*	0.703	1.15 (0.60)	0.37 (0.31)	0.31 (0.27)

* significant at $p<0.05$

Table 25

The df, the t value, and the Effect Size (Cohen's d) of the Response Latencies for the Perspective t-Test Results for the Adjusted Perspective Data for the Original Data Set

	df	*t*	Cohen's *d*
Pre-test and Training	19	6.746*	1.508
Pre-test and Post-test	19	7.030*	1.572
Training and Post-test	19	1.547	0.346

* significant at $p<0.05$

Discussion

Response latencies will be longer as a function of relational complexity

Self, Other, Photo question types

One aim of the current study was to replicate McHugh et al.'s (2007a) findings that response latencies for the different types of questions (Self, Other, and Photo) would be longer the greater the relational complexity of the question. Both McHugh et al. (2007a) and the Original and Max 2.5 data sets from the current study found that there was a significant main effect of the question type on response latencies taken from all the trials from the perspective-taking phases of this study. However, the main effect of the question type was not significant for the McH data set from the current study, which dealt with the errors and outliers in the same way as McHugh et al. (2007a).

Analysis of the three data sets (Original, Max 2.5 and McH) are discussed here and show that the results differ depending on the method of dealing with the long response latencies or incorrect trials. Analyses of the Max 10 and Max 5 data sets (Appendix E-J) showed that most results of the analyses were similar to those of the Max 2.5 data set.

The response latencies from all the perspective-taking trials for the Self questions were very similar to those of the Other questions, and both were longer than those of the Photo questions for all the data sets. This result is only partially consistent with the prediction of Relational Frame Theory that response latencies will be longer for the types of questions that involve more relational complexity, and will be shorter for the types of questions that involve less relational complexity. As predicted, the response latencies from Photo questions, involving the least relational complexity, were shorter than those from both the Self and Other questions. However, Relational Frame Theory suggests that the response latencies of the Other questions should have been longer than those from the Self questions as the Self questions involve less relational complexity. This was not the case, they were in fact very similar, suggesting that they were at a similar level of difficulty. McHugh et al.'s results were consistent with the prediction of Relational Frame Theory as they found that the response latencies from the Other questions were significantly longer than those of the Self questions which, in turn, were also longer than the Photo questions but not significantly so.

The Photo questions do not involve the deictic frames that are used in perspective-taking therefore they should be easier and have shorter response latencies. It does not matter whether it was You or I who took the photo, if it was taken Then or Now, or if it was taken Here or There. In all cases the photo will remain unchanged because the participant simply has to report on its content. On the other hand, both the Self and Other questions require the use of the deictic relations Then-Now and Here-There. The difference between the Self and Other questions is whether the perspective was from the participant (Self) or the other person (Other). The Other questions required the participants to respond in accordance with an if-then relation; they must derive "If I were you, then I would…" (McHugh et al., 2007a) in order to respond to the trial correctly. The Self questions do not require this relation to be made therefore they involve less relational complexity than the Other questions.

Given this, the McHugh et al. (2007a) finding that the Other questions, which involved more relational complexity than the Self questions, had significantly longer mean response latencies than the Self questions, is as would be predicted. However, their finding that the mean response latencies of the Self questions were not significantly different from those of the Photo questions is surprising.

The above findings from the current study were the results of analyses that included the data from all the trials from all the perspective-taking phases combined. The current study required participants to respond to a total of 216 trials and of these, 168 were from the perspective-taking phases which involved a Pre-test, Training, and a Post-test. McHugh et al.'s (2007a) study required participants to respond to one block of 96 trials, all perspective-taking tasks. McHugh et al. (2007a) did not include training in their research, so a better comparison maybe with only the data from the Pre-test of the current study where the procedure was the same as theirs. The effect of question type (Self, Other, Photo) was not significant for the current study's Pre-test data. The Pre-test from the current study contained 48 trials and the study of McHugh et al. (2007a) contained 96 trials. Given there was no feedback on correct trials in either the Pre-test of the current study and the entire study of McHugh et al. (2007a) and given the high degree of accuracy on the task from the start, it is unlikely that the participants would have 'learned' the task if there had been a further 48 trials without feedback, they may however have got 'faster' by becoming more adept at the mechanics of the task. So it maybe that decreases in response latencies from getting faster at the mechanics of the task may mask any differential effects of the response latencies for the question types.

The results of the current study were also analysed by combining the data from the Pre-test and Training phases to allow a comparison to be made between this study and McHugh et al. (2007a) based on the number of trials. The 96 trials from the Pre-test and Training phases may have allowed the mechanics of the task to become sorted which would allow the differences in the complexity of the tasks to appear in the response latencies. The main effect of the type of question (Self, Other, Photo) was not significant for the Original and McH data sets, but was significant for the Max 2.5 data set. Analysis of the response latencies from the Original data set indicated that at the completion of 96 trials, the response latencies of the three question types (Self, Other, and Photo) were similar and that no differences had emerged in the difficulty of the questions (Table 17). This result was consistent with the analysis of the Pre-test phase from the current study. Analysis of the response latencies from the Max 2.5 data set were consistent with the findings of the analyses of all the trials from all the perspective-taking phases combined, and the response latencies indicated that the Self and Other questions were similar in their difficulty, and both more difficult than the Photo questions (Table 17). Analysis of the response latencies from the McH data set indicated that the Self questions were more difficult than the Other and Photo questions, but that the Other and Photo questions were similar (Table 17). The above findings highlight that the method used to deal with the outliers leads to different conclusions. For this reason it is not possible to draw conclusions after 96 trials about the effect (if there is one at this point) of learning the mechanics of completing the task, or if there are any differential effects resulting from the relational complexity of the perspective-taking tasks.

True- or False- Belief

The relational complexity of the question presented in each trial may be influenced by whether the question related to a true- or false-belief. False-belief tasks involve logic NOT and have been thought to be more difficult than true-belief tasks that do not involve logic NOT. According to ToM model, true-beliefs develop at Level 4 whereas false-beliefs develop at the higher Level 5, suggesting that false-beliefs are more difficult because they develop after true-beliefs. To date, investigation into true- and false-belief under the rubric of Relational Frame Theory suggests that there may be an overlap in the relational skills used in understanding both true- and false-belief.

McHugh et al. (2006) did not find significant differences for accuracy in responses involving true- or false-belief, and McHugh et al. (2007a) also did not find there to be a significant difference in the response latencies from the true- and false-belief questions. The current study found that the effect of belief, when the data of all perspective taking phases were combined, was not significant for the Original and McH data sets, but was significant for the Max 2.5 data set. The findings for the Original and McH data sets are consistent with previous Relational Frame Theory accounts of true- and false-beliefs and provide support for the idea that the relational skills involved in understanding true- and false-beliefs overlap. Analysis of the three perspective-taking phases separately indicated that the response latencies from the belief of the question were only significant for the Training phase for all data sets, suggesting that the true-belief questions were easier than the false-belief questions. It is surprising that the differential response latencies from the two types of belief were not maintained and were not evident for the Post-test phase. If training was influential in differentiating the difficulty of the true- and false-belief questions then it would be expected that this learning would generalize to other stimulus variations. However, it is possible that training did not have a differential effect and that the differences found in the mean response latencies of the true- and false-belief questions may have been the result of some other factor. The effect of training with feedback on learning of true- and false-beliefs requires further investigation.

Generalization to novel stimuli will occur

A second aim of the current study was to investigate whether the learning effects (measured by decreases in response latencies) expected as a result of training would generalize to novel perspective-taking stimuli. Generalization is important when learning perspective-taking skills because it would be impossible to teach an individual every instance in which perspective-taking occurs. It is more desirable that an individual understands the relational properties involved in perspective-taking so that this can be applied to novel situations. The Relational Frame Theory account of perspective-taking describes perspective-taking as generalized operant behaviour, therefore it is expected that generalization will occur.

In the current study a novel stimulus set (Set 6) was included in order to test whether generalization occurred as a result of training. The Post-test phase of the experiment contained three sets of questions, one from the Pre-test phase (Set 2), one from the Training phase (Set 4), and the novel set (Set 6). For generalization to occur, training must first improve performance. In the current study evidence for an effect of training would be the response latencies for Set 4 being shorter for those from the Post-test phase, than for those from the Training phase. To show that the training generalised to novel sets the improved performance must also occur for stimuli that were not included in the training. Here this would be the response latencies from Set 6 being similar to those of Sets 2 and 4 in the Post-test.

The mean response latencies from Set 4 in the Post-test were shorter than those from Set 4 in the Training, and the response latencies from the novel Set 6 (in the Post-test) did not consistently differ from those for Sets 2 and 4 also in the Post-test for all participants (Table 5). Thus, these results suggest that generalization to the novel Set 6 occurred for all participants.

Multiple exemplars of the task (practice) may also improve performance on that task. Evidence for this occurring in the current study would be seen if the response latencies for Set 2 for the Post-test were shorter than those of Set 2 in the Pre-test. This was the case (Table 5). Further evidence would be provided by the response latencies for the sets presented during Training (Sets 4 and 5) being shorter than those in the Pre-test (Sets 2 and 3). This was also the case for all participants (Table 5). These results suggest that multiple exemplars may also play a role in improving performance on perspective-taking tasks.

A limitation of the current study is that multiple exemplars and training (multiple exemplars and feedback) were not investigated independently. It may be that simple practice of the task (multiple exemplars) could increase performance on the task and generalize to novel stimuli. Or it may be that the combination of feedback and practice are required to improve performance and produce generalization. If this study had used a control group who were given the same task but did not receive feedback during the Training phase, the effect of feedback could have been investigated as a role in learning about perspective-taking. Future research is needed to investigate this.

Extended training on perspective-taking tasks will decrease response latencies

The third aim of the current study was to investigate the hypothesis that extended training on the perspective-taking tasks would decrease response latencies. The mean response latencies for the group decreased across the three perspective-taking phases (Pre-test, Training, and Post-test) (Figure 2) and this effect was significant showing that the extended training improved the participants' performances.

The analysis of the response latencies for the different question types across the trials from all three perspective-taking phases shows that extended training on the perspective-taking tasks decreased response latencies differentially. At the beginning of the experiment, during the Pre-test phase, the three types of question (Self, Other, Photo) were of similar difficulty (i.e., the response latencies were not significantly different). However, for the Training and Post-test phases, the Photo questions were easier than both the Self and Other type questions for the participants (i.e., response latencies for the Photo questions were significantly different (and shorter) than those for the Other and Self questions (which remained similar) for the Original and Max 2.5 data sets. Note that for the McH data set, the mean response latencies from the Photo questions were shorter than those from the Self and Other questions, but they were not significantly so. The results suggest that at the beginning of the experiment, the different types of questions were of similar difficulty and that training on the perspective-taking tasks decreased response latencies more on the easier Photo questions, than on those of the Self and Other questions that involve more relational complexity.

There are two possible explanations for the finding that response latencies decreased over the perspective-taking phases. First, the practice the participants received by completing multiple exemplars and receiving feedback during the Training phase of the experiment could have improved their performance on the task as a result of learning about being able to see things from differing perspectives. Or, second, it may be that the participants' performances improved in part, as a result of getting better at responding to the task; that is, simply doing the mechanical actions required for responding by pressing the Z or M key faster.

Response latencies on the Reaction Time Tests will decrease over time but will not account for all the changes found on the perspective-taking task

The possibility that decreases in response latencies may have been due in part to the participants simply getting faster at responding to the task was tested by the fourth and final aim of the current study that was to investigate how the response latencies on a simple Reaction Time Test would change with practice and to investigate how any changes in the response latencies on the perspective-taking tasks compared with these.

The response latencies for both the perspective-taking phases and the reaction time phases decreased significantly over the duration of the experiment. The reaction time data suggest that the participants got faster at the task by simply doing the mechanical responses required to complete the task. Thus it is possible that some of the decrease in perspective-taking response latencies was the result of simply getting faster at the mechanics of doing the task.

One possible way to investigate whether perspective-taking response latencies changed independently of reaction time response latencies was to eliminate the reaction time component. This was accomplished by creating an adjusted data set for the perspective-taking phases. The new adjusted data set was generated by subtracting the mean response latency for each reaction time phase from the mean response latency of the perspective-taking phase immediately preceding it for each participant (Table 23). Reaction Time Test 1 was not used in the generation of the adjusted data as the participants' were not given any practice trials prior to this and so it could be argued that the response latencies generated from Reaction Time Test 1 were not a true reflection of reaction time as the participants were becoming familiar with the demands of the task.

There are at least two possible outcomes from removing the simple reaction time from the response latencies of the perspective-taking phases. One is that the remaining response latencies across the perspective-taking tasks would still decrease, suggesting that perspective-taking learning occurred over and above any decreases in simple reaction time. Another is that the remaining response latencies would be similar across the perspective-taking phases, suggesting that decreases in response latencies were due to the participants getting faster at carrying out the mechanics of the task, and not due to learning to perspective-take.

When the reaction time response latencies were removed the new mean response latencies still decreased significantly across the three perspective-taking phases (Table 24). The adjusted Pre-test latencies were significantly longer than those from the Training phase, indicating that participants' performance on the task improved over and above simply getting faster at responding to the task and that the training helped with perspective-taking. The adjusted response latencies from the Training phase were longer than those for the Post-test but the difference was no longer significant. This suggests that performance increases on the perspective-taking tasks after the end of training were mainly a result of participants getting faster in doing the mechanics of the task.

The greatest learning (decreases in response latencies) appeared to occur at the beginning of the experiment and the benefits of extended training appeared to reduce over time. That is, the magnitudes of the decreases in response latencies decreased with each successive phase (Table 23). There are at least two possible explanations for the finding of reduced change in response latencies over the phases.

First, the positive feedback (the word "Correct") that the participants received from responding correctly to the trials in the Training phase may have decreased their response latencies to the trials during the Training phase more than practice alone as experience during the Post-test (repeating the task without feedback). Feedback was not given during the Post-test phase and it may be that without reinforcement (feedback) the performance gains that potentially could have occurred were instead reduced as some form of feedback is more likely to decrease response latencies than no feedback. Examining the data from the last three Reaction Time Tests, shows that the decreases in response latencies between Reaction Time Test 2, 3, and 4 were less than those found with the perspective-taking tasks. This finding supports the idea that the combination of practice and feedback that occurred during the Training phase was more effective at decreasing response latencies than practice alone.

The second possible explanation is that the extended practice (repeating the task) decreased response latencies over time, regardless of feedback, but the benefits of more practice reduced over time. Examination of the mean response latencies of the Sets presented during the perspective-taking phases (Figure 13 & Table 20) show that the group mean response latencies for Sets 2 and 3 presented during the Pre-test were clearly longer than those for the Sets presented during the Training and Post-test phases and the response latencies of the latter two phases were similar (Figure 13). This indicates that repeating the task without feedback (as done so with Sets 2 & 3 in the Pre-test), may have reduced response latencies alone. The repetition of perspective-taking tasks in the Pre-test (from responding to Sets 2 & 3) may have taught the participants about the task. It is possible that when the participants were presented with new stimulus sets in the Training phase (Sets 4 & 5), the participants were able to generalize what they had learnt from the Pre-test phase to the same task in the Training phase. It is also possible that the multiple exemplars and feedback that occurred during the Training phase did not decrease the response latencies any further, which would explain why the response latencies for the Sets presented in the Training and Post-test phases are similar. The differences between the means of the response latencies on successive Reaction Time Tests became smaller for each new test (even though no feedback was given on this task) suggesting that the benefits of more practice reduced over time. Together these findings suggest that it is possible that learning of the perspective-taking task in the Pre-test phase (from repeating the task) may have generalized to the trials in both the Training and Post-test phases, showing that the more practice an individual has the less effect it has on decreasing response latencies.

Discussion on floors of reaction time and perspective-taking tasks

It would be expected that there is maximum speed at which participants can perform the task and so at some point the participants would not be able to respond any faster to the task, regardless of the amount of practice/training completed. A "floor" is the term used when there is no room for further improvements in the measure being used. The reaction time data show that the decrease between the mean response latencies for each Reaction Time Test suggest that the response latencies for the Reaction Time Tests were reaching a floor.

The decreases in mean response latencies between each perspective-taking phase for each type of question (Self, Other, and Photo) (Table 12) suggest that these response latencies for the perspective-taking tasks are also reaching a floor. It is unclear if or how the effect of floors will alter or not alter the differential response latencies found in the relational complexity of the questions of the perspective-taking tasks, but there are two possibilities. Either, all types of questions, regardless of relational complexity, will eventually have similar response latencies (when response latencies for each question type reach a floor), therefore eliminating any differential effects. Or the differential effects of the response latencies between the types of questions with different levels of relational complexity will remain. Future research could investigate the differential changes in response latencies of the different types of questions involving different levels of relational complexity over an extended number of trials.

Outliers and Data Sets

Response latency was used as a measure of learning in this study because it has been argued that response latencies are a more sensitive measure than response accuracy (as used in previous research on perspective-taking, e.g., Rehfeldt et al., 2007; McHugh et al., 2007b; McHugh et al., 2006) because changes in response latencies can still be observed when response accuracy has stabilized (Spencer & Chase, 1996). We would expect that in this study, where the task should be relatively easy for the participants, response latencies would be no longer than a few seconds. Any longer and it could be argued that participants were re-reading or thinking for an extended period of time about the trial rather than making an almost instant decision (reaction) to the question. The statements in each trial were presented word by word, with a new word appearing every 0.512s (consistent with McHugh et al. (2007a), and Sabbagh & Taylor (2000)) and response latency was taken from the appearance of the last word until a response by the participant was made. The steady and regular presentation of the words on the computer screen left plenty of time for reading and this should mean that participants should have read all the words prior to the start of the recording of response latency and so it should have eliminated reading the questioning from the response latency.

The participants' raw data (Appendix K1-K20) show that there were a small proportion of trials that had response latencies much longer than would be required to finish reading and responding to the trial. McHugh et al. (2007a) report that they did not include the data from any trials with errors or with response latencies above 2.5s, however, it is not known how many or how long the response latencies were for any outliers in their study.

The data of the current study were analysed using five different data sets (Original, Max 2.5, and McH are presented, and the Max 10 and Max 5 can be found in Appendix E-J) in order to investigate the effect of different methods of dealing with the outliers on the overall results and to compare the results of the current study with those of McHugh et al. (2007a). The analysis of the results using the three different data sets showed some differences between the sets. The Original and Max 2.5 data sets had only a small number of outliers adjusted across all trials and participants compared to the McH data set, so these outliers tended to lose their effect when the data was analysed as a group. There were more outliers in the Pre-test phase than in the Training, Post-test or any reaction time phases. This may indicate that the outliers with longer response latencies found during the Pre-test were part of the learning process. The McH data set removed any trials where an error was made or the response latency was longer than 2.5s. Dealing with the data in this way meant that some participants had large numbers of trials removed from analysis (e.g. Participant 3 had 52 trials out of a total of 216 removed). It could be argued that removal of data using this method may obscure the results as more often than not, the trials that were removed due to errors or longer response latencies were trials that could be considered to be the more difficult. The different results from the different data sets highlight that the way in which outliers and errors are dealt with can influence the outcomes of the research.

McHugh et al. (2007a) included 10 practice trials before the proper experiment began. In the current study the participants were not given any perspective-taking practice trials, but were given Reaction Time Test 1 prior to the Pre-test phase. Reaction Time Test 1 allowed the participants to become familiar with the requirements of the task but did not allow them to practice on the perspective-taking trials. With this small methodological change the overall findings were still fairly consistent with those of McHugh et al. (2007a). If the practice trials were influential in decreasing the response latencies before the experiment began, then we would expect the response latencies from the first 10 or so trials of the Pre-test in the current study to be longer than those later in the Pre-test and Figure 1 shows that this is the case for all the participants. Although the response latencies at the beginning of the Pre-test were longer than those later in the Pre-test, there were very few extreme outliers. It could also be argued that these longer response latencies were part of the learning process and should be included in analysis. Analysis of the Max 2.5 data set (which adjusted any response latencies above 2.5s) and Original data set (which left the response latencies as there were), showed little difference between the overall results of these two sets suggesting that the existence of the longer response latencies made little or no difference to the overall outcomes of the study, so the inclusion of perspective-taking practice trials was probably not required and would most likely not have altered the results.

Throwing out Participant 21 and keeping Participant 3

Response latency was used as a measure of performance because it was expected that this population of participants would have fully developed perspective-taking skills, and would respond at with a high level of accuracy. The accuracy of responses by the majority of participants was very high and consistent across the three perspective-taking phases (Table 1). This indicates that most participants understood the task and that accuracy was not compromised when participants got faster at responding to the task. There were two participants (Participant 21, and Participant 3) that did not respond with a high level of accuracy as expected.

Participant 21 showed very low accuracy in the perspective-taking phases only responding correctly to 52% of the perspective-taking trials (Table 1), indicating that they were responding at chance level. The trials in the Training phase, where feedback was given, improved this participants' accuracy slightly to 62.5%, but this was not maintained as the accuracy in the Post-test was similar to that in the Pre-test phase (Table 1). Their response latencies for the Reaction Time Tests were much shorter and more accurate than those for the perspective-taking phases, eliminating the idea that this participant was simply not attending to the task. For these reasons, the responses made by this participant suggest that they either did not understand the perspective-taking task, or they were unable to do it; therefore Participant 21's data was not used for analysis.

Participant 3 also responded with accuracy lower than what would have been expected from a participant who was likely to have normally developed perspective-taking skills. Further investigation of Participant 3's raw data (Appendix K3) revealed that the majority of incorrect responses were made on three types of questions. This suggests that Participant 3 had partial perspective-taking skills, as the majority of responses on all other types of questions were correct, and response accuracy on the reaction time phases was also high. The response latencies on the incorrect trials did not affect the overall response latencies for the perspective-taking phases, therefore Participant 3's data were included in the group analysis.

Limitation

A limitation of the current study is that it is unclear if the improvement in performance over the course of the experiment would have been maintained. If it had been possible, a follow up test to the current experiment could have investigated the maintenance of the skill over time. If the response latencies for the follow up were similar to those during the Post-test then it could be said that the perspective-taking skills were maintained. If the response latencies for the follow up were longer, an investigation into how many trials are needed before the response latencies return to the length of those in the Post-test could be carried out. A minimal number of trials would suggest that the skill was maintained but required a booster session to return the perspective-taking performance to that at the end of the current experiment. Whereas if a large number of trials were required, say similar to the number in the current experiment, this would suggest that the benefits of practicing the perspective-taking skills was not maintained over time.

What does my study add back into the literature?

The contribution of the current research to the literature is threefold. First, the current research provides some support for that of McHugh et al. (2007a) that response latencies of perspective-taking tasks involving greater relational complexity are longer than those for tasks involving less relational complexity, consistent with predictions of Relational Frame Theory. This suggests that the less relational complexity a task involves, the easier the task should be. Future research could investigate if training on easier perspective-taking tasks will facilitate increased performance on harder perspective-taking tasks involving more relational complexity.

Second, the current research has shown that generalization to novel stimuli occurs when being trained on that task using multiple exemplars of the task. This was evident from the response latencies for Set 6 presented in the Post-test phase of the experiment. Generalization to novel stimuli is consistent with the predictions of Relational Frame Theory which explains perspective-taking learning as generalized operant behaviour that occurs as a result of generalized patterns of arbitrarily applicable relational responding (McHugh et al., 2007b). The type of relational responding identified in perspective-taking is deictic frames.

Deictic relations are unlike other relations because they require an individual to respond to the relations rather than to formal or nonarbritrary properties (McHugh et al., 2007b). There are no physical or nonarbitrary properties on which to make rules about understanding perspective-taking. In normal development it is thought that children develop perspective-taking skills by responding to a history of questions such as "What am I doing now?" or "What were you doing then?" (McHugh et al., 2007a). Therefore it may be that through trial and error of multiple exemplars, an individual will be able to construct their own 'rules' for understanding perspective-taking. To date, the current study is the only research that has investigated generalization across perspective-taking tasks using different stimulus sets when participants are trained on the tasks using multiple exemplars and feedback. A property of relational frames is that generalized patterns of responding occur so that each new situation does not need to be explicitly taught. It may be that training on one type of perspective-taking task (e.g. deception tasks) could facilitate learning on other perspective-taking task (e.g. false-belief tasks). Future research could investigate if there is a relationship between training on one perspective-taking task and increased learning (generalization) to another perspective-taking task.

Third, the current research has shown that the use of multiple exemplars and training increased participants' performances on the perspective-taking task. This is consistent with results of O'Hora et al. (2002) who found that training decreased response latencies for normally developed participants responding to functional relations involving the frames more-than, less-than, and same-opposite. Together these results would suggest that whether the relational frames include arbitrary or non-arbitrary relations is irrelevant, and that training (using multiple exemplars and feedback) is an effective method to improve performance on tasks involving relational frames.

Training with the use of multiple exemplars may be an effective method for improving perspective-taking skills for individuals who have perspective-taking deficits such as those with autism. Further research is needed to investigate this area.

To summarise, the current research has shown that the use of multiple exemplars in extended training on a perspective-taking task improved performance on that task over and above decreases in reaction time for normally developed adults; and that the learning generalized to novel stimulus sets. The study also shows that response latencies decrease differentially for questions involving more relational complexity (Self and Other) than questions involving less relational complexity (Photo) and provides some support for McHugh et al.'s (2007a) findings that response latencies increase as a function of increased relational complexity.

References

Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D. & Jones, T. (2003). Theory of mind and relational complexity. *Child Development, 74,* 1476-1499.

Alcantara, P. R. (1994). Effects of videotape instructional package on purchasing skills of children with autism. *Exceptional Children, 61,* 40-55.

Charlop-Christy, M. H., & Daneshvar, S. (2003). Using video modeling to teach perspective taking to children with autism. *Journal of Positive Behavioral Interventions, 5,* 12-21.

Charlop-Christy, M. H., Le, L. & Freeman, K. (2001). A comparison of video modeling with in-vivo modeling for teaching children with autism. *Journal of Autism and Developmental Disorders, 30*, 537-552.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawerence Erlbaum Associates, Publishers.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Haldord, G. S., Wilson, W. H. & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803-865.

Hayes, S. C., Barnes-Holmes, D. & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

Horne, P.J. & Lowe, C. F. (1998). On the origins of naming and other symbolic behaviour. *Journal of the Experimental Analysis of Behavior, 65,* 185-241.

Howlin, P., Baron-Cohen, S. & Hadwin, J. (1999). *Teaching children with autism to mind-read: A practical guide.* Chichester, England: Wiley.

LeBlanc, L. A., Coates, A. M., Daneshvar, S., Chalop-Christy, M. H., Morris, C. & Lancaster, B. M. (2003). Using video modelling and reinforcement to teach perspective-taking skills to children with autism. *Journal of Applied Behavior Analysis, 36,* 253-257.

Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition, 40,* 203-218.

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

McHugh, L., Barnes-Holmes, Y., Barnes-Holmes, D. & Stewart, I. (2006). Understanding false belief as generalized operant behavior. *The Psychological Record, 56,* 341-364.

McHugh, L., Barnes-Holmes, Y. & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *The Psychological Record, 54,* 115-144.

McHugh, L., Barnes-Holmes, Y., Barnes-Holmes, D., Whelan, R. & Stewart, I. (2007a). Knowing me, knowing you: Deictic complexity in false-belief understanding. *The Psychological Record*, *57,* 533-542.

McHugh, L., Barnes-Holmes, Y., Barnes-Holmes, D., Stewart, I. & Dymond, S. (2007b). Deictic relational complexity and the development of deception. *The Psychological Record, 57,* 517-531.

O'Hora, D., Roche, B., Barnes-Holmes, D. & Smeets, P. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record, 52,* 51-75.

Rehfeldt, R. A., Dillen, J. E., Ziomek, M. M. & Kowalchuk, R. K. (2007). Assessing relational learning deficits in perspective-taking in children with high-functioning autism spectrum disorder. *The Psychology Record, 57,* 23-47.

Russell, J., Mauthner, N., Sharpe, S. & Tidswell, T. (1991). Strategic deception in a competitive game. *British Journal of Developmental Psychology, 9,* 331-349.

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of the theory-of-mind reasoning: An event-related potential study. *Psychological Science, 11,* 46-50.

Sidman, M. (1971). Reading and auditory-visual equivalences. Journal of Speech and Hearing Research, 14, 5-13

Sodian, B., & Firth, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry*

Spencer, T. J. & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65,* 643-659.

Steele, D. & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56,* 519-555.

Tierney, K. J. & Bracken, M. (1998). Stimulus equivalence and behaviour therapy. In O'Donohue, W. (Ed), *Learning and Behavior Therapy,* 392-402. Allyn and Bacon: Boston.

Vivanti, G., Nadig, A., Oxonoff, S. & Rogers, S. J. (2008). What do children with autism attend to during imitation tasks? *Journal of Experimental Child Psychology, 101,* 186-205.

Appendix A

Research Participants Wanted Notice

Research Participants Wanted

My name is Laura Baker and I am conducting research as part of my Master of Psychology (Behaviour Analysis). I am interested in the way people learn and react to simple perspective-taking tasks.

During this study you will be required to read a number of statements that will be presented to you on a computer screen and make a judgment whether you think the statement is true or false. There will be several phases in the experiment, one of which will involve some practice/training. The experiment will be completed in one session that will take approximately 2hrs. You will be offered several breaks throughout this time.

Students enrolled in PSYC102 or PSYC103 will receive 2% course credit for their participation.

If you are interested in participating or would like further information please contact
Laura Baker: laura_maree@hotmail.com
Supervisor: Mary Foster

Appendix B

Information Sheet

The Task
For the duration of this experiment you will be working at a computer in a room by yourself. I (the researcher) will not be present during the actual task but will be available if required at any stage. There will be 7 phases that will investigate the way people learn and react to simple perspective taking tasks. The tasks will be made up of four reaction time phases, and three perspective taking phases. During the reaction time tasks you will be presented with a statement on the computer screen, and you task is to press one of two keys that will indicate whether you think the statement is true or false. The reaction time task will be repeated throughout the experiment. There will be three perspective taking phases, a pretest, practice/training, and a posttest. In these phases you will be presented with two statements. The second statement will be about the first statement, and you task is to press one of two keys that will indicate whether you think the statement is true or false.

Time
It is expected that the entire task will take approximately 2hrs to complete. There will be opportunities for you to take short breaks during the experiment. The computer will notify you when these are available.

Course Credit
Students enrolled in PSYC102 or PSYC103 may receive 2% course credit for their participation in this experiment. I will have forms available for this at the time of the experiment.

Withdrawal
Any participant has the right to withdraw from the experiment at any time, for whatever reason, without penalty.

You will have the opportunity to ask any questions you may have at the beginning, and at the completion of the experiment.

Several students will be participating in this research. If you would like information regarding the outcomes of this research upon completion, please contact myself on the details below.

If you have any questions, please ask.

Contact details
Laura Baker
Email: laura_maree@hotmail.com

Supervisor: Mary Foster

Appendix C

University of Waikato
Psychology Department

CONSENT FORM

PARTICIPANT’S COPY

Research Project: Does training decrease response latencies of normally developed adult participants on a perspective taking task?

Name of Researcher: Laura Baker

Name of Supervisor (if applicable): Mary Foster

I have received an information sheet about this research project or the researcher has explained the study to me. I have had the chance to ask any questions and discuss my participation with other people. Any questions have been answered to my satisfaction.

I agree to participate in this research project and I understand that I may withdraw at any time. If I have any concerns about this project, I may contact the convenor of the Research and Ethics Committee (Dr Robert Isler, phone: 838 4466 ext. 8401, e-mail r.isler@waikato.ac.nz)

Participant’sName:______________________Signature:_________________Date:_____

University of Waikato
Psychology Department

CONSENT FORM

RESEARCHER’S COPY

Research Project: Does training decrease response latencies of normally developed adult participants on a perspective taking task?

Name of Researcher: Laura Baker

Name of Supervisor (if applicable): Mary Foster

I have received an information sheet about this research project or the researcher has explained the study to me. I have had the chance to ask any questions and discuss my participation with other people. Any questions have been answered to my satisfaction.

I agree to participate in this research project and I understand that I may withdraw at any time. If I have any concerns about this project, I may contact the convenor of the Research and Ethics Committee.

Participant’sName: ___________________Signature:_______________ Date:_________

Appendix D

Stimulus Sets

Reaction Time Test Trial Questions

Set 1

Trial Reference No.	Statement	Colour of stimulus	Correct Response
1001	This is Red	Red	Z
1002	This is Red	Blue	M
1003	This is Blue	Blue	Z
1004	This is Blue	Red	M
1005	This is Yellow	Yellow	Z
1006	This is Yellow	Green	M
1007	This is Green	Green	Z
1008	This is Green	Yellow	M
1009	This is Orange	Orange	Z
1010	This is Orange	Pink	M
1011	This is Pink	Pink	Z
1012	This is Pink	Orange	M

Perspective Taking Sets

Pre-test
Set 2 cookie jar, doll, cookies
Set 3 toy box, pizza, toy

Training
Set 4 smarties box, pencil, smarties
Set 5 dog kennel, tv, dog

Post-test
Set 2 cookie jar, doll, cookies
Set 4 smarties box, pencil, smarties
Set 6 money box, paper clip, money

Perspective Taking Questions

Stimulus Set 2: Cookie jar, doll, cookies				
Trial Reference No.	**Statement 1**	**Statement 2**	**Correct Response**	**Trial Type Perspective/ Belief/ Response**
1013	If you put the doll in the cookie jar and I am there.	You would think the cookie jar contains the DOLL?	Z	Self/True/True
1014	If I put the doll in the cookie jar and you are there.	You would think the cookie jar contains COOKIES?	M	Self/True/False
1015	If you put the doll in the cookie jar and I am not there.	You would think the cookie jar contains the DOLL?	Z	Self/False/True
1016	If I put the doll in the cookie jar and you are not there.	You would think the cookie jar contains the DOLL?	M	Self/False/False
1017	If you put the doll in the cookie jar and I am there.	I would think the cookie jar contains the DOLL?	Z	Other/True/True
1018	If you put the doll in the cookie jar and I am there.	I would think the cookie jar contains COOKIES?	M	Other/True/False
1019	If I put the doll in the cookie jar and you are not there.	I would think the cookie jar contains the DOLL?	Z	Other/False/True
1020	If I put the doll in the cookie jar and you are not there.	I would think the cookie jar contains COOKIES?	M	Other/False/False
1021	If you photograph the doll in the cookie jar and then I take the doll out.	The photograph will show the cookie jar containing the DOLL?	Z	Photograph/True/ True
1022	If I photograph the doll in the cookie jar and then you take the doll out.	The photograph will show the cookie jar containing COOKIES?	M	Photograph/True/ False
1023	If you photograph the doll in the cookie jar and then I do not take the doll out.	The photograph will show the cookie jar containing the DOLL?	Z	Photograph/False/ True
1024	If I photograph the doll in the cookie jar and then you do not take the doll out.	The photograph will show the cookie jar containing COOKIES?	M	Photograph/False/ False

Stimulus Set 3: Toy box, pizza, toy				
Trial Reference No.	**Statement 1**	**Statement 2**	**Correct Response**	**Trial Type Perspective/ Belief/ Response**
1025	If you put the pizza in the toy box and I am there.	You would think the toy box contains PIZZA?	Z	Self/True/True
1026	If I put the pizza in the toy box and you are there.	You would think the toy box contains TOYS?	M	Self/True/False
1027	If you put the pizza in the toy box and I am not there.	You would think the toy box contains PIZZA?	Z	Self/False/True
1028	If I put the pizza in the toy box and you are not there.	You would think the toy box contains PIZZA?	M	Self/False/False
1029	If you put the pizza in the toy box and I am there.	I would think the toy box contains PIZZA?	Z	Other/True/True
1030	If you put the pizza in the toy box and I am there.	I would think the toy box contains TOYS?	M	Other/True/False
1031	If I put the pizza in the toy box and you are not there.	I would think the toy box contains PIZZA?	Z	Other/False/True
1032	If I put the pizza in the toy box and you are not there.	I would think the toy box contains TOYS?	M	Other/False/False
1033	If you photograph pizza in the toy box and then I take the pizza out.	The photograph will show the toy box containing PIZZA?	Z	Photograph/True/ True
1034	If I photograph pizza in the toy box and then you take the pizza out.	The photograph will show the toy box containing TOYS?	M	Photograph/True/ False
1035	If you photograph pizza in the toy box and then I do not take the pizza out.	The photograph will show the toy box containing PIZZA?	Z	Photograph/False/ True
1036	If I photograph pizza in the toy box and then you do not take the pizza out.	The photograph will show the toy box containing TOYS?	M	Photograph/False/ False

Stimulus Set 4: Smarties box, pencils, smarties				
Trial Reference No.	**Statement 1**	**Statement 2**	**Correct Response**	**Trial Type Perspective/ Belief/ Response**
1037	If you put the pencils in the smarties box and I am there.	You would think the smarties box contains PENCILS?	Z	Self/True/True
1038	If I put the pencils in the smarties box and you are there.	You would think the smarties box contains SMARTIES?	M	Self/True/False
1039	If you put the pencils in the smarties box and I am not there.	You would think the smarties box contains PENCILS?	Z	Self/False/True
1040	If I put the pencils in the smarties box and you are not there.	You would think the smarties box contains PENCILS?	M	Self/False/False
1041	If you put the pencils in the smarties box and I am there.	I would think the smarties box contains PENCILS?	Z	Other/True/True
1042	If you put the pencils in the smarties box and I am there.	I would think the smarties box contains SMARTIES?	M	Other/True/False
1043	If I put the pencils in the smarties box and you are not there.	I would think the smarties box contains PENCILS?	Z	Other/False/True
1044	If I put the pencils in the smarties box and you are not there.	I would think the smarties box contains SMARTIES?	M	Other/False/False
1045	If you photograph pencils in the smarties box and then I take the pencils out.	The photograph will show the smarties box containing PENCILS?	Z	Photograph/True/ True
1046	If I photograph pencils in the smarties box and then you take the pencils out.	The photograph will show the smarties box containing SMARTIES?	M	Photograph/True/ False
1047	If you photograph pencils in the smarties box and then I do not take the pencils out.	The photograph will show the smarties box containing PENCILS?	Z	Photograph/False/ True
1048	If I photograph pencils in the smarties box and then you do not take the pencils out.	The photograph will show the smarties box containing SMARTIES?	M	Photograph/False/ False

Stimulus Set 5: Dog kennel, TV, dog				
Trial Reference No.	**Statement 1**	**Statement 2**	**Correct Response**	**Trial Type Perspective/ Belief/ Response**
1049	If you put the TV in the dog kennel and I am there.	You would think the dog kennel contains the TV?	Z	Self/True/True
1050	If I put the TV in the dog kennel and you are there.	You would think the dog kennel contains the DOG?	M	Self/True/False
1051	If you put the TV in the dog kennel and I am not there.	You would think the dog kennel contains the TV?	Z	Self/False/True
1052	If I put the TV in the dog kennel and you are not there.	You would think the dog kennel contains the TV?	M	Self/False/False
1053	If you put the TV in the dog kennel and I am there.	I would think the dog kennel contains the TV?	Z	Other/True/True
1054	If you put the TV in the dog kennel and I am there.	I would think the dog kennel contains the DOG?	M	Other/True/False
1055	If I put the TV in the dog kennel and you are not there.	I would think the dog kennel contains the TV?	Z	Other/False/True
1056	If I put the TV in the dog kennel and you are not there.	I would think the dog kennel contains the DOG?	M	Other/False/False
1057	If you photograph the TV in the dog kennel and then I take the TV out.	The photograph will show the dog kennel containing the TV?	Z	Photograph/True/ True
1058	If I photograph the TV in the dog kennel and then you take the TV out.	The photograph will show the dog kennel containing the DOG?	M	Photograph/True/ False
1059	If you photograph the TV in the dog kennel and then I do not take the TV out.	The photograph will show the dog kennel containing the TV?	Z	Photograph/False/ True
1060	If I photograph the TV in the dog kennel and then you do not take the TV out.	The photograph will show the dog kennel containing the DOG?	M	Photograph/False/ False

Stimulus Set 6: Money box, paper clip, money				
Trial Reference No.	**Statement 1**	**Statement 2**	**Correct Response**	**Trial Type Perspective/ Belief/ Response**
1061	If you put the paper clip in the money box and I am there.	You would think the money box contains the PAPER CLIP?	Z	Self/True/True
1062	If I put the paper clip in the money box and you are there.	You would think the money box contains MONEY?	M	Self/True/False
1063	If you put the paper clip in the money box and I am not there.	You would think the money box contains the PAPER CLIP?	Z	Self/False/True
1064	If I put the paper clip in the money box and you are not there.	You would think the money box contains the PAPER CLIP?	M	Self/False/False
1065	If you put the paper clip in the money box and I am there.	I would think the money box contains the PAPER CLIP?	Z	Other/True/True
1066	If you put the paper clip in the money box and I am there.	I would think the money box contains MONEY?	M	Other/True/False
1067	If I put the paper clip in the money box and you are not there.	I would think the money box contains the PAPER CLIP?	Z	Other/False/True
1068	If I put the paper clip in the money box and you are not there.	I would think the money box contains MONEY?	M	Other/False/False
1069	If you photograph the paper clip in the money box and then I take the paper clip out.	The photograph will show the money box containing the PAPER CLIP?	Z	Photograph/True/ True
1070	If I photograph the paper clip in the money box and then you take the paper clip out.	The photograph will show the money box containing MONEY?	M	Photograph/True/ False
1071	If you photograph the paper clip in the money box and then I do not take the paper clip out.	The photograph will show the money box containing the PAPER CLIP?	Z	Photograph/False/ True
1072	If I photograph the paper clip in the money box and then you do not take the paper clip out.	The photograph will show the money box containing MONEY?	M	Photograph/False/ False

Appendix E

ANOVA's and *t*-tests for all Trials from the Perspective-taking Phases

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for all the Trials from the Perspective-taking Phases (Pre-test, Training, and Post-test) ANOVAs for the Max 10 and Max 5 Data Sets

Data Set	*df*	F	Partial ή²	Pre-test	Training	Post-test
Max 10	2,38	57.534*	0.752	1.64 (0.57)	0.84 (0.30)	0.74 (0.28)
Max 5	2,38	60.522*	0.761	1.57 (0.52)	0.84 (0.30)	0.74 (0.28)

* significant at $p<0.05$

The df, t value, and Effect Size (Cohen's d) of the Response Latencies for all the Trials from the Perspective-taking Phases t-tests for the Max 10 and Max 5 Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Pre-test and	Max 10	19	7.541*	1.686
Training	Max 5	19	8.263*	1.848
Pre-test and	Max 10	19	8.026*	1.795
Post-test	Max 5	19	7.843*	1.754
Training and	Max 10	19	2.519*	0.563
Post-test	Max 5	19	2.430*	0.543

* significant at $p<0.05$

Appendix F

ANOVAs and *t*-tests for the Question Type (Self, Other, and Photo) from all the Trials from the Perspective-taking Phases

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Type (Self, Other, and Photo) from all the Trials from the Perspective-taking Phases ANOVAs Results for the Max 10 and Max 5 Data Sets

Data Set	*df*	F	Partial ή²	Self	Other	Photo
Max 10	2,38	6.743*	0.262	1.07 (0.32)	1.05 (0.37)	0.96 (0.29)
Max 5	2,38	9.672*	0.337	1.04 (0.31)	1.03 (0.35)	0.94 (0.28)

* significant at $p<0.05$

The df, the t value, and Effect Size (Cohen's d) for the Question Type (Self, Other, and Photo) of the Response Latencies for all the Trials from the Perspective-taking Phases t-tests Results for the Max 10 and Max 5 Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Max 10	19	0.627	0.140
	Max 5	19	0.434	0.097
Self & Photo	Max 10	19	4.055*	0.907
	Max 5	19	4.521*	1.011
Other & Photo	Max 10	19	2.275*	0.509
	Max 5	19	2.952*	0.660

* significant at $p<0.05$

Appendix G

ANOVAs and *t*-tests for the Question Type (Self, Other, and Photo) for the Pre-test, Training and Post-test Phases

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from each of the Pre-test, Training, and Post-test Phases ANOVAs Results for the Max 10 and Max 5 Data Sets

Phase	*df*	F	Partial ή²	Self	Other	Photo
Max 10 Data Set						
Pre-test	2,38	0.453	0.023	1.68 (0.54)	1.66 (0.65)	1.60 (0.64)
Training	2,38	5.374*	0.220	0.91 (0.29)	0.87 (0.42)	0.74 (0.27)
Post-test	2,38	3.479*	0.155	0.77 (0.37)	0.77 (0.28)	0.68 (0.24)
Max 5 Data Set						
Pre-test	2,38	0.994	0.050	1.61 (0.48)	1.60 (0.61)	1.52 (0.58)
Training	2,38	5.334*	0.219	0.90 (0.29)	0.87 (0.42)	0.74 (0.27)
Post-test	2,38	3.245	0.146	0.77 (0.37)	0.77 (0.27)	0.68 (0.24)

* significant at $p<0.05$

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from the Pre-test Phase t-tests Results for the Max 10 and Max 5 Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Max 10	19	0.279	0.062
	Max 5	19	0.053	0.012
Self & Photo	Max 10	19	0.993	0.222
	Max 5	19	1.335	0.298
Other & Photo	Max 10	19	0.578	0.129
	Max 5	19	1.059	0.237

* significant at $p<0.05$

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from the Training Phase t-tests Results for the Max 10 and Max 5 Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Max 10	19	0.058	0.130
	Max 5	19	0.437	0.098
Self & Photo	Max 10	19	3.620*	0.809
	Max 5	19	3.679*	0.823
Other & Photo	Max 10	19	2.542*	0.568
	Max 5	19	2.542*	0.568

* significant at $p<0.05$

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Question Type (Self, Other, and Photo) for the Trials from the Post-test Phase t-tests Results for the Max 10 and Max 5 Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Max 10	19	0.000	0.000
	Max 5	19	0.086	0.019
Self & Photo	Max 10	19	2.082	0.466
	Max 5	19	2.090	0.467
Other & Photo	Max 10	19	2.949*	0.659
	Max 5	19	2.873*	0.643

* significant at $p<0.05$

Appendix H

ANOVAs and *t*-tests for the Question Type (Self, Other, and Photo) from the Trials from the Combined Pre-test and Training Phases

The df, the F value, the Effect Size (Partial ή²), Mean and Standard Deviation (in brackets) of the Response Latencies for the Perspective Question Type (Self, Other, or Photo) for all Trials from the Combined Pre-test and Training Phases, ANOVAs Results for the Max 10, and Max 5 Data Sets

Data Set	*df*	F	Partial ή²	Self	Other	Photo
Max 10	2,38	3.557*	0.158	1.29 (0.35)	1.26 (0.49)	1.17 (0.38)
Max 5	2,38	6.589*	0.268	0.96 (0.18)	0.87 (0.16)	0.89 (0.15)

* significant at $p<0.05$

The df, the t value, and Effect Size (Cohen's d) of the Response Latencies for the Perspective Question Type (Self, Other, or Photo) for all Trials from the Combined Pre-test and Training Phases, t-tests Results for the Max 10, and Max 5 Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Self & Other	Max 10	19	0.590	0.132
	Max 5	18	3.772*	0.865
Self & Photo	Max 10	19	3.728*	0.833
	Max 5	18	2.491*	0.571
Other & Photo	Max 10	19	1.578	0.353
	Max 5	18	-1.007	-0.231

* significant at $p<0.05$

Appendix I

t-tests for the Question Belief for the Trials from all the Perspective-taking Phases, and for the Trials from the Pre-test, Training, and Post-test Phases

t-tests for the Question Belief (True and False) for the Trials from all the Perspective-taking Phases

The df, the t value, the Effect Size (Cohen's d), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Belief (True and False) for all the Trials from the Perspective-taking Phases t-tests for the Max 10 and Max 5 Data Sets

Data Set	*df*	*t*	Cohen's *d*	True Belief	False Belief
Max 10	19	-1.643	-0.367	1.00 (0.32)	1.05 (0.33)
Max 5	19	-1.675	-0.375	0.98 (0.31)	1.03 (0.32)

* significant at $p<0.05$

t-tests for the Question Belief (True and False) for the Trials from the Pre-test, Training, and Post-test Phases

The df, the t value, the Effect Size (Cohen's d), Mean and Standard Deviation (in brackets) of the Response Latencies for the Question Belief (True and False) for the Trials from the Pre-test, Training, and Post-test Phases t-tests Results for the Max 10 and Max 5 Data Sets

Phase	*df*	*t*	Cohen's *d*	True Belief	False Belief
Max 10 Data Set					
Pre-test	19	-0.781	-0.175	1.61 (0.59)	1.68 (0.61)
Training	19	-2.320*	-0.519	0.79 (0.31)	0.89 (0.33)
Post-test	19	0.170	0.038	0.74 (0.32)	0.74 (0.26)
Max 5 Data Set					
Pre-test	19	-0.577	-0.129	1.55 (0.54)	1.60 (0.56)
Training	19	-2.416*	-0.540	0.78 (0.31)	0.89 (0.33)
Post-test	19	0.090	0.020	0.74 (0.31)	0.74 (0.26)

* significant at $p<0.05$

Appendix J

Means, Standard Deviations, ANOVAs and t-tests for the Trials from Each Set from the Perspective-taking Phases

Means and Standard Deviations (in brackets) of the Response Latencies for Each Set from the Perspective-taking Phases (Pre-test, Training, and Post-test) for the Max 10 and Max 5 Data Sets

Data Set	Pre-test		Training		Post-test		
	Set 2	Set 3	Set 4	Set 5	Set 2	Set 4	Set 6
Max 10	1.71(0.72)	1.57(0.51)	0.91(0.37)	0.77(0.28)	0.70(0.29)	0.82(0.29)	0.69(0.29)
Max 5	1.63(0.63)	1.51(0.47)	0.91(0.37)	0.76(0.28)	0.70(0.27)	0.82(0.29)	0.69(0.29)

The df, the F value, the Effect Size (Partial ή²) of the Response Latencies for the Sets from the Perspective-taking Phases ANOVAs Results for the Max 10 and Max 5 Data Sets

Data Set	*df*	F	Partial ή²
Max 10	6,114	41.494*	0.686
Max 5	6,114	46.890*	0.712

* significant at $p<0.05$

The df, the F value, the Effect Size (Partial ή²) of the Response Latencies for the Sets from the Perspective-taking Phases t-tests Results for the Max 10 and Max 5 Data Sets

	Data Set	*df*	*t*	Cohen's *d*
Set 2 pre-test &	Max 10	19	6.987*	1.562
Set 2 post-test	Max 5	19	7.463*	1.669
Set 4 training &	Max 10	19	1.424	0.318
Set 4 post-test	Max 5	19	1.389	0.311
Set 2 pre-test &	Max 10	19	6.777*	1.515
Set 6 post-test	Max 5	19	7.126*	1.593
Set 3 pre-test &	Max 10	19	8.315*	1.859
Set 6 post-test	Max 5	19	8.980*	2.008

* significant at $p<0.05$