Published in final edited form as:

Aphasiology. 2008 January 1; 22(1): 42–61. doi:10.1080/02687030600830999.

A different story on “Theory of Mind” deficit in adults with right hemisphere brain damage

Connie A. Tompkins,

University of Pittsburgh, Pittsburgh, PA, USA

Victoria L. Scharp,

University of Pittsburgh, Pittsburgh, PA, USA

Wiltrud Fassbinder,

University of Pittsburgh, Pittsburgh, PA, USA

Kimberly M. Meigh, and

University of Pittsburgh, Pittsburgh, PA, USA

Elizabeth M. Armstrong

Macquarie University, Sydney, Australia

Abstract

Background—Difficulties in social cognition and interaction can characterise adults with unilateral right hemisphere brain damage (RHD). Some pertinent evidence involves their apparently poor reasoning from a “Theory of Mind” perspective, which requires a capacity to attribute thoughts, beliefs, and intentions in order to understand other people’s behaviour. Theory of Mind is typically assessed with tasks that induce conflicting mental representations. Prior research with a commonly used text task reported that adults with RHD were less accurate in drawing causal inferences about mental states than at making non-mental-state causal inferences from control texts. However, the Theory of Mind and control texts differed in the number and nature of competing discourse entity representations. This stimulus discrepancy, together with the explicit measure of causal inferencing, likely put the adults with RHD at a disadvantage on the Theory of Mind texts.

Aims—This study revisited the question of Theory of Mind deficit in adults with RHD. The aforementioned Theory of Mind texts were used but new control texts were written to address stimulus discrepancies, and causal inferencing was assessed relatively implicitly. Adults with RHD were hypothesised not to display a Theory of Mind deficit under these conditions.

Methods & Procedures—The participants were 22 adults with unilateral RHD from cerebrovascular accident, and 38 adults without brain damage. Participants listened to spoken texts that targeted either mental-state or non-mental-state causal inferences. Each text was followed by spoken True/False probe sentences, to gauge target inference comprehension. Both accuracy and RT data were recorded. Data were analysed with mixed, two-way Analyses of Variance (Group by Text Type).

Outcomes & Results—There was a main effect of Text Type in both accuracy and RT analyses, with a performance advantage for the Theory of Mind/mental-state inference stimuli. The control group was faster at responding, and primed more for the target inferences, than the RHD group. The overall advantage for Theory of Mind texts was traceable to one highly conventional inference:

someone tells a white lie to be polite. Particularly poor performance in mental-state causal inferencing was not related to neglect or lesion site for the group with RHD.

Conclusions—With appropriate stimulus controls and a relatively implicit measure of causal inferencing, this study found no “Theory of Mind” deficit for adults with RHD. The utility of the “Theory of Mind” construct is questioned. A better understanding of the social communication difficulties of adults with RHD will enhance clinical management in the future.

Right hemisphere brain damage (RHD) in adults can engender an array of deficits in communication and social interaction. Among these are prominent difficulties in pragmatic functioning and discourse interpretation (see, e.g., Brownell, Griffin, Winner, Friedman, & Happé, 2000; Myers, 1999; Tompkins, Fassbinder, Lehman-Blake, & Baumgaertner, 2002). Some investigators posit that at least part of the essential RHD communicative profile reflects an impairment of social cognition (e.g., Brownell et al., 2000; Brownell & Martino, 1998; Sabbagh, 1999). “Theory of mind” (ToM) has become a popular investigative target in this regard.

Theory of Mind involves a capacity to attribute mental states in order to understand the behaviour of others (e.g., she wants to persuade, to be polite, to make a joke). It entails both a basic ability to interpret affective nonverbal signals, such as tone of voice and facial expression, and an apparently separate cognitive dimension, for reasoning about others’ thoughts, beliefs, and intentions (Sabbagh, 2004; Shamay-Tsoory, Tomer, Berger, Goldsher, & Aharon-Peretz, 2005). Potential deficits in perceiving nonverbal social and emotional cues are well substantiated in the RHD literature (see, e.g., Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000; Baum & Pell, 1999; Pell, 2006). Fewer studies have examined the cognitive aspect of ToM processing in adults with RHD, but each investigation has reported difficulties with ToM reasoning (e.g., Brownell, Pincus, Blum, Rehak, & Winner, 1997; Happé, Brownell, & Winner, 1999; Siegal, Carrington, & Radel, 1996; Winner, Brownell, Happé, Blum, & Pincus, 1998).

This paper revisits the question of “cognitive” ToM processing deficit in adults with RHD. Individuals with RHD are the focus in this paper, given the importance of the ToM construct in social cognition accounts of RHD, but it should be emphasised that various regions of both cerebral hemispheres contribute to the distributed neural circuitry that underlies ToM abilities (e.g., den Ouden, Frith, Frith, & Blakemore, 2005; Gallagher & Frith, 2003; Sabbagh, 2004; Shamay-Tsoory et al., 2005). This is not particularly surprising, because reasoning from a ToM perspective undoubtedly rests on a variety of component processes (e.g., Brownell et al., 2000).

Behavioural, cognitive, and neuroscience analyses of ToM processing are still in their infancy, and as a result, component analyses of ToM abilities are currently underdeveloped. Some rather general conceptual distinctions have been suggested, however. One contrasts representational aspects of ToM with executive selection (e.g., Apperly, Samson, Chiavarino, & Humphreys, 2004) and/or salience-marking processes (Brownell et al., 2000). Other proposed ToM distinctions include those between social-perceptual versus social-cognitive aspects (Tager-Flusberg & Sullivan, 2000), and between affective and cognitive facets, as noted above. Apart from theoretical challenges, the ToM research base is limited by methodological concerns. One of the most fundamental is that ToM distinctions and/or deficits are too often inferred from comparisons across questionably equated stimuli or tasks. This concern forms the main focus of the current study, and is elaborated below.

An analysis of ToM tasks from both the developmental and adult literatures indicates that performance typically hinges on the facility with which a comprehender resolves conflicting mental representations. For example, various versions of the classic false-belief task (e.g., Wimmer & Perner, 1983) introduce a change in a location (an item is moved from one place

to another while a character is out of the room) or of an item in a location (a box is filled with an unexpected kind of candy). The detection of irony (Shamay-Tsoory et al., 2005) and deception (e.g., Brune & Bodenstein, 2005; Stuss, Gallup, & Alexander, 2001) similarly requires an awareness of a mismatch between a proposition or action and some other explicit, observable context or circumstance. Identifying faux pas (Shamay-Tsoory et al., 2005; Stone, Baron-Cohen, & Knight, 1998) involves an appreciation of how an act or remark violates some behavioural norm. And narrative tasks, whether they use pictured (Brune & Bodenstein, 2005) or linguistic input (e.g., Happé et al., 1999; Winner et al., 1998), incorporate multiple characters and character perspectives in addition to that adopted by the comprehender. Indeed, because such tasks require participants to override salient expectations or perspectives, a variety of investigators have examined the extent to which they rely on inhibitory control (e.g., Leslie, Friedman, & German, 2004; Russell, 2005; Sabbagh, 2006).

For adults with RHD, ToM ability often is assessed with texts that are built around explicitly competing propositions. One prominent example comes from Happé and colleagues (1999), whose stimuli have been used to investigate ToM processing in a broad range of populations (Fletcher et al., 1995; Happé, 1994; Happé, Malhi, & Checkley, 2001; Snowden et al., 2003; Tchanturia, Happé, Godley, Treasure, Bara-Carril, & Schmidt, 2004). These ToM stimuli feature a character whose motives or beliefs explain a textual conflict (e.g., Peter hates his aunt's new hat but tells her he likes it; while walking alone at night, a woman is approached by a man who wants to ask the time but she gives him her purse). The study participant is asked "Why" the character did what s/he did. The correct answer requires an inference about context-general social conventions (e.g., Peter is being polite) or context-specific and at times emotionally laden scenarios (the woman does not want to be accosted). Happé and colleagues reported that adults with RHD were disproportionately impaired on these "mental causal inference" stimuli, in relation to control stimuli designed to require a "non-mental causal inference" (e.g., a burglar carefully avoids the detector beam that would set off a burglar alarm. Then he steps on something soft and furry that screeches and runs away, and the burglar alarm goes off. Target inference: the burglar disturbed a cat, which then went through the detector beam).

Inadequate stimulus control rather than mental inference requirements may account for RHD deficits on this type of task. For example, Happé and colleagues' (1999) ToM stimuli, but not their control stimuli, contain an explicit contradiction. It is well established that RHD adults' language comprehension problems are particularly evident when the stimulus material induces competing interpretations (e.g., Tompkins et al., 2002). Furthermore, seven of eight ToM stimuli include at least two characters and shift multiple times between their perspectives, while six of eight control texts have only one character. Thus there are important differences between the experimental and control stimuli in the number and nature of competing discourse entity representations and in the processing needed to integrate them.

The current study was designed to revisit the attribution of mental-state causal inferencing deficits to adults with RHD (hereafter, "mental inference"). Happé and colleagues' (1999) ToM stories were used, but new control texts were developed to address the aforementioned stimulus discrepancies. These control texts incorporated an explicit propositional conflict that could be resolved with a non-mental-state causal inference (hereafter, "non-mental inference"). Numbers of characters, character mentions, and perspective shifts were controlled across ToM and non-mental texts. Balancing the stimulus sets across these dimensions was expected to eliminate the differential comprehension demands of the original stimuli—demands that most likely put Happé and colleagues' RHD adults at a particular disadvantage on the ToM stories. We hypothesised that when requirements were equated for the representation and processing of textual contradictions, adults with RHD would not demonstrate the previously reported discrepancy between ToM and non-mental inferencing abilities. We also evaluated causal

inferencing in a relatively implicit manner, in order to reduce potential confounds to interpretation that accompany highly explicit methods, such as the original open-ended question measure (Tompkins & Baumgaertner, 1998; Tompkins & Lehman, 1998).

METHOD

Participants

A total of 60 adults participated in this study. Of these, 22 had unilateral RHD due to cerebrovascular accident, and 38 had no known brain damage or neurological impairment. Table 1 provides biographical and clinical information for each participant group. The groups did not differ in demographic characteristics, but the group with RHD was reliably poorer on the clinical/neuropsychological measures, as would be expected.

Participants with RHD were recruited from eight acute care hospitals and rehabilitation facilities. All potential participants with RHD met the following criteria: unilateral hemispheric lesion confirmed by CT or MRI scan report; at least 4 months post-onset of CVA; age between 40 and 85 years; and a minimum of 8 years of formal education. Exclusions were based on medically documented evidence of bilateral lesions, brainstem or cerebellar damage, premorbid seizure disorders, head injuries requiring hospitalisation, problems with drugs and/or alcohol, a potentially cognitively deteriorating condition such as Alzheimer's or Parkinson's disease, or psychiatric illness. Clinical CT or MRI reports showed a range of within-hemisphere lesion sites (see Table 1).

Potential participants were interviewed and tested for several other selection criteria. By self-report they were premorbidly right-handed, monolingual American English speakers (see Tompkins, Bloise, Timko, & Baumgaertner, 1994, for operationalisations). They also passed a pure-tone air conduction hearing screening (35dB HL at 500, 1000, and 2000 Hz; values within 0.5 standard deviation of Harford and Dodds' (1982) means for ambulatory, non-institutionalised older men). Those who passed the hearing screening in only one ear were given a repetition task, consisting of 12 words loaded with fricative consonants. More than one repetition error was grounds for exclusion from the study.

Individuals in the non-brain-damaged (NBD) control group were recruited from the laboratory's research registry, and also through media advertisements, senior citizen groups, hospital volunteer departments, a university website on clinical research, and control participants' referrals of family members or friends. These individuals met the same biographical and behavioural inclusion criteria as participants with RHD. Before being enrolled in the study, they were interviewed to rule out previous neurological episodes or conditions or problems with drugs and/or alcohol. In addition, as a cognitive screen, all control group participants passed the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975) with at least 28 out of 30 possible points.

Overview

Participants heard spoken narrative stimuli that targeted either a mental or a non-mental causal inference. Each experimental text was followed by two spoken sentence probes. Target inference comprehension was gauged relatively implicitly with a speeded verification task, in which participants indicated as quickly as possible whether the sentence probe was True or False in relation to its preceding narrative. The Appendix provides sample narratives with their sentence verification probes and the full stimulus set is available at <http://www.pitt.edu/~neurogen>

Stimuli

Except as otherwise noted, the stimulus validation results reported below reflect data from a minimum of 10 socio-demographically appropriate respondents per validation task. More detail on stimulus validation is available on request.

Happé and colleagues' (1999) "mental causal inference" (ToM) stimuli express a conflict that can be resolved by drawing an inference about a character's motives or beliefs (e.g., Mrs Robinson, while walking at night, is approached by a man who wants to ask her the time. Mrs Robinson fearfully hands him her purse). Two texts each induce a single, highly probable inference about the reason for this conflict, through (a) double bluff situations, where a character with a clear reason to deceive actually tells the truth; (b) attempts to persuade by invoking false consequences; (c) white lies for socially appropriate ends; and (d) mistakes, in which a second character's innocent motives are misinterpreted by the protagonist (e.g., Mrs Robinson). One of the original eight stimuli was excluded from this study because pilot subjects performed on average less than 70% correct when this text was presented in auditory form, rather than for unlimited reading time in visual form, as in the Happé et al. study. This "double bluff" text was also more abstract than the others (involving what a prisoner will divulge about the position of his army's tanks, when captured by the enemy). The ToM stories were modified slightly to replace British English terms with American English; to change some character names to those more common in American English or to avoid unwanted associations (e.g., "Mrs Robinson"); and to avoid gender pronoun confusion. The stimuli are quite heterogeneous linguistically, and that was not changed for this study.

Seven "physical causal inference" (PHYS) stories were developed to include an explicit contradiction (e.g., Bill leaves home to get Chinese takeout for dinner. Bill comes home with pizza for dinner). Other information in the stimulus text (e.g., the restaurant was boarded up) made this conflict resolvable via a single, highly probable inference that, as per Trabasso and Suh's (1993) criteria, had a physical and objectively judged basis (e.g., the Chinese restaurant had gone out of business). There were three general themes for the PHYS text inferences: (a) something is broken/not working (e.g., Bob, carrying a heavy load, is climbing eight flights of stairs to his office because the elevator is out of order); (b) an unexpected event occurs (e.g., Bill and the Chinese restaurant); and (c) something is lost or misplaced (e.g., Jean puts her keys in her purse but shortly after the keys are not there, because they fell out in the car).

The PHYS texts were designed and validated to be comparable, on average, with the ToM stimuli along a number of dimensions (see Table 2). These include syntactic complexity, discourse structure, range of verb usage (Armstrong, 2001), several measures of length, cohesion (connectedness of the text in terms of repetition of information, semantically related information, and use of pronominal reference; Armstrong, 1991), and conjunction analysis (Halliday & Hasan, 1976). The ToM and PHYS texts were also equated for the number of intervening clauses between the contradictory propositions, and were balanced for number of characters, character mentions, and shifts of reference between characters. Pilot testing confirmed that the causal inferences that resolved the textual conflicts were similarly highly probable for ToM and PHYS texts. Additional piloting was done to rate degree of conflict for the two contradictory propositions extracted from each narrative. These ratings were similar for ToM and PHYS texts, and much higher than for filler texts that contained no overt contradiction. As expected, these same stimulus excerpts differed substantially in ratings of how easy it was to make a "mental picture" of the explanation for the conflict. PHYS text inferences were judged much more imageable than ToM text inferences.

To assess causal inferencing, each stimulus text was paired with three types of sentence verification probes, one experimental, one comparison, and two filler. *Experimental probes* for the TOM texts were derived from the correct answers to the question "Why did this

happen?” from a multiple-choice version of Happé and colleagues’ (1999) task that was developed for adults with aphasia. Experimental probes for the PHYS texts were developed from pilot subjects’ highly probable responses to the same “Why” question. *Comparison probes* were developed to be unrelated to the texts, so by definition, the correct probe verification response would be “False”. This necessitated a “False” response for the experimental probes as well. Thus the experimental sentence probes were written and validated to assert the opposite of the target causal inference (e.g., “Mrs Jones thinks the man wants to buy her purse”; “Bill found that the Chinese restaurant was in business”). Comparison probes were validated to be unrelated to the preceding texts (e.g., “Mrs Jones thinks the man wants to fix her purse”; “Bill found that the Chinese restaurant was in fashion”). Experimental and comparison probes differed in only a key lexical item or two-word phrase, which were comparable across text types in terms of critical properties (see Table 2). *Filler probes* (e.g., “Mrs Jones thinks the man wants to steal her purse”) were used to balance the overall number of True and False responses in the study.

Both the ToM- and PHYS-text probe sentences consist of three major elements: Subject-Verb-Object or Subject-Verb-Adverbial. All probe sentences include some degree of embedding. The ToM probes involve embedded clauses in the object position, using reduced and unreduced that-clauses typically associated with mental verbs. The PHYS probes include such embedding as well, but also involve relative clauses that postmodify nouns in Subject and Adverbial elements of the clause (e.g., “The money that Sam needed was in his coat”).

Each text was presented twice, followed by two probe sentences. For ToM and PHYS texts, the first sentence was either its experimental or its comparison probe. The second probe was always a filler.

A total of 18 filler texts were also written, to disguise crucial stimulus and task characteristics. All filler texts differed in story theme from the ToM and PHYS texts (e.g., buying a house; getting a haircut). Eight of the filler texts contained no propositional conflict. To minimise participants’ expectation for verbatim repetition of texts and probe sentences, 10 of the 18 filler texts were written in alternate versions, in which character names, settings, and other details were changed. All filler texts were presented twice in the study, either verbatim or in their alternate versions. The first probe for each filler text trial was predominantly “True”, to balance the number of times that the correct answer to the initial sentence probe was True or False. The two sentence probes for each filler text were arranged to balance the patterns of responses for each two-probe combination (e.g., True–True; True–False, etc.) throughout the study.

As developed from the Happé et al. multiple choice items, correct performance was possible on three of the ToM sentence probes without considering a character’s mental state (i.e., Peter was not rude, regardless of how he wanted to act). Two design features were included to induce participants to attend to the entire probe sentence, particularly the early portion that contained the mental state verb. First, 44% of the filler probes contained an internal incongruity (e.g., “Peter did not know how to be polite”). For these items, ignoring the first part of the stimulus would yield an error. Second, a sentence recognition task was used after 50% of the second filler probes. In this task, participants heard the same filler probe sentence (on 40% of trials) or one that differed by a single word. Their task was to indicate if the sentence was exactly the same as the one just before it. On 75% of the “different” items, the difference occurred in the first third of the sentence. For 80% of those items the difference was in the main verb position.

Task construction

A practised female speaker audio-recorded the narrative stimuli, and a practised male speaker recorded the sentence verification probes, all at an average speaking rate of about four syllables/second. Stimuli were produced without undue emphasis on any lexical element. All recordings

were made with an Audio-Technica ATR20 vocal/instrument microphone with a constant microphone-to-mouth distance (~4 inches). Recording was done in a double-walled, sound-treated booth. Stimuli were recorded onto a Dell Optiplex GX260 with a Creative SB Live! Value (WDM) sound card using Sound Forge v4.5 software at a sampling rate of 22.05 KHz with 16-bit resolution. The first author and several assistants collaborated with the speakers to achieve recording consistency.

Stimuli were assembled using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). A single trial consisted of a trial number, a 500-ms silent interval; the narrative; a 500-ms silent interval; the first sentence verification probe; a 4000-ms silent interval; and the second sentence probe. On trials with a sentence recognition probe, that probe occurred 4000 ms after the offset of the preceding probe.

The 64 trials for this task were pseudorandomly arranged into eight blocks of eight trials each, according to the following criteria: none of the texts or probes was repeated in any single session; the number of experimental stories was balanced across sessions and trial blocks; each block of trials began with a filler text; no more than two experimental texts occurred in a row; and no more than three texts with an initial True or False probe occurred in sequence.

Experimental apparatus and procedures

All stimuli were delivered via a Dell Inspiron 5150 notebook computer, through high quality supraural earphones (Beyerdynamic DT150) at a comfortable loudness level selected by the participant via Quick Mixer v1.7.2. Participants responded to each sentence probe by pressing one of two labelled buttons (True/False) on a manual response box. E-Prime software generated and stored accuracy data and millisecond RTs.

Testing required four to five sessions, lasting up to 90 minutes, over a period of about 3–7 weeks. Order of stimulus block presentation within and across sessions was counterbalanced for each participant. Stimulus blocks were interspersed with screening and clinical measures, and with experimental tasks for other studies, to maximally separate presentations of repeated stimuli.

Participants were tested in a quiet room, either in their homes or in the first author's laboratory. Five examiners were trained to perform the testing, and only one examiner worked with each participant. Each examiner tested people from both groups.

Participants received extensive orientation, instruction, and practice in performing the sentence probe verification task. They were also trained to respond with a single finger of the right hand, and to return that finger between trials to a designated location equidistant from the response buttons. Spoken and gestural reminders about speed and consistency were provided throughout the training. Training and practice continued until participants responded with assurance and RTs had stabilised. In each new session, participants received additional practice items before presentation of the first block of stimuli. Spoken and gestural reminders about response speed and consistency were provided prior to each block of trials.

To encourage rapid responding, a response deadline (standard Windows bell) was presented on 60% of filler sentence probes. Response deadlines were individually set, and were based on the average RT for correct responses during practice trials.

RESULTS

Preliminary data analysis

Independent *t*-tests were computed separately for each group, to evaluate accuracy and RT performance for gender differences. Dependent *t*-tests were also computed within group, separately for each text type, to examine for order effects of experimental vs comparison probe presentation. None of these analyses was significant (all $p > .08$).

Primary data analysis

This study addressed presumed deficits in mental causal inferencing by adults with RHD. Existing ToM texts were equated across critical dimensions with new control texts that induced a non-mental, physical causal inference (PHYS), and the probe verification task provided a relatively implicit measure of causal inferencing. Table 3 provides group accuracy and RT data.

One participant with RHD was excluded from all analyses because she did not understand the task. During prolonged demonstration and practice she performed erratically, making many errors, responding before the probe sentence was presented, or not responding at all. This participant was the oldest in the RHD group (by 2 years) and made the most errors on the auditory working memory measure (Tompkins et al., 1994); otherwise her performance on the clinical measures was unremarkable. Her particular difficulties cannot be attributed clearly to a deficit in auditory attentional control, however. One other participant with RHD was equally poor on the working memory task but well within the range of RHD group performance on the dependent measures.

Accuracy data were analysed with two-way ANOVA (Group by Text Type [ToM, PHYS]). In the analysis for experimental probes, only the main effect of Text Type was significant, $F(1, 57) = 38.53, p < .05$, with performance more accurate on the ToM/mental causal inference texts. For comparison probes, both main effects were significant, Group $F(1, 57) = 12.93, p < .05$; Text Type $F(1, 57) = 7.95, p < .05$, but these effects were qualified by a significant Group by Story Type interaction, $F(1, 57) = 5.10, p < .05$. For the group with RHD, sentence probe responses were more accurate for ToM stimuli than for PHYS stimuli, while NBD participants' accuracy did not differ across text types. This interaction most likely reflects a ceiling effect for the NBD group. To adjust for the different accuracy patterns on experimental and comparison probes, experimental probe data were submitted to two-way Analysis of Covariance, with comparison probe performance the covariate. Only the main effect for Text Type was significant, $F(1, 55) = 6.77, p < .05$. Again, performance was better for ToM than for PHYS text inferences.

RT data were analysed only for accurate trials. Four additional participants with RHD and four from the NBD group were excluded from RT analyses because they achieved less than 66% accuracy for one Text Type (the PHYS condition, in each case). The excluded individuals were no different in any obvious way from the participants whose data remained in the RT analyses. Outlying RT values also were excluded from RT analyses (RT $> \pm 3$ SD from each individual respondent's condition mean). For the RHD group, this involved 4.5% of ToM and 7.6% of PHYS trials. For the NBD group, 6.3% of ToM and 7.4% of PHYS trials were excluded.

Using mean RT as the dependent measure, two-way ANOVA indicated significant main effects of Group and Text Type, for both types of sentence probes—experimental: Group $F(1, 46) = 17.43, p < .05$; Text Type $F(1, 46) = 9.09, p < .05$; comparison: Group $F(1, 46) = 11.03, p < .05$; Text Type $F(1, 46) = 31.98, p < .05$. As would be expected, participants in the NBD group responded more quickly than those in the RHD group. And once again, a performance advantage was evident for probes associated with the ToM stories.

RT priming was also evaluated, using a proportion measure calculated for each trial (experimental RT/comparison RT) to adjust for inter-individual differences in basic response speed. The smaller the proportion, the greater the RT advantage for experimental probes, which represents the extent of “priming” of the mental and non-mental causal inferences. A two-way ANOVA indicated only a significant Group effect, $F(1, 46) = 4.84$; $p < .05$. The NBD group exhibited more RT priming overall than did the RHD group.

Assessment of variations in performance

Text Type differences—Performance differences in causal inferencing were inspected within Text Type, to determine whether the results of the main analyses could be attributed to any particular stimulus item. Performance on each ToM stimulus was highly accurate, within and between groups (range 95–100%). Accuracy on the PHYS trials was much more variable (57–95% for RHD; 68–97% for NBD), but with several exceptions, the rank order of difficulty was similar for both groups. The most difficult stimulus for both groups involved a man who threw his wallet into his coat pocket on the way to the store, but whose pocket was empty when he got to the checkout counter. The sentence probe for this stimulus contained an embedded relative clause (“The money that Sam needed was in his coat/hand”), but so did another PHYS text probe that was performed substantially better. When this “most difficult” stimulus was excluded from the main analyses, the Text Type effect was still significant. Moreover, the text itself did not stand out on any of the textual variables listed in Table 2.

Turning to RT priming for ToM stimuli, the RHD group exhibited the most RT priming for one of the social convention inferences (Peter, who lied to be polite; RT proportion = 0.76), and this was the second most primed item for the NBD group (RT proportion = 0.86). The second “most primed” item for adults with RHD had a proportion score of 0.92 (1.10 for the NBD group); this text involved an intent to persuade someone to do something that was against the rules. However, the RHD group exhibited the least priming for the other “white lie” and “persuasion” texts. For PHYS stimuli, both groups primed the least on the least accurate stimulus, described above, but otherwise there was little group overlap in the rank order of RT priming for PHYS trials. Again, removing this most difficult stimulus from the RT priming analysis did not change the original result.

When the easiest ToM stimulus was removed from the original analyses (Peter and his white lie), all prior main effects for Text Type vanished. Thus, this one item was responsible for what otherwise appeared to be a performance advantage for ToM processing.

Individual performance differences—Causal inferencing variation between participants was evaluated first by correlating performance on the dependent measures with the continuous demographic and clinical variables listed in Table 1, and also with pure tone averages as an index of hearing sensitivity. Results were assessed against Cohen’s (1988) rule of thumb for a large effect size in the behavioural sciences ($r > 0.5$). For the RHD group, there were no correlations that even came close to this criterion. For the NBD group, greater vocabulary knowledge, as assessed by the *Peabody Picture Vocabulary Test – Revised* (Dunn & Dunn, 1991), was associated with more accurate responses to experimental probes in both stimulus sets—ToM; $r(38) = .54$; PHYS; $r(38) = .69$. Greater RT priming for the PHYS experimental probes was also associated with better vocabulary, $r(32) = -.52$, and with lower scores on immediate and delayed story retell (Bayles & Tomoeda, 1993); $r(38) = .51$ and $.60$, respectively.

A bimodal clinical description variable, presence or absence of neglect in the RHD group, was assessed next for its potential relationship to performance variation. Participants who fell below the cut-off for neglect on the *Behavioural Inattention Test* (Wilson, Cockburn, & Halligan, 1987) performed well within 1 SD of the RHD group means on all dependent measures, indicating that neglect was not a factor for our sample.

Differences in mental causal inferencing performance were also assessed by within-hemisphere lesion site for the RHD group. Because sample sizes were small in each lesion subgroup, these results were not analysed statistically. Mean performance accuracy was highly similar across lesion subgroups, ranging from 93% (“cortical anterior”) to 100% (“cortical posterior”, “cortical mixed”, “cortical + subcortical”). This pattern of results was the same whether or not participants with lacunar infarcts were included. Mean RT priming was also similar across most subgroups, with ToM advantages ranging from 5% to 13% in most cases. By contrast, a substantial advantage for PHYS text priming was evident for two subgroups that consisted of a single individual each (“cortical mixed” and “cortical + subcortical”).

Finally, data were inspected to identify members of the RHD group who were particularly deficient at mental causal inferencing. The criterion for this designation was performance at or less than the lower 2 SD bound of the NBD group mean on ToM texts. There were no such poor performers on the accuracy measure, and only two who were particularly low in RT priming. One of these was the single “cortical mixed” individual just noted above. An MRI scan was available for this participant, taken contemporaneously with the behavioural testing. The Automated Labelling Pathway procedure for lesion localisation (Wu, Carmichael, Carter, Figurski, Lopez-Garcia, & Aizenstein, in press) identified damage involving Brodmann’s areas 12, 25, and 47 along with small portions of 10 and 11. Clinical description of the lesion in the second case indicated a “lucency in the right middle cerebral artery distribution compatible with infarct”. Both of these individuals fell at or better than the RHD group mean on the demographic and clinical variables in Table 1.

Sentence recognition task performance

A sentence recognition task was performed after 50% of the filler trials in this study, to induce participants to attend to all portions of the probes in the sentence verification task. Not surprisingly, given the divided attention demands of this task, performance of the participants with RHD, $M(SD) = 12.67(3.0)$, was significantly worse than that of the NBD group, $M(SD) = 15.4(2.9)$; $t(57) = -3.39$; $p < .05$.

DISCUSSION

This study evaluated causal inferencing about mental states by adults with RHD, using Happé and colleagues’ (1999) ToM stimuli and new control stimuli designed to address potentially important discrepancies in Happé et al.’s original ToM and control texts. When assessed implicitly with mental and physical causal inference stimuli balanced for numbers of characters, character mentions, and character perspective shifts, adults with RHD evidenced no ToM deficit. This is contrary to the results reported by Happé and colleagues.

Based on another aspect of the original Happé et al. (1999) study, one might still want to attribute ToM deficit to adults with RHD. Happé and colleagues also reported RHD deficits in interpreting cartoon stimuli when the humour depended on mental states (false beliefs or ignorance) in relation to non-mental-state cartoons. However, again these results might reflect factors other than ToM impairment. As illustrated in both examples in the Happé et al. article, there are potential differences in the visual clarity and configuration of the two types of cartoons that might put adults with RHD at a disadvantage on the ToM stimuli. In one example, the essential aspect of the ToM cartoon is partially obscured (by both the rungs of a staircase and part of a doorframe) and is backgrounded by both of these objects from the focal point of the scene, whereas the key figure of the non-mental-state cartoon is visually unobscured and in the centre of the scene. In the other example, interpretation of the ToM cartoon depends on detecting subtle indications of gaze direction in an otherwise dark and detailed background, while the physical anomaly of the non-mental-state cartoon is depicted by clear, bold lines against a white background. Happé et al. reported only that the ToM and non-mental-state

cartoon sets contained the same number of stimuli that show facial expressions. With these concerns, we regard as premature the conclusion that adults with RHD have a modality-independent deficit in cognitive TOM processing.

In the current investigation, what appeared to be a consistent performance disadvantage for the PHYS stimulus set is actually attributable to one easy ToM item, and text type differences disappeared when analyses were redone without this item. The mental causal inference for this story was highly conventional, involving a child who told a “white lie” in order to be polite. Context-general, world-knowledge mental inferences of this sort are likely to be more resistant to (right hemisphere) brain damage than are context-variable mental inferences (e.g., Brownell & Martino, 1998), which were targeted in most of the rest of the ToM stimuli. The sentence probe for this “white lie” item also assessed a first-order ToM inference (A thinks/believes/intends X), rather than a second-order inference (A thinks/believes/intends that B thinks/believes/intends X), and second-order mental inferences may be more vulnerable to RHD (e.g., Winner et al., 1998; but see also Siegal et al., 1996). The other “white lie” story in the ToM stimulus set did not stand out as particularly easy and it also probed a first-order inference, so conventionality and low inference complexity are not unequivocally associated with better ToM inferencing.

Happé and colleagues (1999) assessed accuracy of causal inferencing by asking participants explicitly why something happened, whereas this study’s sentence verification task provides a more implicit measure of causal inferencing. These differences in assessment methods cannot account for the discrepant findings on RHD adults’ relative ease of mental versus non-mental causal inferencing, because any potential advantage conferred by the more implicit method would have affected performance on both text types. However, the more implicit measure in this study was expected to improve RHD group accuracy overall (e.g., Tompkins & Baumgaertner, 1998; Tompkins, Boada, & McGarry, 1992; Tompkins & Lehman, 1998). Indeed, there was no accuracy performance gap between groups.

An index of “reading time” bolstered Happé and colleagues’ (1999) conclusion about the link between RHD and ToM deficit. Participants in their study were given unlimited time to read each text silently, before indicating that they were ready for the probe question. Their participants with RHD took longer with the ToM texts than did the NBD group, but there was no group difference in reading times for the control stories. Disregarding the difficulty of interpreting such a measure as indicative of a processing impairment, the RHD difference for ToM versus control texts could reflect the discrepancy between text types in inducing competing interpretations. As discussed previously, this factor can create particular performance difficulties for adults with RHD. Such is the case for the RT measures in this study, in which both ToM and PHYS texts were built around explicit contradictions. The group with RHD performed less well than our NBD group on the raw RT measure of causal inferencing, as well as the RT priming measure that took into account individual differences in basic response speed.

The most abstract of the original ToM texts was excluded from this study, but this does not appear to account for the cross-study differences in outcome. The individual participant data provided in Happé et al. (1999) suggest that more than one ToM story created difficulty for many in their RHD group. In addition, Happé and colleagues made no mention of any particularly difficult or easy stimulus. Returning to this investigation, there was also no effect of the potential confound that a number of the ToM sentence probes could be answered without considering mental state. Both accuracy and RT performance were consistent across ToM sentence probes, suggesting that our task manipulations were effective in controlling for this potential problem.

It is possible that this study's sample of participants with RHD had less substantial brain damage than the group in Happé et al. (1999). Individual participant data provided in Happé et al. indicate that 8 of 14 in the RHD group were particularly poor performers on the ToM task, as gauged by this study's criterion of scores at least 2 SD below the NBD group mean. This contrasts with 2 of 21 RHD participants in the current study, for the RT priming measure. However, this comparison is not strictly valid, due to the methodological differences in assessing mental causal inferencing. As noted above, the less explicit measure in this study most likely boosted RHD accuracy overall. Despite the lack of group accuracy differences, the sample of adults with RHD in this study was not simply "normal". They were significantly impaired on the clinical description measures listed in Table 1, and they performed more poorly than the NBD group on the sentence recognition task. The relative lack of such ancillary clinical data for Happé and colleagues' group with RHD makes it difficult to evaluate potential sample severity differences between the two studies. Happé et al. do indicate that none of their participants with RHD had neglect, which indicates that their reported ToM deficit was not attributable to neglect. This is consistent with the current study's finding that the presence of neglect did not necessitate poor experimental performance.

Turning to potential neural substrates of ToM deficit, one of the particularly poor performers with RHD in this study, in terms of RT priming for mental causal inferences, had a lesion that encroached on ventromedial/orbitofrontal brain regions (Brodmann's areas 12, 25, 47, and small portions of 10, 11). These regions have been reported by some to be necessary for ToM performance (e.g., Stuss et al., 2001), but interpreted by others as more important for assessing general stimulus coherence than ToM inferencing (Ferstl & von Cramon, 2002). We cannot add to this debate, for several reasons. This study's other particularly poor performer had a middle cerebral artery lesion, which more than likely did not affect these brain regions. Also, the lesion for one of Happé and colleagues' participants involved both the middle cerebral artery and the anterior communicating artery, and as such, could have affected the basal forebrain. However, this participant was far from the most impaired on the TOM task.

Although lesion–performance correlates are as yet unclear, another cross-study lesion difference should be considered in trying to account for the discrepancy in results between the current investigation and that of Happé et al. (1999). It appears that Happé et al. did not include participants with purely thalamic/basal ganglia lesions (although that cannot be determined unequivocally from a description of MCA lesion), while this study did. However, the subcortical subgroup's ToM task performance in this study was very similar to that of the subgroups whose lesions were characterised as right cortical anterior, cortical posterior, cortical + subcortical, or simply middle cerebral artery. Of course nothing confident can be concluded from such gross indicators. But from the combined observations above, one might speculate that right hemisphere lesion site per se may not be a crucial predictor of mental inferencing ability on story tasks like that of Happé and colleagues. If so that would not be particularly surprising, in light of the potential for differential performance on basic components of such tasks, and of the distributed neural circuitry that underlies ToM reasoning (e.g., den Ouden et al., 2005; Gallagher & Frith, 2003; Shamay-Tsoory et al., 2005).

Happé et al.'s (1999) RHD participant data indicate a wide range of individual differences in mental causal inferencing based on the ToM texts. One participant with RHD had a perfect score on the ToM measure, while two performed with only about 30% accuracy. Individual differences of this sort are common when investigators select participants by presence of a hemispheric lesion and not by a clinical diagnosis or behavioural characteristic. We believe that such heterogeneity can be illuminating if it can be interpreted with reference to methodological or performance correlates. The five best RHD performers in Happé and colleagues' data had the highest levels of education. This could perhaps be aligned with our finding that NBD adults with better vocabularies were more accurate at causal inferencing. But

there were no education or vocabulary effects for our participants with RHD, and education did not clearly distinguish the three worst performers in Happé et al.'s study from those with lesser ToM-inferencing impairment. It would be interesting to know whether the exceptionally good or poor ToM performances in Happé and colleagues' RHD data could be linked to any other individual difference variables.

In the current study, individual differences in performance were also evident, and were evaluated extensively. None of the demographic or clinical measures was associated with performance differences among the participants with RHD. As just noted, vocabulary knowledge was implicated in the NBD group, with a particularly strong association for the physical causal inference texts. It is possible that the PHYS texts included more difficult vocabulary, or a larger range of vocabulary sophistication, than the ToM stimuli. Immediate and delayed text recall also predicted ToM, but not PHYS, performance for the NBD control group. It is not clear why textual memory skills should be linked with one story type and not the other. Unfortunately, for the group with RHD, this study provides no bases from which to draw even tentative hypotheses about the behavioural correlates of mental causal inferencing. The two individuals who were particularly poor on the RT priming measure of ToM did not have neglect, and were well within the range of scores on the other clinical neuropsychological subject description measures.

This study is, of course, limited by its focus on only one type of ToM task. A variety of story tasks have been used to assess ToM (e.g., Siegel et al., 1996; Stone et al., 1998; Zaitchik, Koff, Brownell, Winner, & Albert, 2006), along with cartoon tasks (e.g., Happé et al., 1999) and other measures of perspective taking and deception (e.g., Stuss et al., 2001; Zaitchik et al., 2006). Recently, Apperly, Samson, and Humphreys (2005) evaluated ToM assessments and suggested the need to develop new measures that target potential component mechanisms and processes of the construct. Future conceptualisation and investigation of ToM processing should also take into account potential modulators of ToM task performance. For example, Rutherford (2004) found that even temporary, experimentally induced social status differences influenced reasoning on TOM tasks, without affecting performance on a control task.

Another issue for future consideration concerns the ecological validity of the TOM construct, and its centrality in theories of social cognition. Keysar, Lin, and Barr (2003) have reported that adults with well-developed ToM abilities do not routinely rely on them to interpret what other people do. Harris, Todorov, and Fiske (2005) herald the value of attribution theory to expand the conceptual underpinnings of, and evidence base on, dispositional inferencing (see also Brownell & Martino, 1998).

In sum, with appropriate stimulus controls in place and a relatively implicit measure of performance, this study found that adults with RHD did not evidence an impairment in mental-state causal inferencing on a commonly used textual measure of ToM. The results of this study, along with those of Happé and colleagues (1999), suggest that variation in such mental inference performance may not be unambiguously linked to site of lesion within the right hemisphere. It is possible that the sort of mental causal inferencing deficit reported originally by Happé and colleagues for adults with RHD would be evident in participants with more extensive brain damage, and/or on other assessments of ToM. It is also possible, however, that the focus on ToM per se may not be very useful for conceptualising the communicative and social interaction deficits of adults with RHD: a construct at once too broad in relation to the component processes involved, and too narrow in relation to more overarching theories of social cognition and attribution.

Acknowledgments

This project was supported in part by grant # DC01820 from the National Institute on Deafness and Other Communicative Disorders. The authors are indebted to Denise Balason, Meghan Capellini, Bethany Peters, Anita Lewis, Sara Byers, and Annie Palaika for their assistance.

References

- Adolphs R, Damasio H, Tranel D, Cooper G, Damasio AR. A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience* 2000;20(7):2683–2690. [PubMed: 10729349]
- Apperly IA, Samson D, Chiavarino C, Humphreys GW. Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience* 2004;16(10):1773–1784. [PubMed: 15701227]
- Apperly IA, Samson D, Humphreys GW. Domain-specificity and theory of mind: Evaluating neuropsychological evidence. *Trends in Cognitive Sciences* 2005;9(12):572–577. [PubMed: 16271505]
- Armstrong EM. The potential of cohesion analysis in the analysis and treatment of aphasic discourse. *Clinical Linguistics and Phonetics* 1991;5(1):39–51.
- Armstrong EM. Connecting lexical patterns of verb usage with discourse meanings in aphasia. *Aphasiology* 2001;15(10–11):1029–1045.
- Baum SR, Pell MD. The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology* 1999;13(8):581–608.
- Bayler, KA.; Tomoeda, CK. *Arizona Battery for Communication Disorders of Dementia*. Tucson, AZ: Canyonlands Publishing; 1993.
- Benton, AL.; Hamsher, KdeS; Varney, NR.; Spreen, O. *Visual Form Discrimination*. New York, NY: Oxford University Press; 1983.
- Benton, AL.; Sivan, AB.; Hamsher, KdeS; Varney, NR.; Spreen, O. *Judgement of Line Orientation*. New York, NY: Oxford University Press; 1983.
- Brownell, H.; Griffin, R.; Winner, E.; Friedman, O.; Happé, F. Cerebral lateralization and theory of mind. In: Baron-Cohen, S.; Tager-Flusberg, H.; Cohen, D., editors. *Understanding other minds: Perspectives from developmental cognitive neuroscience*. Vol. 2. Oxford, UK: Oxford University Press; 2000. p. 306–333.
- Brownell, H.; Martino, G. Deficits in inference and social cognition: The effects of right hemisphere brain damage on discourse. In: Beeman, M.; Chiarello, C., editors. *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 1998. p. 309–328.
- Brownell H, Pincus D, Blum A, Rehak A, Winner E. The effects of right-hemisphere brain damage on patient's use of terms of personal reference. *Brain and Language* 1997;57:60–79. [PubMed: 9126407]
- Brune M, Bodenstein. Proverb comprehension reconsidered—“Theory of mind” and the pragmatic use of language in schizophrenia. *Schizophrenia Research* 2005;75:233–239. [PubMed: 15885515]
- Cohen, J. *Statistical power analysis for the behavioural sciences*. Vol. 2. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1988.
- den Ouden HEM, Frith U, Frith C, Blakemore SJ. Thinking about intentions. *NeuroImage* 2005;28(4):787–796. [PubMed: 15964210]
- Dunn, L.; Dunn, L. *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: American Guidance Service; 1991.
- Forstl EC, von Cramon DY. What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *NeuroImage* 2002;17(3):1599–1612. [PubMed: 12414298]
- Fletcher PC, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, et al. Other minds in the brain: A functional imaging study of “theory of mind” in story comprehension. *Cognition* 1995;57:109–128. [PubMed: 8556839]

- Folstein MR, Folstein SE, McHugh PR. Mini-Mental State: A practical guide for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;12:189–198. [PubMed: 1202204]
- Gallagher HL, Frith CD. Functional imaging of “theory of mind”. *Trends in Cognitive Sciences* 2003;7(2):77–83. [PubMed: 12584026]
- Halliday, MAK.; Hasan, R. *Cohesion in English*. London: Longman; 1976.
- Happé FGE. An advanced test of theory of mind: Understanding of story characters’ thoughts and feelings by able autistic, mentally handicapped and normal children and adults. *Journal of Autism and Developmental Disorders* 1994;24:129–154. [PubMed: 8040158]
- Happé FGE, Brownell HH, Winner E. Acquired “theory of mind” impairments following stroke. *Cognition* 1999;70:211–240. [PubMed: 10384736]
- Happé FGE, Malhi GS, Checkley S. Acquired mind-blindness following frontal lobe surgery? A single case study of impaired “theory of mind” in a patient treated with stereotactic anterior capsulotomy. *Neuropsychologia* 2001;39:83–90. [PubMed: 11115657]
- Harford ER, Dodds E. Hearing status of ambulatory senior citizens. *Ear & Hearing* 1982;3(3):105–109. [PubMed: 7095317]
- Harris LT, Todorov A, Fiske ST. Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. *NeuroImage* 2005;28(4):763–769. [PubMed: 16046148]
- Keysar B, Lin S, Barr DJ. Limits on theory of mind use in adults. *Cognition* 2003;89(1):25–41. [PubMed: 12893123]
- Leslie AM, Friedman O, German TP. Core mechanisms in “theory of mind”. *Trends in Cognitive Sciences* 2004;8(12):528–533. [PubMed: 15556021]
- Myers, P. *Right hemisphere damage: Disorders of communication and cognition*. San Diego, CA: Singular Publishing Group, Inc; 1999.
- Pell MD. Cerebral mechanisms for understanding emotional prosody in speech. *Brain and Language* 2006;96(2):221–234. [PubMed: 15913754]
- Russell J. Justifying all the fuss about false belief. *Trends in Cognitive Sciences* 2005;9(7):307–308. [PubMed: 15951225]
- Rutherford MD. The effect of social role on theory of mind reasoning. *British Journal of Psychology* 2004;95(1):91–103. [PubMed: 15005870]
- Sabbagh MA. Communicative intentions and language: Evidence from right-hemisphere damage and autism. *Brain and Language* 1999;70:29–69. [PubMed: 10534371]
- Sabbagh MA. Understanding orbitofrontal contributions of theory-of-mind reasoning: Implications for autism. *Brain and Cognition* 2004;55:209–219. [PubMed: 15134854]
- Sabbagh MA. The development of executive functioning and theory of mind a comparison of Chinese and U.S. preschoolers. *Psychological Science* 2006;17(1):74–81. [PubMed: 16371147]
- Schneider, W.; Eschman, A.; Zuccolotto, A. *E-prime user’s Guide*. Pittsburgh, PA: Psychology Software Tools Inc; 2002.
- Shamay-Tsoory SG, Tomer R, Berger BD, Goldsher D, Aharon-Peretz J. Impaired “affective theory of mind” is associated with right ventromedial prefrontal damage. *Cognitive and Behavioural Neurology* 2005;18(1):55–67.
- Siegal M, Carrington J, Radel M. Theory of mind and pragmatic understanding following right hemisphere damage. *Brain and Language* 1996;53(1):40–50. [PubMed: 8722898]
- Snowden JS, Gibbons ZC, Blackshaw A, Doubleday E, Thompson J, Craufurd D, et al. Social cognition in frontotemporal dementia and Huntington’s disease. *Neuropsychologia* 2003;41(6):688–701. [PubMed: 12591026]
- Stone VE, Baron-Cohen S, Knight RT. Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience* 1998;10(5):640–656. [PubMed: 9802997]
- Stuss DT, Gallup GG, Alexander MP. The frontal lobes are necessary for “theory of mind. *Brain* 2001;124:279–286. [PubMed: 11157555]
- Tager-Flusberg H, Sullivan K. A componential view of theory of mind: Evidence from Williams syndrome. *Cognition* 2000;76(1):59–90. [PubMed: 10822043]

- Tchanturia K, Happé FGE, Godley J, Treasure J, Bara-Carril N, Schmidt U. "Theory of mind" in anorexia nervosa. *European Eating Disorders Review* 2004;12:361–366.
- Tompkins CA, Baumgaertner A. Clinical value of online measures for adults with right hemisphere brain damage. *American Journal of Speech-Language Pathology* 1998;7:68–74.
- Tompkins CA, Bloise CGR, Timko ML, Baumgaertner A. Working memory and inference revision in brain-damaged and normally aging adults. *Journal of Speech and Hearing Research* 1994;37:896–912. [PubMed: 7967574]
- Tompkins CA, Boada R, McGarry K. The access and processing of familiar idioms by brain-damaged and normally aging adults. *Journal of Speech and Hearing Research* 1992;35:626–637. [PubMed: 1608254]
- Tompkins, CA.; Fassbinder, W.; Lehman-Blake, MT.; Baumgaertner, A. The nature and implications of right hemisphere language disorders: Issues in search of answers. In: Hillis, A., editor. *Handbook of adult language disorders: Integrating cognitive neuropsychology, neurology, and rehabilitation*. New York: Psychology Press; 2002. p. 429-448.
- Tompkins CA, Lehman MT. Interpreting intended meanings after right hemisphere brain damage: An analysis of evidence, potential accounts, and clinical implications. *Topics in Stroke Rehabilitation* 1998;5:29–47.
- Trabasso T, Suh S. Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes* 1993;16:3–34.
- Wilson, B.; Cockburn, J.; Halligan, P. *Behavioural Inattention Test*. Windsor, UK: Thames Valley Test Company; 1987.
- Wimmer H, Perner J. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 1983;13:103–128. [PubMed: 6681741]
- Winner E, Brownell HH, Happé F, Blum A, Pincus D. Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language* 1998;62:89–106. [PubMed: 9570881]
- Wu M, Carmichael O, Carter CS, Figurski JL, Lopez-Garcia P, Aizenstein HJ. Quantitative comparison of neuroimage registration by AIR, SPM, and a fully deformable model. *Human Brain Mapping*. (in press).
- Zaitchik D, Koff E, Brownell H, Winner E, Albert M. Inference of beliefs and emotions in patients with Alzheimer's disease. *Neuropsychology* 2006;20(1):11–20. [PubMed: 16460218]

APPENDIX

Sample narratives with sentence probes

"Mental inference" (ToM) stimulus (as modified from Happé et al., 1999):

Tom, a burglar who has just robbed a shop, is making his getaway. As he is running home, Mary Davis, a police officer on her beat, sees him drop his glove. She doesn't know that Tom is a burglar, she just wants to tell him he dropped his glove. But when she shouts out to the burglar, "Hey, you! Stop!", he turns around, sees the officer, and gives himself up. He puts his hands up and admits that he robbed the local shop.

Experimental probe: Tom thinks the officer knows he is innocent.

Comparison probe: Tom thinks the officer knows he is studious.

"Physical inference" (PHYS) stimulus:

Bob shares an office with his co-worker Sandy. The office is on the eighth floor of a building. Sandy always walks up the eight steep flights of stairs to their office, and never sees anyone else in the stairwell. Bob refuses to walk and always takes the elevator because he carries many files every day and thinks that eight flights is just too far to walk. One morning, Sandy saw many other employees in the stairwell, including Bob climbing the stairs with his heavy briefcase.

Experimental Probe: Bob saw that the elevator was working.

Comparison Probe: Bob saw that the elevator was leaving.

TABLE 1

Demographic and clinical characteristics of two participant groups

Characteristics	RHD (<i>n</i> =22)	NBD (<i>n</i> =38)
<i>Age (years)</i>		
Mean (<i>SD</i>)	64.4 (10.3)	60.4 (9.5)
Range	42–79	45–84
<i>Gender</i>		
Male	13	19
Female	9	19
<i>Education (years)</i>		
Mean (<i>SD</i>)	14.6 (3.2)	13.9 (2.2)
Range	9–22	12–20
<i># per lesion site (CT/MRI report)</i>		Not applicable
Right cortical anterior	3	
Right cortical posterior	1	
Right cortical mixed	2	
Right subcortical	8	
Right cortical + subcortical	2	
Right MCA	6	
<i>Lesion type (CT/MRI report)</i>		Not applicable
Thromboembolic	11	
Lacunar	2	
Haemorrhagic	9	
<i>Months post-onset</i>		Not applicable
Mean (<i>SD</i>)	65.7 (52.2)	
Range	4–167	
<i>PPVT-R^a</i>		
Mean (<i>SD</i>)	157.1 (11.1)	163 (11.1)
Range	132–173	115–174
<i>Auditory Working Memory for Language^b</i>		
<i>*Word recall errors</i>		
Mean (<i>SD</i>)	12.4 (6.3)	5.0 (4.6)
Range	2–24	0–16
<i>True/false errors</i>		
Mean (<i>SD</i>)	1.0 (0.02)	1.0 (0.02)
Range	0.9–1	0.9–1
<i>*Behavioural Inattention Test^c</i>		
Mean (<i>SD</i>)	138 (14.5)	144 (2.8)
Range	85–146	133–146
<i>*Visual Form Discrimination^d</i>		
Mean (<i>SD</i>)	28.4 (3.5)	30.3 (2.2)
Range	20–32	24–32

Characteristics	RHD (<i>n</i> =22)	NBD (<i>n</i> =38)
<i>*Judgement of Line Orientation^e</i>		
Mean (<i>SD</i>)	23 (4.8)	27.1 (4.2)
Range	11–30	16–33
<i>ABCD^f Story Retell</i>		
<i>*Immediate</i>		
Mean (<i>SD</i>)	12.5 (2.6)	14.4 (2.1)
Range	7–17	9–17
<i>Delayed</i>		
Mean (<i>SD</i>)	12.5 (3.1)	13.6 (2.5)
Range	5–17	7–17

RHD =right hemisphere brain damage; NBD =non-brain-damaged; MCA =middle cerebral artery; PPVT-R =Peabody Picture Vocabulary Test – Revised; ABCD =Arizona Battery for Communication Disorders of Dementia.

^aDunn & Dunn (1991; maximum =175).

^bTompkins et al. (1994; maximum errors = 42).

^cWilson et al. (1987; maximum =146; neglect cut-off =129).

^dBenton, Hamsher, Varney, & Spreen (1983; maximum =32).

^eBenton, Sivan, Hamsher, Varney, & Spreen (1983; age- and gender-corrected score; maximum =35).

^fBayles & Tomoeda (1993; maximum =17).

* significant group difference by independent *t*-test (*p* <.05).

TABLE 2

Stimulus characterisation data (*M*, range)

	Mental inference (ToM) stimuli	Physical inference (PHYS) stimuli
<i>Stimulus texts</i>		
Length (clauses)	15.1 (10–19)	15.7 (11–21)
Length (words)	91.0 (73–108)	86.7 (66–111)
# T-units	6.1 (4–8)	6.6 (5–10)
Clauses/T-unit	2.5 (2.1–2.7)	2.4 (1.8–3)
<i>Text structure components</i>		
Initiating event	1.0 (1–1)	1.0 (1–1)
Complicating event	1.0 (1–1)	1.0 (1–1)
Resolution	0.0 (0–0)	0.3 (0–1)
<i>Verb range^a</i>		
# Material	10.3 (7–12)	7.6 (3–10)
# Relational	2.1 (0–6)	3.6 (1–7)
# Mental	2.3 (0–5)	3.6 (1–6)
# Verbal	1.4 (0–4)	2.1 (1–3)
# Behavioural	0.1 (0–1)	0.1 (0–1)
% Cohesion ^b	70.4 (68–75)	81.4 (75–85)
# Intervening clauses	4.9 (0–7)	4.9 (1–8)
<i># Character references</i>		
Character 1	10.7 (6–19)	10.6 (6–18)
Character 2	6.0 (1–10)	5.6 (2–7)
Shifts	6.9 (2–12)	6.7 (4–11)
<i>Sentence Probes</i>		
Length (clauses)	2.4 (2–3)	2.1 (2–3)
Length (syllables)	10.5 (7–11)	11.2 (9–13)
<i>Key lexical contrasts^c</i>		
<i>Phoneme count</i>		
Experimental	4.3 (2–7)	4.2 (3–6)
Comparison	4.5 (2–7)	4.3 (3–6)
<i>Syllable count</i>		
Experimental	1.7 (1–3)	1.6 (1–3)
Comparison	1.7 (1–3)	1.6 (1–3)
<i>Mean RT (ms)</i>		
Experimental	613 (569–661)	609 (527–644)
Comparison	609 (563–792)	625 (558–704)

^aArmstrong (2001).^bArmstrong (1991).^cFrom Washington University's English Lexicon Project website (www.ellexiconwustl.edu).

TABLE 3Group accuracy and response time (RT) data for experimental and comparison sentence probes (*M, SD*)

Group	Mental inference (ToM) stimuli	Physical inference (PHYS) stimuli
<i>Accuracy (Maximum =7)</i>		
<i>RHD</i>		
Experimental	6.81 (0.51)	5.67 (1.27)
Comparison	6.48 (0.75)	6.00 (1.05)
<i>NBD</i>		
Experimental	6.89 (0.31)	6.21 (1.02)
Comparison	6.81 (0.46)	6.76 (0.49)
<i>RT (ms)</i>		
<i>RHD</i>		
Experimental	1091 (317)	1226 (317)
Comparison	1084 (292)	1274 (442)
<i>NBD</i>		
Experimental	756 (264)	836 (335)
Comparison	804 (269)	804 (279)
<i>RT priming^a</i>		
<i>RHD</i>	1.02 (0.25)	1.03 (0.30)
<i>NBD</i>	0.96 (0.20)	0.86 (0.22)

^aRT priming =RT experimental/RT comparison; smaller number =greater priming for the experimental (causal inference) sentence probe. This proportion measure is calculated for each trial and then averaged, and as such cannot be generated from the mean RT data, above.