Cognition 138 (2015) 21–34

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

From self to social cognition: Theory of Mind mechanisms and their relation to Executive Functioning

Elisabeth E.F. Bradford *, Ines Jentzsch *, Juan-Carlos Gomez *

School of Psychology & Neuroscience, University of St Andrews, St Andrews, Scotland KY16 9JP, UK

ARTICLE INFO

Article history:
Received 30 January 2014
Revised 27 January 2015
Accepted 1 February 2015
Available online 19 February 2015

Keywords:
Theory of Mind
False belief
Belief attribution
Social cognition

ABSTRACT

'Theory of Mind' refers to the ability to attribute mental states to oneself and other people (Premack & Woodruff, 1978). This study examined the extent to which 'Self' and 'Other' belief-attribution processes within the Theory of Mind (ToM) mechanism could be distinguished behaviourally, and whether these separable components differentially related to Executive Functioning (EF) abilities. A computerized false-belief task, utilizing a matched-design to allow direct comparison of self-oriented vs. other-oriented belief-attribution, was used to assess ToM, and a face-image Stroop task was employed to assess EF, within a population of typically-developed adults. Results revealed significantly longer reaction times when attributing beliefs to other people as opposed to recognizing and attributing beliefs to oneself. Intriguingly, results revealed that 'perspective-shift' requirements (i.e. changing from adoption of the 'self' perspective to the perspective of the 'other', or vice versa) across false-belief trials influenced reaction times. Reaction times were significantly longer when the perspective shift was from self-to-other than from other-to-self. It is suggested that the 'self' forms the stem of understanding the 'other', and is therefore processed regardless of ultimate task demands; in contrast, the 'other' perspective is only processed when explicitly required. We conclude that adopting another person's perspective, even when their belief state is matched to one's own, requires more cognitive effort than recalling and reflecting on self-oriented belief-states.

1. Introduction

The ability to understand and attribute mental states, including intentions, knowledge and desires, to both ourselves and other people, is referred to as possession of a 'Theory of Mind' (Premack & Woodruff, 1978; Baron-Cohen, Leslie, & Frith, 1985). These 'mentalizing' abilities form an essential and fundamental role in many social and communicative interactions, allowing successful and mutual exchanges of information between individuals (Hamilton, 2009; Ahmed & Miller, 2011). The mechanisms underlying Theory of Mind (ToM) abilities are not as yet clear. Recently, evidence has supported the notion of a modular structure underlying ToM abilities, with separate component parts involved in specific, differing mentalizing processes (e.g. Bodden et al., 2010; Decety & Sommerville, 2003; Harari, Shamay-Tsoory, Ravid, & Levkovitz, 2010). One such proposed delineation is between self-oriented and other-person oriented mental state attribution, where the ability to reflect on one's own mental states ('self') may utilize distinct mechanisms from those used in attributing and understanding mental states of the 'other' (e.g. Decety & Sommerville, 2003; Hartwright, Apperly, & Hansen, 2012; Jardri et al., 2011; Jeannerod & Anquetil, 2008).

The purpose of the present study was to determine whether 'Self' and 'Other' belief-attribution processes, a

* Tel.: +44 (0)1334 463 060 (E.E.F. Bradford).
E-mail addresses: eefb@st-andrews.ac.uk (E.E.F. Bradford), ij7@st-andrews.ac.uk (I. Jentzsch), jg5@st-andrews.ac.uk (J.-C. Gomez).

http://dx.doi.org/10.1016/j.cognition.2015.02.001
0010-0277/© 2015 The Authors. Published by Elsevier B.V.

part of the ToM mechanism, could be differentiated using behavioural measures. Some prior research has examined the 'Self'/'Other' distinction within ToM, as will be discussed below. However, the present study is the first, to our knowledge, to use a matched design in a false-belief task, allowing creation of directly comparable conditions of self/other belief-attribution processes. A secondary aim of the current study was to assess the extent to which these ToM components, if found to be separable, are driven by differing aspects of Executive Functioning, as some prior research has begun to indicate (e.g. Ahmed & Miller, 2011; Brent, Rios, Happé, & Charman, 2004; German & Hehman, 2006). Executive Function (EF) refers to a set of cognitive processes that regulate, control and manage other cognitive processes, including inhibition, working memory, cognitive flexibility, and planning (Miyake et al., 2000; Miyake & Friedman, 2012; Carlson & Moses, 2001). Evidence has suggested a strong association between ToM and EF abilities (e.g. Ozonoff, Pennington, & Rogers, 1991; Sabbagh, Xu, Carlson, Moses, & Lee, 2006; Ozonoff & McEvoy, 1994), and Ahmed and Miller (2011) suggested that by examining the relationship between ToM and EF, researchers may be able to gain better insight and understanding of the mechanisms underlying ToM.

False-belief tasks are one of the tests most often used to assess ToM abilities in both typically and atypically developed individuals (Brewer, 1991). False-belief tasks involve scenarios in which individuals are shown a situation where reality states differ from belief states, and where a clear distinction between one's own current belief states and the current belief states of another individual is created. One of the first false-belief tasks designed to assess self and other belief-attribution abilities was Perner, Leekham, and Wimmer's (1987) 'Smarties' task. In this task, children were shown a box of sweets ('Smarties') and asked what they thought would be inside. On responding sweets/chocolate, children were shown that the box actually contained pencils. The pencils were then re-hidden, and children were asked three critical questions, akin to the following: '*What did you think was in the box, before you saw inside?*' (self-oriented belief attribution); '*What would your teacher, who hasn't seen inside, think was in the box?*' (other-oriented belief attribution); and '*What was really in the box?*' (reality test).

Converging evidence, from both the Smarties task and other verbally explicit false-belief paradigms (e.g. Wimmer & Perner, 1983; Gopnik & Astington, 1988; Williams & Happé, 2010), suggests that a developmental shift is undertaken by typically developing children at the age of about 4 years; prior to this age, children tend to fail ToM tasks, unable to inhibit their own current knowledge states. From the age of 4-years, however, a rapid improvement in ToM abilities is seen, with children successfully able to recognize separate and differing mental states of other people, acknowledging, for example, that they themselves know that there are pencils in the box, but another person, who hasn't seen inside, would think it contains chocolates (Wimmer & Perner, 1983; Perner et al., 1987; Wellman, Cross, & Watson, 2001; Doherty, 2009). The seemingly simultaneous emergence of these capacities may indicate a single 'ToM' mechanism, with no detectable differentiation in the development of self vs. other-oriented belief-attribution abilities.

However, cases in which ToM abilities fail to fully develop or are disrupted due to illness or injury provide evidence for the occurrence of deficits that may differentially affect self and other ToM. For instance, a particular focus of past research has been Autistic Spectrum Conditions (ASC), a defining feature of which is difficulties with ToM abilities (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Brent et al., 2004; Frith, 1989; Hillier & Allinson, 2002; Lombardo et al., 2010). Individuals diagnosed with ASC are often found to display egocentric behaviours in ToM tasks akin to those seen in typically developing children prior to the age of 4-years. Results about the type of ToM deficits experienced by individuals with ASC have been mixed, with suggestions of specific deficits in attributing mental states to other people (e.g. Hutchins, Prelock, & Bonazinga, 2011; Wimmer & Perner, 1983), problems with reflecting on one's own mental states (e.g. Williams & Happé, 2010), or deficits in both self-oriented and other-person oriented belief-attribution (e.g. Carlson, Moses, & Breton, 2002; Brent et al., 2004; Perner, Frith, Leslie, & Leekham, 1989). Most commonly, however, individuals with ASC are suggested to show a particular deficit in their social cognition abilities, specifically reduced in their capacity to comprehend and understand differing mental states of other people, which supports the notion of differentiation between the 'self' and 'other' in ToM processes (Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Tager-Flusberg, 2007).

Additional evidence for the distinction between 'self' and 'other' has been offered by Samson, Apperly, Kathirgamanathan, and Humphreys (2005), who present a case study of an adult patient (WBA) with a right fronto-temporal brain lesion. Patient WBA was found to exhibit a specific problem with inhibiting the 'self-perspective', whilst retaining the ability to infer someone else's perspective. The patient could understand that other people's knowledge may vary from his own, and could successfully attribute mental states to them, but only if his own personal knowledge did not contradict the other person's, or was not too salient. For instance, if patient WBA knew the true location of an object in a false-belief task, he was unable to inhibit an egocentric pre-potent response; however, if he did not know the true location of the object, although he knew that the object had been moved and the naïve other was therefore in possession of a false-belief, he was able to successfully attribute a belief to the other person. In this way, Samson et al. (2005) argued that there is a distinction between processing of the 'self' perspective and processing of the 'other' perspective.

To further explore the extent of this differentiation, some prior literature has focused on the extent to which 'self' and 'other' perspectives are processed automatically, with results suggesting that the very presence of a secondary agent (the 'other') can influence the behaviour of an individual (e.g. Samson, Apperly, Braithwaite, & Andrews, 2010; Kovács, Téglás, & Endress, 2010). Samson et al. (2010), for instance, reported a study in which participants were slower to report the number of dots visible in a room when a secondary agent could not see all the dots; they suggest this implies that individuals automatically take note of the fact that others may have different knowledge states given their different point of view, even when this information is irrelevant to the task at hand. Kovács et al. (2010) reported a study in which both adults' and 7-month-old infants' behaviours (reaction times and looking times, respectively) appeared to be influenced by an automatic encoding of another agent's beliefs, even when the beliefs of the agent were irrelevant to the actual task.

Qureshi, Apperly, and Samson (2010) suggest a distinction between *awareness* of differing knowledge states and explicit *adoption* of another person's perspective; simply knowing that someone's mental states differ from one's own is not necessarily the same as being explicitly aware of what the other person's mental state *is* (see also: Schneider, Slaughter, & Dux, 2014). Supporting this dissociation between 'awareness' and 'adoption' of another person's mental states, Apperly, Riggs, Simpson, Samson, and Chiavarino (2006) conducted a study in which participants were either instructed to track an individual's beliefs, or were not instructed to do so but were given unexpected questions about the other person's beliefs. Participants watched a video of an object being placed in a box. An actor then left the scene, and the object was moved to a different location, creating a false-belief state for the actor. In one condition of this task, participants were instructed to keep track of where the object was located (reality state). In a second condition, participants were instructed to track where the object really was and, importantly, where the actor thought the object was (belief-attribution). Results revealed that when instructed to track the actor's belief states, there was no difference in time taken for participants to identify either the true location of the object, or to identify where the actor believed the object to be. In contrast, in the condition in which they were simply told to track the location of the object (and not the belief state of the actor), participants took significantly longer to identify where the actor believed the object was compared to identifying where the object was actually located. Contradicting Samson et al. (2010) and Kovács et al. (2010), these results suggest that other-oriented belief-tracking is not an automatic process, and may only be undertaken when explicitly required. Back and Apperly (2010) further support these findings, suggesting that others' beliefs are not automatically inferred, but rather require explicit motivation to track this information.

As discussed above, previous studies have suggested that a differentiation between 'Self' and 'Other' is present when considering and attributing mental states, and there remains some debate as to the extent to which the 'other' perspective is automatically processed. However, an issue with these studies is that the 'self' and 'other' conditions are often difficult to compare directly, because different task requirements are used for self/other conditions. For instance, in Samson et al.'s (2010) study, response times where compared between conditions in which a secondary agent's view was either matched to the participant's (i.e. they could both see the same number of dots on a wall) or mis-matched (i.e. the participant could see a different number of dots to the agent). Similarly in Apperly et al.'s (2006) study, the comparative conditions refer to other-oriented belief states and reality states, which may arguably be resolved from recall of a memory, rather than requiring an explicit judgement of one's own belief states. The current study sought to address these problems by developing a false-belief task in which belief-states were created from either the other-perspective or the self-perspective, and remained matched in structure and formation (i.e. the same sentence set-up and task requirements were introduced, whether ultimately the final question was self or other oriented). In this way self and other belief-attribution processes could be explicitly and directly compared under the same task requirements.

Much documented evidence points to a strong association between ToM abilities and Executive Functioning (EF) capacities (e.g. Ozonoff et al., 1991; Ozonoff & McEvoy, 1994; Sabbagh et al., 2006). In typically developed adults, it has been demonstrated that a heavier cognitive load can influence success on ToM tasks, often resulting in more egocentric errors, supporting the notion of a relationship between ToM and EF (e.g. Bull, Philips, & Conway, 2008; Epley, Keysar, Van Boven, & Gilovich, 2004). Keysar, Barr, Balin, and Brauner (2000) report a study in which two adults (a participant and a confederate, the 'director') sat on opposite sides of a table, with a vertical grid between them. Some of the slots on the grid were occluded so that only the participant, but not the director, could see them, creating distinct self and other-person perspectives. The task required participants to follow the director's instructions in reassigning positions to various objects. Some objects were unique (e.g. a truck) and others were ambiguous, such as three candles (small, medium, large). In the case of the candles, the smallest candle, from the participant's view, may have been occluded so that, if the director said '*move the small candle*', he would be referring to the medium sized candle, which is the smallest from his perspective. Using eye-tracking equipment, Keysar et al. (2000) found that participants were sometimes biased by an egocentric interpretation of instructions – even when they clearly knew that information was shared or privileged, they still focused on the occluded objects as potential solutions. These findings have been replicated across different studies, suggesting a resilient effect of egocentric biases in adults in implicit tests of ToM (e.g. Apperly et al., 2010; Surtees, Butterfill, & Apperly, 2012). These results indicate that even in typically developed adults, who arguably should possess the cognitive mechanisms necessary for successful ToM expression, EF abilities (e.g. inhibition, attention direction) may not automatically be recruited when ToM is assessed implicitly, and may only be utilized when explicitly required.

Some contradictory evidence about the relationship between ToM and EF has also been reported. Fine, Lumsden, and Blair (2001) report a case study of patient B.M., who suffered from amygdala damage and had been diagnosed with schizophrenia and Asperger's syndrome. Patient B.M. was shown to be severely impaired in his ability to represent mental states (ToM) but did not show any impairment in his EF abilities (Fine et al., 2001). However, in the majority of cases, studies continue to find a link between EF and ToM abilities, both within clinical populations (e.g. Joseph & Tager-Flusberg, 2004; Ozonoff & McEvoy, 1994), and in non-clinical populations (e.g. Carlson et al., 2002; Hughes, 1998). Further, the EF-ToM relationship has been demonstrated across both Western and non-Western societies, such as Sabbagh et al.'s (2006) study showing a relationship between ToM and EF in pre-schoolers from both the United States and China.

Whilst a link between EF abilities and ToM expression abilities has been mostly supported, the specifics of this relationship are as yet unclear. For instance, if it is indeed the case that ToM is constructed of separable component parts, it is possible that some specific aspects of ToM will be more demanding of EF than others. Further, different aspects of EF may be more important than others in the expression of ToM capacities (e.g. Bull et al., 2008; German & Hehman, 2006). For instance, Carlson et al. (2002) suggest that inhibition and working memory have the strongest relationship with general ToM abilities (i.e. inhibitory control as opposed to planning abilities). These results were supported by German and Hehman (2006) who found that increasing executive demands resulted in decreased performance on ToM tasks in both younger and older adults, and that inhibition abilities and processing speed were key in determining ToM task performance. They suggest that deficits in ToM may be the result of a deficit in EF abilities, rather than necessarily a deficit in the ToM mechanism specifically. McKinnon and Moscovitch (2007) also assessed performance of older and younger adults on ToM tasks and EF tasks. Results revealed that for all adults, ToM performance varied with the demands placed on EF abilities, with decreased ToM performance co-occurring with increased EF demands (e.g. working memory).

A potential differentiation that has been suggested within EF is that of affective versus non-affective EF (e.g. Reeck & Egner, 2011). Reeck and Egner (2011) found that affective distracters in a Stroop task influenced participant's performance, regardless of whether the aim of the task was to identify an emotion (task-relevant) or the gender (task-irrelevant) of a presented face. In contrast, non-affective distracters only influenced performance when the task was relevant (i.e. identify gender) and not when the task was irrelevant to the distracter (i.e. identify emotion). Given this asymmetry between affective and non-affective stimuli, it is argued that they may activate separable EF abilities; affective tasks may tap into more mentalizing abilities due to the need to attribute an emotional state to a static face, than non-affective tasks, where observation of traits may be enough to attribute a specified gender, without requiring any mental state attribution. The emotional tasks could therefore arguably require entirely separate processing from non-emotional tasks, or additional processing to meet task demands, potentially resulting in mobilisation of extra EF resources. This distinction provides an opportunity to explore whether these two components differentially relate to ToM abilities, further allowing investigation of whether these two EF abilities are indeed separable, or whether they are simply reflecting two aspects of one overarching EF capacity.

The current study sought to explore the extent to which self-oriented and other-oriented ToM could be differentiated using behavioural measures (reaction times and error rates) in a non-clinical, typically developed adult population, and the extent to which these ToM components differentially relate to EF abilities. Studying this relationship in adults will allow insights about the importance of EF in the activation and maintenance of ToM abilities in adulthood, when both of these abilities are arguably fully developed (Ahmed & Miller, 2011; Anderson, Levein, & Jacobs, 2002; Apperly, Samson, & Humphreys, 2009). To assess self/other differentiations, a novel task based on the standard false-belief paradigm was used, in which participants completed a computerized version of a false-belief task, followed by probe questions about either their own belief state or that of another person. The novelty of this task is that, as far as the authors are aware, this is the first task in which self and other belief-attribution questions are explicitly matched in design to allow direct comparison between the two conditions. Further, the study adopts a novel approach in distinguishing the requirement for participants to explicitly reflect on their own prior *beliefs*, rather than potentially utilizing recall of prior memory, without necessarily requiring belief-attribution. A face/word image Stroop task was also employed, with affective and non-affective conditions, to investigate the extent to which different ToM components (self and other) may relate to separable EF abilities.

If, as Kovács et al. (2010) suggest, the perspectives and beliefs of other people are automatically processed alongside the self-perspective, regardless of ultimate task demands, we would not expect there to be any significant difference in response times between self or other-oriented belief attribution trials in our Self/Other Differentiation Task, as, in both cases, the response should already be known to participants. However, if the 'other' is only processed when the task explicitly requires it (e.g. Apperly et al., 2006; Back & Apperly, 2010), longer response times would be expected in the 'other' perspective conditions than in the 'self' perspective conditions, as the 'other' would not have been considered until required. Within the Self/Other Differentiation task, participants are asked to solve problems from either their own or another person's perspective. In some trials, participants are asked to switch perspectives, where they would begin a trial by answering a dilemma from one perspective (e.g. '*Where would John look for some bread?*' – other-oriented) and end a trial by answering a probe question from another perspective (e.g. '*What did you think was in the box?*' – self-oriented).

Of interest in this study is the effect of the presence or absence of a perspective shift within a trial, and whether being required to shift between self- and other-oriented perspectives within a single trial influences response times. If it is the case that the 'other' is always processed, no significant difference in reaction times should be found between self-to-other and other-to-self trials in perspective shift conditions, as both self and other perspectives should have already been considered, regardless of task requirements. If, however, only the 'self' is automatically processed, unless task demands explicitly require the 'other' to be considered, then reaction times should be faster in the other-to-self perspective shift trials than in the self-to-other perspective shift trials; in the former, the 'self' will have already been processed, to allow understanding of the 'other' perspective, whereas in the latter, consideration of the perspective of the other has not yet been required.

Prior research has suggested that affective and non-affective Stroop tasks may tap into different abilities, primarily due to the finding that affectively salient stimuli are prioritized above non-affective stimuli in cognitive processing, capturing attention and disrupting task performance at a higher rate (Reeck & Egner, 2011). In the current study, correlation analysis will be conducted to explore the relationship between performance on the Self/Other Differentiation task and affective and non-affective versions of the Stroop task (EF abilities), to assess whether these two components differentially relate to ToM abilities. It is predicted that higher levels of cognitive control, as indicated by affective and non-affective Stroop scores, will positively correlate with mentalizing performance (i.e. belief attribution assessed using response times at the probe stage of the Self/Other Differentiation Task). That is, better performance on the Stroop task (i.e. higher levels of inhibition abilities) would correlate with better performance on the ToM task (i.e. ability to switch between Self/Other perspectives when attributing beliefs). Further, affective Stroop task scores may more strongly correlate with ToM task performance than non-affective Stroop tasks, as they may tap into more 'mentalizing' abilities (identifying emotion expressions) than the non-affective task (gender identification).

2. Method

2.1. Participants

Sixty-two adult participants (50 females; mean age 22.8 years, range 18–55) were recruited from St Andrews University, Scotland. All participants were required to be either native or fluent in English, and were reimbursed at a rate of £5 per hour. All participants gave informed consent and this study was approved for use in human subjects in accordance with the University of St Andrews Research Ethics committee.

2.2. Procedure

Participants were tested in a single session, lasting between 60 and 90 min. Participants completed the Self/Other Differentiation Task and the Stroop task as part of a larger task battery, which will not be presented in this paper. All tasks were separated by a break in the session. The order of tasks was randomized across participants. After completion of testing, participants received a written and verbal debriefing of the study's aims.

2.3. Measures

2.3.1. Self/Other Differentiation Task

This computerized task was programmed using E-Prime. The task consisted of 12 practice trials and 60 test trials, each consisting of three stages: dilemma stage, contents revelation stage, and probe stage. Fig. 1 provides a pictorial depiction of each of these stages and a timeline illustrating presentation order. The practice and test trials both followed the same format, but only the test trials required belief attribution. Participants were asked to respond as quickly and accurately as possible, both before the practice trials began and again before the test trials commenced. A fixation cross (+) was shown for 1000 ms between each trial. Fig. 2 illustrates each of the different test conditions in the Self/Other Differentiation Task.

Dilemma Stage – used to identify a belief-state, in the dilemma stage participants were asked to select a container from three images presented in a horizontal line, in which they would look for a specific object (e.g. '*Ruth has been saving up some money at home. Where would she look for it?*' with the options of a plastic container, piggy bank, or cat carrier presented below). Dilemmas could be either self-oriented ('*Where would you look?*'; 30 trials) or other-oriented ('*Where would John look?*'; 30 trials). In practice trials, participants were simply asked to select a specific object (e.g. '*Select the yellow bag.*'). The questions were shown for 1500 ms before the image options were displayed. Image answer options were displayed for a maximum of 5000 ms. If participants did not select the correct container or failed to make a selection in this time, an 'X' was displayed for 1000 ms, before resetting the dilemma until the correct choice was made.

Contents Revelation Stage – having selected the correct container at the dilemma stage, contents of the container was then revealed. Contents could either be expected (e.g. coins in the piggy-bank) or unexpected (e.g. earrings in the piggy-bank). Contents were shown for 2000 ms. Following self/other-oriented dilemmas, half of each were followed by expected contents, the other half by unexpected contents. In practice trials, contents was always expected (e.g. a wallet in a bag).

Probe Stage – the critical probe stage was used to assess belief-attribution abilities. Participants were asked to identify what either they themselves (self-oriented) or another person (other-oriented) believed was in the container, before they saw inside. The probe questions were matched in terms of syntax and syllables, so that each sentence had at least one paired match in the opposite condition (Self vs. Other). For self-oriented questions, the probe could, for example, be either: '*Before you saw what was inside, what did you think was in the container?*' or '*What would you think was in the container if you hadn't seen inside?*'. For other-oriented questions, the probe could be, for example: '*Before John saw what was inside, what did he think was in the container?*' or '*What would someone who hasn't seen inside think was in the container?*'. There were an equal number of question forms across each trial type (see Table 1 for details of trial types and the number of trials in each condition). All names used in 'Other' conditions were one syllable, ensuring the lengths of sentences were not altered.

Answers were again given by selection of one image from three presented in a horizontal line. Distracter questions were also included at the probe stage (e.g. '*What colour was the egg box?*'), to try and reduce participant's anticipation of the correct answer until they saw the probe

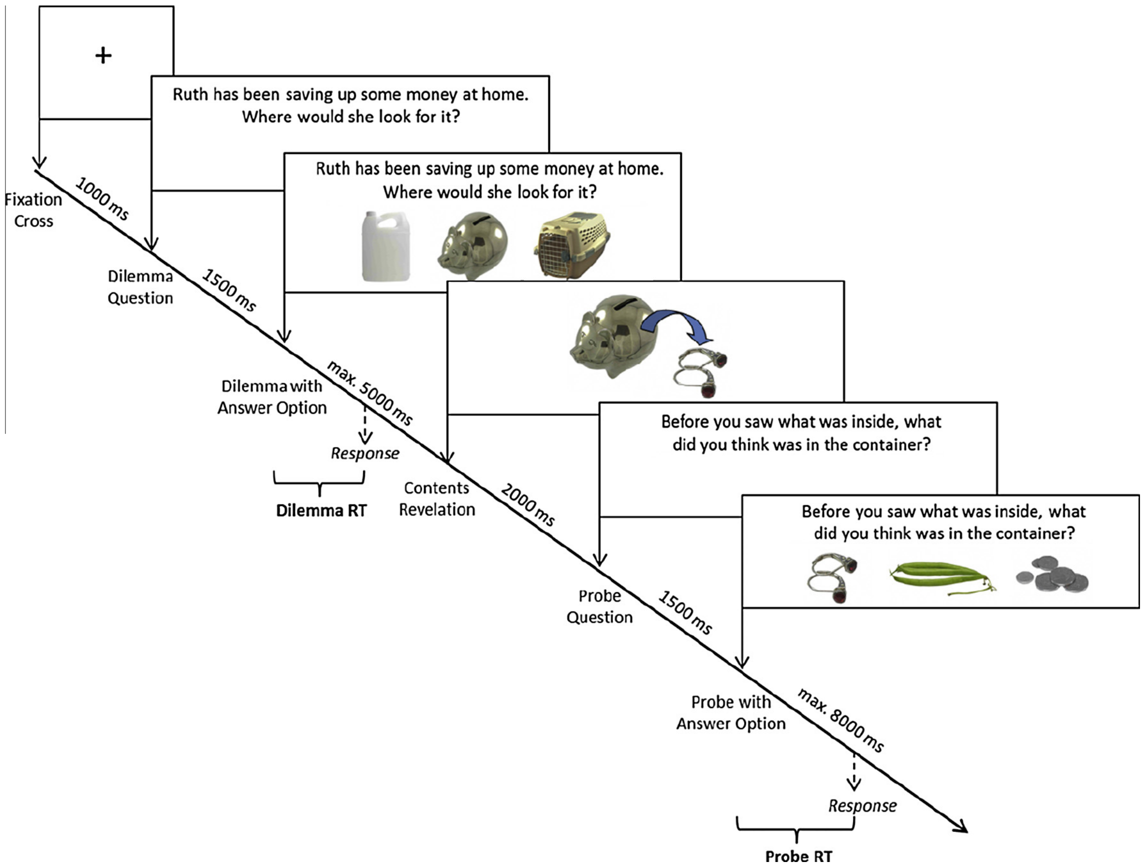

Fig. 1. Illustration of the different stages of a single trial in the Self/Other Differentiation Task: the Dilemma Stage ('Self' or 'Other'), Contents Revelation Stage (Expected or Unexpected), and Probe Question Stage ('Self', 'Other', or Distracter).

question. Two reality test questions were included within the distracter questions (i.e. '*What was actually in the container?*'), one following expected contents and one following unexpected contents, to ensure participants understood the task. The probe question was displayed for 1500 ms before the three answer options were revealed. Image answer options were displayed for a maximum of 8000 ms. If an incorrect response (including no response) was recorded, the task continued on to the next trial without resetting. In practice trials, participants were asked to select the picture that showed what had been in the bag (i.e. reality/true belief).

Stimuli were presented on a 17 inch CRT monitor. The keyboard was used to record responses, with participants pressing the key that corresponded with the spatial location of the correct image. Keys used were 1, 2, and 3 from the number pad on the right side of the keyboard; if the correct answer was on the left, the correct response would be '1', the centre '2' and the right '3'. Coloured images, from the Bank of Standardized Stimuli (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and the Amsterdam Library of Object Images (Geusebroek, Burghouts, & Smeulders, 2005), were presented on a white background, in the centre of the screen. Questions were presented in black text, using 'calibri' font. Reaction time was recorded at both dilemma and probe stages, from the moment that the three images of potential answers were presented.

2.3.2. Stroop Task

Stimuli consisted of black and white face images from Ekman and Friesen (1976), with word labels placed directly beneath the image. The stimulus set consisted of four female and four male individuals' faces, each with both happy and sad expressions, resulting in a total of 16 distinct face stimuli. Images were standardized and cropped to an oval shape consisting only of the main facial features, removing non-face information, such as hair, which could have facilitated performance during gender discrimination. Each of the 16 face images was paired with four different word labels (gender – 'male', 'female', and emotion – 'happy', 'sad') that were placed directly beneath the images (black ink, upper case lettering), resulting in 64 face-word stimuli sets. Fig. 3 provides examples of the types of stimuli employed. Stimulus presentation and data collection used E-Prime software.

There were two task types within this experiment: gender-identification task and emotion-expression identification task. All participants took part in both tasks, with task order counter-balanced between participants. On each trial, participants were presented with a face-word

Fig. 2. Examples of each of the different test conditions in the Self/Other Differentiation task. Letters indicate different stages of individual trials: (a) Dilemma Question (Self/Other) – used to establish a belief state, before image answer options (b) were shown. (c) Contents Revelation – creating either a false-belief (unexpected contents) or a true-belief (expected contents) scenario. (d) Probe Question (Self/Other), after which image answer options (e) were added. Distracter trials were also included at the Probe stage, but only critical test trials (Self/Other) are shown in Fig. 2.

Table 1
Number of trials in each condition combination in the Self/Other Differentiation Task. Reality test questions are indicated in parenthesis.

	Self dilemma			Other dilemma			
	Self probe	Other probe	Distracter	Self probe	Other probe	Distracter	
Expected contents	4	4	6 (1)	4	4	7	*30*
Unexpected contents	4	4	7	4	4	6 (1)	*30*
Total	*8*	*8*	*14*	*8*	*8*	*14*	*60*

stimulus and instructed to make a judgement of the face, as quickly and accurately as possible, whilst ignoring the word label. In the gender trials, participants were asked to make a gender judgement (is the face male or female?), and in the emotion trials, participants were asked to make an expression judgment (is the face happy or sad?). Responses were made via keyboard presses using the index fingers from the right and left hands on keys 'Z' and 'M'. Stimuli were presented until a response was recorded, or for a maximum of 2000 ms if no response was made. Between trials, a fixation cross (+) was presented for 1000 ms, in the centre of the screen.

The distracter labels shown underneath the face image stimuli could be either semantically congruent or incongruent with the target face image. For instance, a happy male face could be accompanied by a congruent 'male'/'happy' label, or by an incongruent 'female'/'sad' label. The distracter could also be either task-relevant or task-irrelevant (i.e. gender words in the emotion-identification task, emotion words in the gender-identification task). Participants completed four blocks of 64 trials within each task type; within these, two blocks included task-relevant distracters, and two blocks included task-irrelevant distracters. Within both task-relevant and task-irrelevant blocks, half of the trials used congruent stimuli, the other half incongruent stimuli, with stimuli presented in a randomized order. Therefore, there were a total of 256 test trials in each of the gender and emotion identification tasks. At the beginning of each task type (i.e. emotion/gender identification), participants completed a short practice session (8 trials). Order of block presentation (i.e. task relevant/task irrelevant) within the gender/emotion tasks was randomized across participants.

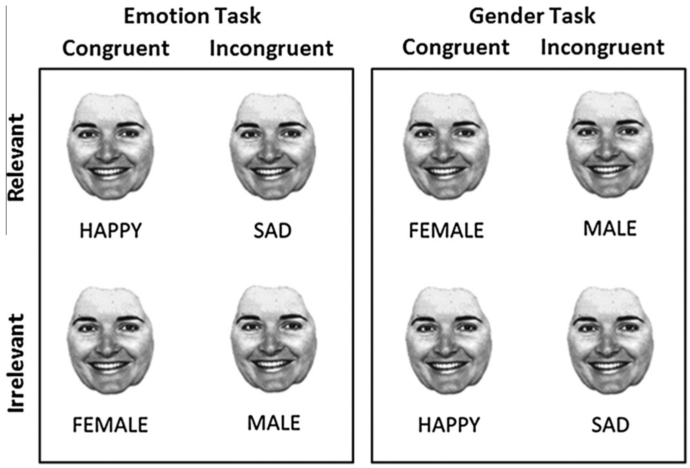

Fig. 3. Examples of experimental stimuli for the emotion identification task and the gender identification task. For simplicity, only one face-image (female, happy) is shown here, in each of the task conditions.

3. Results

Participants were screened for outliers in overall reaction times. That is, we checked for individual participants' mean response times that were outside three standard deviations from the overall reaction time mean. No outliers were found, so all 62 participants were retained for analysis.

3.1. Self/Other Differentiation Task

3.1.1. Response time data

Only correct responses were considered for response time (RT) analyses. RT data were separately analysed for Dilemma and Probe stages. That is, if an error was committed at the Dilemma stage, the RT to the dilemma was excluded from analysis of dilemma RTs. However, following this error, if the answer was correct at probe stage, the Probe stage RT was included in analysis, as these were regarded as two separate time points.[1] All correct responses given within the set response window were included in analysis.

3.1.1.1. Dilemma Stage. A *t*-test revealed a significant difference in RT to dilemma questions as a function of dilemma type, $t(61) = -5.99$, $p < .001$, $r = .15$, with faster responses to self-oriented dilemmas ($M = 1582$ ms) than to other-oriented dilemmas ($M = 1763$ ms).

3.1.1.2. Probe Stage. For analysis of RT to the probe question, the relationship between Dilemma Type and Probe Type was taken into account, resulting in a factor of 'Perspective Shift'. For trials in which there was no perspective shift, both the Dilemma and Probe addressed either the self (Self-Self) or the other (Other-Other). For trials in which there was a perspective shift between Dilemma and Probe stages, the shift could be from either the 'self' at dilemma stage to the 'other' at probe stage (Self-Other), or from the other at the dilemma stage to the self at the probe stage (Other-Self).

RTs to probe questions were analysed using a 2 (Perspective Shift: No Shift vs. Shift) × 2 (Contents Type: Expected vs. Unexpected) × 2 (Probe Type: Self vs. Other) Repeated-Measures ANOVA.

There was a significant main effect of Probe Type, $F(1, 61) = 17.78$, $p < .001$, *partial* $\eta^2 = .23$, with faster responses to self-oriented probes ($M = 932$ ms) than other-oriented probes ($M = 1093$ ms). This effect was modulated by Perspective Shift, resulting in a significant interaction between Perspective Shift and Probe Type, $F(1, 61) = 5.27$, $p = .03$, *partial* $\eta^2 = .08$. This interaction was due to larger Probe Type effects (Other minus Self) in perspective shift trials ($M = 244$ ms) than in no-perspective shift trials ($M = 78$ ms). In other words, within perspective-shift conditions, it was harder to shift from Self-to-Other ($M = 1139$ ms) than from Other-to-Self ($M = 895$ ms), $t(61) = 4.04$, $p < .001$. In no-perspective shift conditions, there was no significant difference between Self-to-Self trials ($M = 969$ ms) and Other-to-Other trials ($M = 1047$ ms), $t(61) = -1.78$, $p = .16$. Fig. 4 illustrates the different RTs (and error rates) in response to the probe question in each condition of the Self/Other Differentiation task.

This perspective-shift effect tended to be further modulated by contents type (expected vs. unexpected), resulting in a trend towards a three-way interaction between Perspective Shift, Contents Type, and Probe Type, $F(1, 61) = 2.94$, $p = .09$, *partial* $\eta^2 = .05$. Post-hoc analysis suggested that the interaction between perspective shift and probe type was only present for expected contents trials, $F(1, 61) = 6.75$, $p = .02$, *partial* $\eta^2 = .10$, but not for unexpected contents trials, $F(1, 61) = .46$, $p = .50$. No other effects reached significant levels (all p's > .05).

[1] Additional analysis was conducted on probe stage RTs in which trials where an error was made at the dilemma stage (and answered correctly at probe stage) were excluded from analysis. Results demonstrated the same pattern of findings: a significant main effect of Probe Type, $F(1, 61) = 18.18$, $p < .001$, *partial* $\eta^2 = .23$, and a significant interaction between Perspective Shift and Probe Type, $F(1, 61) = 4.90$, $p = .03$, *partial* $\eta^2 = .07$. Additionally, a significant interaction between Perspective Shift and Contents Type was found, $F(1, 61) = 4.78$, $p = .03$, *partial* $\eta^2 = .07$.

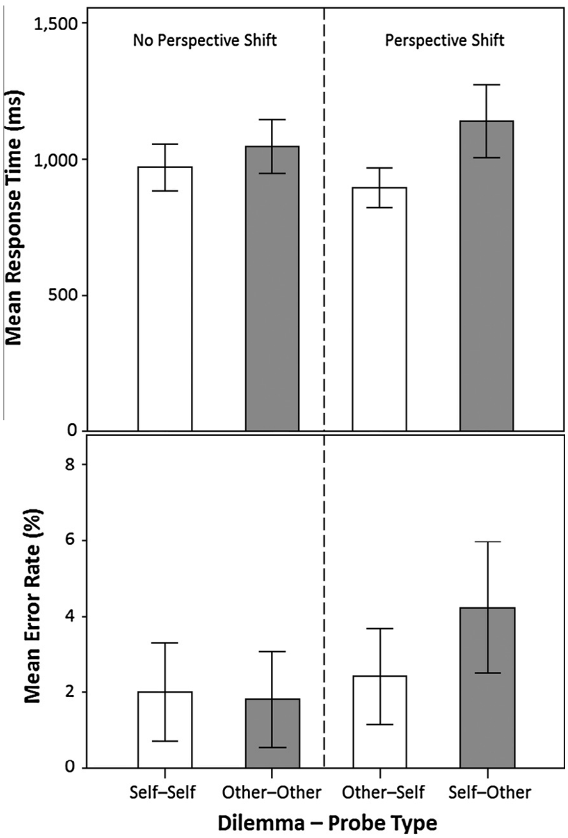

Fig. 4. The upper graph presents mean response time (in ms), and the lower graph presents accuracy data (percentage of errors) to the probe question in each test condition of the Self/Other Differentiation task. Error bars represent standard error of the mean.

3.1.2. Error rate data

3.1.2.1. Dilemma Stage. Analysis of error rates in response to the dilemma question revealed no significant difference between Self-Oriented Dilemmas ($M = 2.92\%$ errors) and Other-Oriented Dilemmas ($M = 2.37\%$ errors), $t(61) = 1.14$, $p = .26$.

3.1.2.2. Probe Stage. Analysis of error rates in response to the probe question revealed a significant main effect of contents type, $F(1, 61) = 18.08$, $p < .001$, *partial* $\eta^2 = .23$, with less errors in expected contents trials ($M = 1.61\%$ errors) than in unexpected contents trials ($M = 5.45\%$ errors). There was a significant three-way interaction between Perspective-Shift, Contents, and Probe Type, $F(1, 61) = 6.76$, $p = .01$, *partial* $\eta^2 = .10$. Post-hoc analysis suggested that this was because an interaction between Perspective-Shifting and Probe-Type was only present in unexpected contents conditions, $F(1, 61) = 5.41$, $p = .02$, *partial* $\eta^2 = .08$, but not for expected contents conditions, $F(1, 61) = 1.15$, $p = .29$. No other effects were significant (all p's > .05).

3.2. Stroop task

Only correct responses given within the set response window were included in RT analysis.

3.2.1. Response time data

A 2 (Task Type: Emotion vs. Gender) × 2 (Task Relevance of Distracter: Relevant vs. Irrelevant) × 2 (Distracter Congruence: Congruent vs. Incongruent) Repeated-Measures ANOVA was conducted on RT data.

Analyses of RT data revealed a main effect of task type, $F(1, 61) = 26.85$, $p < .001$, *partial* $\eta^2 = .31$, with slower responses in emotion-identification tasks ($M = 594$ ms) than in gender-identification tasks ($M = 547$ ms). A main effect of distracter congruency was also found, $F(1, 61) = 45.55$, $p < .001$, *partial* $\eta^2 = .43$, with faster responses in congruent trials ($M = 565$ ms) than incongruent trials ($M = 576$ ms).

There was no significant main effect of distracter relevance, $F(1,61) = 2.68$, $p = .11$. However, distracter relevance interacted with distracter congruence, $F(1,61) = 8.03$, $p = .006$, *partial* $\eta^2 = .12$, as the Stroop effect (incongruent minus congruent) was present only for task-relevant distracters (M = 17 ms), $t(61) = -6.63$, $p < .001$, but not for task-irrelevant distracters ($M = 5$ ms), $t(61) = -2.07$, $p = .22$. There was a significant interaction between Task Type and Distracter Congruence, $F(1,61) = 4.14$, $p = .046$, *partial* $\eta^2 = .06$, due to the presence of a larger Stroop effect in the emotion-identification task ($M = 23$ ms) than in the gender-identification task ($M = 12$ ms). No other interactions were significant. Fig. 5 illustrates these findings.

3.2.2. Error rate data

Analyses of error rates revealed a significant main effect of Task Type, $F(1, 61) = 13.18$, $p = .001$, *partial* $\eta^2 = .18$, with less errors in the gender-identification task ($M = 4.37\%$ errors) than in the emotion-identification task ($M = 5.78\%$ errors). There was also a significant main effect of Distracter Congruence, $F(1, 61) = 11.36$, $p = .001$, *partial* $\eta^2 = .16$, with less errors following congruent distracters ($M = 4.62\%$ errors) than incongruent distracters ($M = 5.53\%$ errors).

There was a significant interaction between Distracter Relevance and Distracter Congruence, $F(1, 61) = 14.07$, $p < .001$, *partial* $\eta^2 = .19$. Post-hoc analysis showed that this was because only task relevant distracter trials showed a significant difference in error rates between congruent ($M = 4.01\%$ errors) and incongruent ($M = 5.73\%$ errors) trials, $t(61) = -4.56$, $p < .001$. There was no significant difference in error rates between congruent ($M = 5.23\%$ errors) and incongruent ($M = 5.33\%$ errors) trials in task irrelevant distracter trials, $t(61) = -.325$, $p = .74$. No other effects were significant (all p's > .05; see Fig. 5).

3.3. Correlations: Self/Other Differentiation Task and Stroop task measures

For the purposes of correlational analysis, an Affective (Emotion-identification) Stroop score and a Non-Affective (Gender-Identification) Stroop score were calculated for each participant, from task relevant distracter conditions. This score was created by subtracting congruent trials RT from incongruent trials RTs.

Three scores were calculated for correlations in the Self/Other Differentiation Task, based on RT to both the

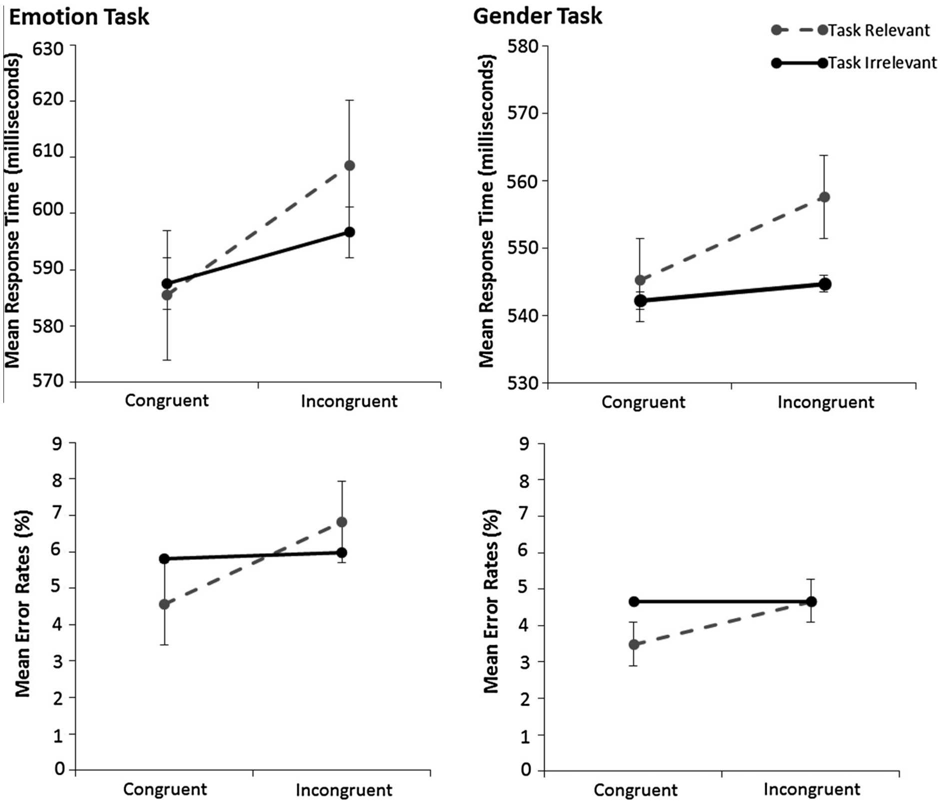

Fig. 5. Mean response times (top graphs) and error rates (lower graphs) by distracter relevance and congruence, presented separately for the emotion-identification task and the gender-identification task. Error bars represent standard error of the mean.

dilemma stage and the critical probe stage (i.e. belief attribution):

(1) Dilemma Score – the difference in RT between dilemma types: (Other minus Self).
(2) Probe Score – the difference in RT between probe types: (Other minus Self).
(3) Perspective-Shift score – the difference in RT between trials in which a perspective-shift was required and trials in which no perspective-shift was required:

$$(\text{Self-Other} - \text{Other-Self}) - (\text{Other-Other} - \text{Self-Self})$$

There was no significant relationship between the emotion-identification/gender-identification Stroop scores and either dilemma difference scores (p's > .05) or probe difference scores (p's > .05).

There was a significant positive relationship between perspective-shift scores and performance on the emotion-identification Stroop task, $r = .27$, $p = .03$ (2-tailed). There was no significant correlation between perspective-shift scores and performance on the gender-identification Stroop task, $r = .17$, $p = .20$ (2-tailed). This suggests that smaller perspective-shift scores (i.e. smaller difference in RT between perspective-shift types) were correlated with smaller affective Stroop scores; that is, higher levels of cognitive control in the affective Stroop trials correlated with faster 'perspective-switching' in the Self/Other Differentiation Task. Fig. 6 provides a scatterplot illustrating the significant correlation between perspective-shift scores and the affective Stroop task scores.

4. Discussion

In this study, trials were manipulated so that mental state attribution was required in regards to either one's own previously held beliefs, or to another person's currently held beliefs. In this context, 'mental state attribution' refers to the concept of holding two potentially contrasting belief states in one's own mind (i.e. possession, or awareness, of a false-belief state). When the contents of a container were revealed to be unexpected (e.g. marbles in a sugar jar), participants would be in possession of a true belief – knowledge that inside the jar are marbles – whilst also being aware that someone else, who hasn't seen inside the jar, would hold a false-belief that there is sugar inside. The results of this study support the idea of a measurable dissociation between self and other-person belief attribution processes at a behavioural level in a healthy adult population. More specifically, results showed that RTs were significantly longer when questions (probes) referred to other-people compared to when they referred to the self. In trials which did not require a perspective-shift between dilemma and probe stage, there was no

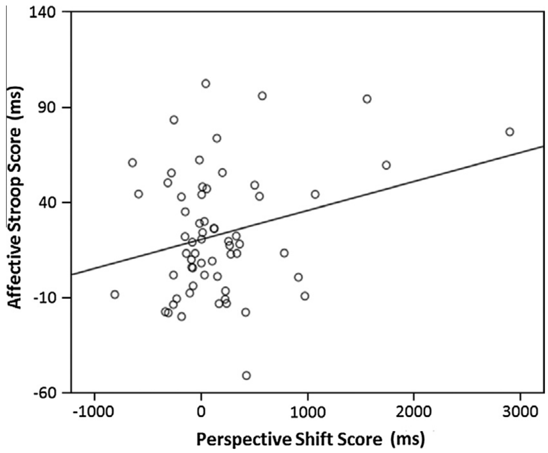

Fig. 6. Scatterplot showing correlations between perspective-shift scores from the Self/Other Differentiation task and the Emotion-Identification Stroop Task.

significant difference in response times between self or other probe conditions. This suggests that once the 'self' or 'other' perspective had been adopted at the dilemma stage, it was then maintained from this perspective throughout the trial, allowing quick 'other' or 'self' perspective assessments to be made by simply using the belief states computed at the dilemma stage.

Most interestingly, however, the requirement of a perspective-shift within a trial – and, further, the *type* of perspective-shift required – significantly affected response times. When participants were required to shift from their 'Self' perspective at dilemma stages to thinking of the 'Other' perspective at probe stages, reaction times were significantly longer than when the shift was from thinking of the 'Other' to thinking of the 'Self'. This finding suggests that in ToM, the self-perspective is always processed first, regardless of whether the ultimate goal is to refer to the self or to the other. More specifically, when one first processes a task from another person's perspective, before having to switch to the 'self' later, behavioural responses are as efficient as when the self perspective was processed from the beginning. This suggests that the Self-perspective has already been spontaneously 'processed' at the dilemma stage, even if the task itself required adoption of the 'Other' perspective. Therefore, the Self perspective was already available when required at the probe stage in the shift from Other-to-Self within a trial. In contrast, when a task began from the first person perspective at the dilemma stage (self-oriented), switching to the Other-person perspective at the probe stage, RTs were significantly slower and more error prone, suggesting that the 'Other' perspective had not yet been processed; the additional cognitive processing necessary to compute the perspective of the 'other' was only engaged in when explicitly prompted, at the probe stage.

These results imply both a significant overlap between cognitive mechanisms responsible for self and other-person belief attribution, and some distinct mechanisms involved in either self *or* other-person belief attribution (Jeannerod & Anquetil, 2008). More specifically, we propose that the 'self' perspective is always processed, regardless of the ultimate task demands (i.e. whether it is to orientate to self or other perspectives), whilst the 'other' perspective is only processed when explicitly required by the task. These results contradict Kovács et al.'s (2010) suggestion that the 'other' view is processed automatically, but support Apperly et al.'s (2006) proposal that other's beliefs are not tracked automatically, and are only considered when motivation to track another person's beliefs, whether directly or indirectly, is present; that is, our results support a distinction between 'awareness' of differing knowledge states and explicit adoption of another person's perspective (e.g. Schneider et al., 2014).

In our study, participants were explicitly asked to consider another agent's belief state which, ultimately, was matched to the participant's own belief state (i.e. what they believed to be in a container), and thus were required to adopt the perspective of the 'other', as opposed to simply being aware of the presence of a differing perspective or state. This differs from studies such as Samson et al. (2010) and Kovács et al. (2010), in which the presence of a secondary agent, irrelevant to the actual task, appeared to influence behaviours. In those prior studies, the 'other' view both differed from the 'self' view (e.g. the dots seen by the participant and avatar in Samson et al.'s task) and was not directly relevant to the task. Therefore, participants simply needed to demonstrate awareness of the presence of a differing state, which may have slowed response times, but does not necessarily reflect actual adoption and understanding of *what* the other person's mental state may be. The current task, in contrast, required both awareness of the existence of a different mental state and adoption of another agent's mental state.

One possible explanation for the longer response times in the other-oriented conditions may be a simple 'surprise' response resulting from the sudden appearance of the 'other', which is absent in the self-oriented conditions. However, it could be argued that this delay in response times does not merely reflect 'surprise', but is indicative of the additional cognitive effort required for the explicit differentiation between 'self' and 'other'. As previously stated, in this task there was no difference in answers to probes between the self and other perspective trials – both the knowledgeable self and naïve 'other' share the belief of what was in a specific container before seeing inside (e.g. sugar in a sugar jar), whether the contents was ultimately expected or unexpected. For instance, a participant may have been asked where they would look for some cookies, identifying a biscuit jar. They may then have been shown that there were either cookies (true belief) or elastic bands (false belief) inside. In both cases, before seeing inside, both the self and other-person believe there to be cookies in the jar, regardless of whether they then become more knowledgeable about the true contents (self) or not (other). Therefore, the key differentiation between self and other oriented trials was the perspective participants were prompted to respond from at probe stage – the answer being the same in both conditions (e.g. believing there to be sugar in the sugar jar), but from either the self or other perspective. It is further noted that trials in which a perspective shift occurred from Other-to-Self across a dilemma to probe were significantly faster, and more accurate, than Self-to-Other trials, suggesting it is not the presence and surprising nature of a perspective shift itself that elongates responses and increases errors, but rather the cognitive effort associated with the specific shift from the self-perspective to the other-perspective.

A perhaps intriguing result of this study was that contents – expected vs. unexpected – did not significantly influence response times, despite being a key feature of false belief tasks. However, analysis of error rates revealed that more errors were made by participants following the unexpected contents revelation (i.e. false-belief scenarios) compared to the expected contents conditions (true-belief scenarios). The belief-attribution task used in this study was designed for typically developed, healthy adults to complete and was not intended to be difficult; that is, each of the questions was easily answerable, and relatively simple to compute. Results suggest that, whilst the contents manipulation worked in creating false-belief scenarios, with participants making more mistakes when holding two contrasting beliefs at one time as demonstrated by the error rates (i.e. true-knowledge and false-belief about the contents of a container), once a solution had been reached (i.e. awareness and knowledge of true/false beliefs), behaviour was not significantly affected by the type of belief involved, but rather by the type of perspective shift required (self vs. other oriented).

The present study also adapted a face-word Stroop paradigm to explore the effect of affective vs. non-affective EF abilities, and how these may differentially relate to ToM capacities. Utilizing a within-subjects design, we found that participants were slower identifying emotions than gender in faces, supporting prior findings that affective stimuli are differentially processed compared to non-affective stimuli (e.g. LeDoux, 2000; Ohman & Mineka, 2001). We also found that the Stroop effect (longer response time for incongruent trials), for both the emotion expression (affective) and gender-identification (non-affective) tasks occurred only when distracters were task-relevant. This supports prior findings that suggest that distracters only influence task performance when relevant to the task at hand, as, otherwise, they do not create any cognitive conflict that needs to be resolved, provoking a significant difference in response times between congruent and incongruent trials (Reeck & Egner, 2011).

Correlational analysis suggested that there was a positive relationship between affective Stroop task scores and the 'perspective-shifting' abilities required for the Self/Other Differentiation Task, whilst non-affective Stroop task scores were not significantly correlated with perspective shifting performance. This supports the notion that there might be separate forms of EF that may be independently related to ToM expression. The positive correlation found between Affective Stroop scores and perspective-shifting suggests that participants who were faster at shifting between self and other perspectives were also better at ignoring irrelevant emotional distracters. This, however, was not due to a generalized superior EF skill to ignore distracters, because the non-affective Stroop scores did not significantly correlate with perspective shift performance. The affective Stroop task may be tapping into a specific aspect of EF that is relevant for mentalizing abilities (Hartwright et al., 2012). However, a note of caution must be added: whilst failing to reach significance, the non-affective Stroop task showed a correlation value with perspective-shifting abilities in the same direction as the affective-Stroop task, and we cannot rule out the possibility that with more power some significant correlation could be found. Additionally, examination of Fig. 6 reveals that there were some extreme values within our correlation analysis; although the results of all participants included were screened prior to analysis, exclusion of the most extreme score from the correlation (see far right in Fig. 6) would result in the disappearance of the significant correlation between affective Stroop scores and perspective-shifting capacities, which suggests that the relation between affective Stroop EF and perspective shifting abilities might have been driven by a few particular individuals. Further study is required to more firmly establish the extent to which these two EF abilities both diverge and overlap and how this may vary in the population.

Overall, our results support findings from recent neuroimaging studies using fMRI (Rothmayr et al., 2011; Van der Meer, Groenewold, Nolen, Pijnenborg, & Aleman, 2011) and PET (Happé et al., 1996) techniques, which have found differences in cognitive activation when task requirements (perspective shift/state type) are varied; that is, different forms of ToM (such as Self/Other or Affective/Cognitive) activate different neural mechanisms, seemingly drawing on separable, although likely interconnected, processes to successfully utilize ToM capacities. Our results support this notion that different types of mental state attribution recruit both common and distinct cognitive mechanisms, reinforcing the idea that there is not a single overarching 'mentalising mechanism' that accounts for all ToM processing, but rather a system of different processes activated depending on task demands (such as Self vs. Other ToM or Affective vs. Cognitive ToM). Our findings suggest that the ability to understand the self-perspective forms a prerequisite for understanding other people's perspectives. The 'self' perspective would be always processed, regardless of task demands, whilst the 'other' perspective would be processed only when explicitly necessary. The cognitive mechanisms associated with self-oriented belief processing are also activated when the task demands an other-person orientation, although additional and specific 'other-oriented' mechanisms are also recruited (Decety & Sommerville, 2003); in other words, self-understanding forms a core component of ToM belief-attribution abilities (Hobson, 2010; Happé et al., 1996; Frith & Frith, 2003).

As noted previously, Samson et al. (2010) report contradictory findings, suggesting that individuals automatically keep track of other people's knowledge states, even when not specifically required to do so (see also Keysar et al., 2000; Kovács et al., 2010). However, Qureshi et al. (2010) suggest that whilst we may automatically keep track of the *knowledge* states of others, it requires cognitive effort to deliberately take their *perspective*. This explicit adoption of the other's perspective is what our study demanded and what might require a previous computation of the Self perspective. Through the use of a matched design, allowing direct comparison of 'Self' and 'Other' oriented processing, our results also support the findings of Apperly et al. (2006) and Back and Apperly (2010), adding strength to the suggestion that other-oriented belief-attribution is not an automatic process.

5. Conclusion

This paper explored the extent to which Self-Oriented and Other-Oriented belief attribution abilities – two potentially separable components of the ToM mechanism – could be differentiated behaviourally in a healthy, adult population. Results revealed that when asked to consider another person's belief-state (e.g. '*What did Jane think was in the box?*'), participants took much longer to answer than when they were asked to reflect on their own prior belief state (e.g. '*What did you think was in the box?*'). This suggests that, even when initial beliefs states are matched (e.g. both the 'self' and the 'other' believed there to be sugar in a sugar bowl, before seeing inside), it takes longer to assess what the 'other' would believe to be in the container compared to what oneself believed to be in the container. Particularly interesting, the results of the current study also showed a specific effect of shifting perspectives (Self-to-Other/Other-to-Self) across a trial. Shifting from one's own perspective (*'Where would you look?*') to the perspective of another person ('*What would Jane think?*') is more difficult, taking significantly longer and resulting in more mistakes, than shifting from another person's perspective (*'Where would Jane look?*') back to one's own perspective ('*What did you think?*'). These findings suggest that the 'Self' perspective is always processed, regardless of whether ultimate task demands require a participant to orient to the 'Self' or 'Other' perspective, allowing the 'Self' perspective to simply be recalled in Other-to-Self trials. In contrast, the 'Other' perspective is only processed when explicitly necessary, thus requiring extra time to evaluate and attribute this belief-state in Self-to-Other trials.

Acknowledgements

This work was supported by the Economic and Social Research Council [grant number ES/J500136/1] in the form of a three-year PhD studentship awarded to Elisabeth Bradford.

References

Ahmed, F. S., & Miller, L. S. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*, 667–678.

Anderson, V., Levein, H. S., & Jacobs, R. (2002). Executive functions after frontal lobe injury: A developmental perspective. In D. T. Stuss & R. T. Knights (Eds.), *Principles of frontal lobe function* (pp. 1–7). New York, N.Y., U.S.: Oxford University Press.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from Adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology, 63*, 1201–1217.

Apperly, I. A., Riggs, K. J., Simpson, A., Samson, D., & Chiavarino, C. (2006). Is belief reasoning automatic? *Psychological Science, 17*(10), 841–844.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology, 45*, 190–201.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition, 115*, 54–70.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'Theory of Mind'? *Cognition, 21*, 37–46.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (Eds.). (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.. Oxford, UK: Oxford University Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the Eyes' test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241–251.

Bodden, M. E., Mollenhauer, B., Trenkwalder, C., Cabanel, N., Eggert, K. M., Unger, M. M., et al. (2010). Affective and cognitive theory of mind in patients with Parkinson's disease. *Parkinsonism and Related Disorders, 16*, 466–470.

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism, 8*, 283–299.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*, 475–482.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE, 5*, 1–13.

Bull, R., Philips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition, 107*, 663–672.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

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

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *TRENDS in Cognitive Sciences, 7*, 527–533.

Doherty, M. J. (2009). *Theory of mind: How children understand others' thoughts and feelings*. Hove, UK: Psychology Press.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologist's Press.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339.

Fine, C., Lumsden, J., & Blair, R. J. R. (2001). Dissociation between 'Theory of Mind' and executive functions in a patient with early left amygdala damage. *Brain, 124*, 287–298.

Frith, U. (1989). *Autism: Exploring the enigma*. Oxford, UK: Basil Blackwell.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London (Series B), 358*, 459–473.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'Theory of Mind': Evidence from compromised belief-desire reasoning in old age. *Cognition, 101*, 129–152.

Geusebroek, J., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam library of object images. *International Journal of Computer Vision, 61*, 103–112.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false-belief and the appearance-reality distinction. *Child Development, 59*, 26–37.

Hamilton, A. F. (2009). Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*, 881–892.

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). 'Theory of Mind' in the brain: Evidence from a PET scan study of Asperger syndrome. *Clinical Neuroscience and Neuropathology, 8*, 197–201.

Harari, H., Shamay-Tsoory, S. G., Ravid, M., & Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Research, 175*, 277–279.

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2012). Multiple roles for executive control in belief-desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning. *NeuroImage, 61*, 921–930.

Hillier, A., & Allinson, L. (2002). Beyond expectations: Autism, understanding embarrassment, and the relationship with theory of mind. *Autism, 6*, 299–314.

Hobson, R. P. (2010). Explaining autism: Ten reasons to focus on the developing self. *Autism, 14*, 391–407.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2011). Psychometric evaluation of the Theory of Mind Inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 42*, 327–341.

Jardri, R., Pins, D., Lafargue, G., Very, E., Ameller, A., Delmaire, C., et al. (2011). Increased overlap between the brain areas involved in self-other distinction in schizophrenia. *PLoS ONE, 6*, 1–9.

Jeannerod, M., & Anquetil, T. (2008). Putting oneself in the perspective of the other: A framework for self-other differentiation. *Social Neuroscience, 3*, 356–367.

Joseph, R. M., & Tager-Flusberg, H. (2004). The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology, 16*, 137–155.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32–38.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social self: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155–184.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., et al. (2010). Atypical neural self-representation in autism. *Brain, 133*, 611–624.

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition, 102*, 179–218.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483–522.

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology, 6*, 415–431.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*, 1081–1105.

Perner, J., Frith, U., Leslie, A. M., & Leekham, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60*, 689–700.

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

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

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117*, 230–236.

Reeck, C., & Egner, T. (2011). Affective privilege: Asymmetric interference by emotional distracters. *Frontiers in Psychology, 2*, 1–7.

Rothmayr, C., Sodian, B., Hajak, G., Dohnel, K., Meinhardt, J., & Sommer, M. (2011). Common and distinct neural networks for false-belief reasoning and inhibitory control. *NeuroImage, 56*, 1705–1713.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. H., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S preschoolers. *Psychological Science, 17*, 74–81.

Samson, D., Apperly, I. A., Braithwaite, J. J., & Andrews, B. J. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1255–1266.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain, 128*, 1102–1111.

Schneider, D., Slaughter, V. P., & Dux, P. E. (2014). What do we know about implicit false-belief tracking? *Psychonomic Bulletin & Review*, 1–12.

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology, 30*, 75–86.

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science, 16*(6), 311–315.

Van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: Neural basis of distinct processes underlying theory of mind. *NeuroImage, 56*, 2364–2374.

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

Williams, D., & Happé, F. (2010). Representing intentions in self and other: Studies of autism and typical development. *Developmental Science, 13*, 307–319.

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