Right Brain Damage and Inference Revision Revisited

Cynthia G. R. Bloise and Connie A. Tompkins

In theoretical and experimental attempts to characterize the communicative consequences of right hemisphere brain damage (RHD), inferencing ability has received much attention (Brownell, Potter, Bihrle, & Gardner, 1986; Joanette & Goulet, 1987; Myers, 1991; Tompkins & Mateer, 1985). For instance, Brownell et al. (1986) examined RHD adults' ability to draw inferences required to form a linguistic bridge between apparently discordant propositions. From another perspective, Tompkins and Mateer (1985) assessed RHD subjects' facility with affective inferences, implied by a character's attitude in a given situation. Both of these studies indicated that RHD adults are capable of drawing inferences from single sentences or from internally consistent discourse units. However, RHD patients have more difficulty when a revision of their initial interpretation is required for comprehension.

Because these two studies have associated RHD with inference revision deficits, we planned to examine the correlates of these deficits. However, our confidence in the original findings was reduced after a thorough review uncovered several questions of internal and external validity, which are described in the Method section. Given our concerns, we extensively modified both studies to investigate more carefully the association between RHD and inference revision deficits, intending to replicate the original findings before pursuing more information about the cognitive and behavioral correlates of these deficits. The nature and results of our replication efforts are described here.

METHOD

Subjects

Thirty-two adults participated in this study; 16 had unilateral RHD as a result of cerebrovascular accident (CVA), and 16 were control subjects without known neurologic impairment. Subjects met stringent medical and behavioral inclusion criteria concerning radiologic documentation of unilateral lesion, auditory and visual processing ability adequate to perform experimental tasks, premorbid right handedness, and yes/no reliability. Descriptive data are summarized in Table 1. The groups did not differ in socio-demographic characteristics.

Stimuli and Tasks

Subjects performed two auditory tasks. The *Bridging Inferences task* (BI), using stimuli adapted from Brownell et al. (1986), included sentence pairs presented as single units. The first sentence set up one inference, but the second sentence implied a different interpretation; thus, an inference revision was necessary to interpret the pair's meaning. After each sentence pair, subjects answered yes/no questions about facts, initial inferences, and revisions of initial inferences (see sample sentence pairs and questions in Table 2).

The *Attitudinal Inferences task* (AI), using stimuli from Tompkins and Mateer (1985), consisted of stories that focused on a main character's attitude. Half were congruent in mood throughout, but half included an incongruent final comment from the character (a semantically positive comment following negative depiction of events) that required an inference revision for successful resolution (e.g., "the character was not positive, he was sarcastic"). Subjects answered pairs of yes/no questions about facts and inferences after these stories, similar to the procedure used in the *Boston Diagnostic Aphasia Examination* (Goodglass & Kaplan, 1983) Complex Ideational Materials subtest (see sample stories and questions in Table 2).

As indicated earlier, we had a number of concerns about the validity of the original studies from which these tasks were modified. These concerns and the changes that we made to address them are indicated in Table 3. The major questions that we had about the Brownell et al. (1986) study centered on stimulus validation procedures, subject inclusion and exclusion criteria, subject description, stimulus presentation mode, and sample size. For the Tompkins and Mateer (1985) study, we also questioned the stimulus presentation mode and sample size. We were also concerned

TABLE 1. DEMOGRAPHIC AND CLINICAL CHARACTERISTICS (MEANS AND RANGES, OR FREQUENCY COUNTS) OF TWO SUBJECT GROUPS

	RHD ($N = 16$)	*Control* ($N = 16$)
Age	64.3 (51–78)	65.6 (51–77)
Education (years)	12.1 (8–17)	12.6 (8–17)
Gender	9 Male 7 Female	8 Male 8 Female
Estimated IQ	106 (93–116)	107 (94–119)
Months Post-Onset of CVA	28.5 (11–40)	
Etiology of CVA	9 Thromboembolic 5 Hemorrhagic 2 Lacunar	
Auditory Comprehension (Overall BDAE %ile)	95 (86–98)	97 (94–98)
Phrase-Length Ratio	4.5 (.9–10)	4.6 (.9–12)
Percent Literal Concepts ("Cookie Theft")	64 (44–81)	59 (50–73)
*Tonal Memory	6.3 (2–10)	8.1 (1–10)
*Judgment of Line Orientation	17.3 (5–27)	23.3 (16–30)
Visual Neglect		
Draw from memory	3 Yes	
Medical records data	9 Yes	
Hemiparesis	14 Yes	

Note: RHD = Right-hemisphere-damaged.
* significantly different, $p < .05$.

about the scoring criterion and the generalizability of results from the young seizure patient population that was studied.

Procedures

All stimuli (sentences, stories, and questions) were extensively pretested and validated with socio-demographically appropriate judges. Additional validation of the original BI stimuli was conducted for two purposes. First,

TABLE 2. SAMPLE EXPERIMENTAL STIMULI FOR TWO INFERENCING TASKS

Bridging Inferences task (Brownell et al., 1986)

Experimental Sentence Pair:

Barbara became too bored to finish the history book.
She had already spent five years writing it.

Experimental Questions:

Factual

1. Did Barbara grow tired of watching movies?
2. Was Barbara tired of the book?

Inferential

1. Did Barbara grow tired of reading the book? (initial inference)
2. Did Barbara grow tired of writing the book? (inference revision)

Attitudinal Inferences task (Tompkins & Mateer, 1985)

Congruent Story:

Nan invited her new neighbor, Mark, to a party. He told *hilarious* stories and *everyone* enjoyed listening. Her husband said to her, "Good decision. He's really fun to have around."

Incongruent Story:

Nan invited her new neighbor, Mark, to a party. He told *boring* stories and *no one* enjoyed listening. Her husband said to her, "Good decision. He's really fun to have around."

Questions for each version: (One point scored only when each member of a question pair is answered correctly).

Factual

Pair 1. Did Nan invite her new [neighbor/boss]?
Pair 2. Did the guests [like/dislike] Mark's stories?

Inferential

Pair 1. Did Nan's party go [well/poorly]?
Pair 2. Might Nan [invite/refuse to invite] Mark again?

to verify that an inference revision was required, we determined subjects' most likely interpretations of each initial misleading sentence. Second, to determine how likely it would be for normally aging adults to draw the less preferred interpretation of an experimental sentence pair, we established the plausibility of the second (revised) interpretation. Details of these validation procedures are provided in the Appendix. Stories and questions for the AI task were read to the validating judges as well. They indicated that the materials were clear and unambiguous. Story type and question type for the AI task were reliably determined in the original investigation.

TABLE 3. VALIDITY CONCERNS ABOUT TWO PREVIOUS INFERENCING STUDIES AND MODIFICATIONS MADE TO ADDRESS THEM IN THIS STUDY

Bridging Inferences task (Brownell et al., 1986)

Validity Concern	*Modification*
—Inadequate stimulus validation	Extensive validation with socio-demographically appropriate judges; 9 of 15 original stimuli needed to be modified and 3 were discarded
—Incomplete/unreported subject inclusion and exclusion criteria	Strict criteria for variables that might affect performance (e.g., hearing)
—Incomplete subject descriptions	Extensive biographical, medical, and behavioral description
—Live-voice presentation of experimental materials, with possible emphatic inconsistency	Audiotaped presentation of materials, for consistency
—Small patient sample ($N=8$)	Larger patient sample ($N=16$)

Attitudinal Inferences task (Tompkins & Mateer, 1985)

Validity Concern	*Modification*
—Live-voice presentation of experimental questions, with possible emphatic inconsistency	Audiotaped presentation of materials, for consistency
—Liberal scoring criterion, may have inflated chance accuracy	Conservative scoring criterion, requiring correct answers to paired questions
—Young, seizure patient sample and their matched controls	Older, CVA patient sample with matched controls
—Small patient sample ($N=8$)	Larger patient sample ($N=16$)

For the experimental tasks, all stimuli were presented via audiotape for consistency of production. Experimental sentence or story order and question order were conditionally randomized, so that no more than three experimental items of the same type (e.g., BI experimental vs. filler items; AI congruent vs. incongruent stories) occurred in sequence. Filler items that did not require inference revision were interspersed with experimental stimuli to minimize the formation of response strategies. Testing was completed in a quiet room, with descriptive and experimental tasks given in a predetermined order. Subjects responded to the questions following each stimulus by pressing buttons labelled *yes* or *no*.

RESULTS

Preliminary analyses indicated that there were no gender differences in accuracy, and the order experimental trial block presentation had no effect. Accordingly these factors were not considered further in the major analyses, which are outlined for each task below.

Bridging Inferences Task

Accuracy data were submitted to two different two-way ANOVAs (Group by Question Type), with repeated measures on Question Type (24 factual vs. 24 inferential questions for one analysis; 12 initial inference vs. 12 inference revision questions for the second analysis). For the analysis of factual vs. inferential questions, only the Condition main effect was significant (F (1,30) = 118.4; $p < .001$). All subjects were more accurate at answering factual (RHD M = 22.1, SD = 1.7; Control M = 22.6; SD = 1.4) than inferential questions (RHD M = 16.9, SD = 2.1; Control M = 18.1, SD = 2.8).

For the second analysis, the 24 inference questions were split into 12 examining initial inferences and 12 regarding inference revisions. Means and standard deviations for initial inference questions were 8.50 (1.2) and 8.75 (1.6) for RHD and control subjects, respectively; for inference revision questions the means and standard deviations were 8.44 (1.8) and 9.38 (1.5). No significant differences were identified using two-way ANOVA.

Attitudinal Inferences Task

Accuracy data were submitted to two-way ANOVA (Group by Question Type), with repeated measures on the Condition variable (10 congruent story factual questions), 10 congruent story inferential questions, 10 incongruent story factual questions, 10 incongruent story inferential questions. Group mean data are provided in Table 4. Only the Condition main effect was significant (F (3,90) = 10.7; $p < .001$); post-hoc comparison of means using the Tukey A procedure indicated that inferential questions about incongruent stories were more difficult than any other questions, for both subject groups.

DISCUSSION

Unlike the original studies, there were no group differences in RHD adults' performance on questions pertaining to revised linguistic or attitudinal

TABLE 4. GROUP MEAN ACCURACY (AND SD) FOR THE ATTITUDINAL INFERENCES TASK

Question Type	*RHD*	*Control*
Congruent/Factual	8.6 (1.4)	9.1 (1.2)
Congruent/Inferential	8.4 (1.7)	9.4 (1.0)
Incongruent/Factual	8.1 (1.8)	8.8 (1.8)
Incongruent/Inferential	7.1 (2.3)	7.8 (2.5)

inferences. Generally, nonsignificant results must be interpreted with caution, because they might simply be an artifact of low experimental power. However, because this study used twice as many subjects as either of the original studies, size effects achieved in the original studies should have been detected easily. Apparently, low power is not the only reason for nonsignificant results.

In interpreting the discrepancy in results, it must be noted that the RHD subjects in this study were relatively mild as a group. That is, behavioral indicators such as performance on neglect and visual perceptual tasks showed high-level functioning relative to population expectations. Group differences might be observed if a greater number of severely impaired RHD subjects were included. In fact, the two RHD subjects who did fall below the normal range on the experimental tasks were among the most impaired on other clinical indicators of RHD (neglect, judgment of line orientation) and on a sentence span measure of auditory working memory (Bloise, 1990). Because Brownell et al. (1986) did not describe their subjects sufficiently, and because Tompkins and Mateer (1985) studied a seizure sample rather than a CVA sample, we cannot ascertain whether severity or some other sampling difference might account for the discrepancy in results.

Several important methodologic changes were intentionally introduced in this study to increase confidence in the experimental findings (see Table 3). We had expected the major findings of the original studies to be replicated despite these extensive methodological modifications, perhaps partly because findings such as these are cited in the literature so often that we had come to accept them as "givens." Our failure to reproduce the original findings reminds us again that the results of the small sample investigations typical of our discipline should be held at arm's length and scrutinized carefully. To enhance the replicability and generalizability of our findings, we should strive consistently to provide procedural detail,

including subject description information (Brookshire, 1983; Tompkins, in press), that is sufficient for others to use in developing their plans and procedures and in interpreting their results.

With the enhanced experimental controls used in this study, the results raise questions about the generality of the previously observed association between RHD and inferencing in general, or inference revision deficits in particular. However, further careful study is warranted to explore issues related to levels of processing required in various inferencing tasks, to assess performance with other types of inferences, and to examine the influence of severity on inferencing performance after right hemisphere brain damage (Bloise, 1990).

There is a critical need to develop or identify sensitive, valid, and reliable indicators of severity for the RHD population. Clinical aphasiologists agree that some of the traditional behavioral and clinical indicators (e.g., presence of contralateral neglect) have not been very informative. They identify the potential value of lower-level sensory and perceptual inference measures (Myers, 1991), attentional allocation tasks (Robin & Rizzo, 1989), working memory measures (Bloise, 1990), a quantifiable indicator of visual neglect (Wilson, Cockburn, & Halligan, 1987) and some of Benton's standardized visuospatial and visuoperceptual tests (Benton, Hamsher, Varney, & Spreen, 1983), among others. Many of these tests and procedures require further development. But progress in this area is imperative to begin to understand and predict covariation among the assorted cognitive, perceptual, and communicative impairments frequently observed in RHD adults and to capture the behavioral heterogeneity that typifies this population.

ACKNOWLEDGMENTS

Preparation of this manuscript was supported by grant #DC00453 from the National Institute on Deafness and Other Communication Disorders. We thank Frank Bloise, Richard Boada, Kathrine McGarry & Susan Ranier for their assistance with this project.

REFERENCES

Benton, A. L., Hamsher, K. deS., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual.* New York: Oxford University Press.

Bloise, C. G. R. (1990). Working memory and inference revision in right hemisphere-damaged adults. Unpublished master's thesis, University of Pittsburgh, PA.

Brookshire, R. H. (1983). Subject description and generality of results in experiments with aphasic adults. *Journal of Speech & Hearing Disorders, 48,* 342–346.

Brownell, H. H., Potter, H. H., Bihrle, A. M., and Gardner, H. (1986). Inference deficits in right brain-damaged patients. *Brain and Language, 227,* 310–321.

Goodglass, H., & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination; Assessment of aphasia and related disorders* (2nd ed.). Philadelphia PA: Lea & Febiger.

Joanette, Y., & Goulet, P. (1987, October). Inferencing deficit in right brain damage: Absence of evidence. Paper presented at 10th European Conference of the International Neuropsychological Society, Barcelona.

Myers, P. S. (1991). Inference failure: The underlying deficit in right hemisphere communication disorders. In T. E. Prescott (Ed.), *Clinical aphasiology* (Vol. 20, pp. 167–180). Austin, TX: PRO-ED.

Robin, D. A., & Rizzo, M. (1989). The effect of focal cerebral lesions on intramodal and cross-modal orienting of attention. In T. E. Prescott (Ed.), *Clinical aphasiology* (Vol. 18, pp. 61–74). Austin, TX: PRO-ED.

Tompkins, C. A. (in press). Improving aphasia treatment research: Some methodological considerations. In *Treatment of aphasia: Research and research needs* [Monograph] NIH (NIDCD): Washington, DC

Tompkins, C. A., & Mateer, C. A. (1985). Right hemisphere appreciation of prosodic and linguistic indications of implicit attitude. *Brain and Language, 24,* 185–203.

Wilson, B., Cockburn, J., & Halligan, P. (1987). *The behavioural inattention test.* Titchfield, England: Thames Valley Test Company.

APPENDIX
STIMULUS VALIDATION PROCEDURES FOR BRIDGING INFERENCES TASK.

1. Fifteen sentence pairs from Brownell et al. (1986), in which misleading information was presented in the first sentence of the pair, were evaluated in their written form by 8 normal judges similar to the expected distribution of experimental subjects in age, gender, and education. The inferences required were not based on the ambiguity of one word but resulted from the meaning of each sentence as a whole.

2. One rating session was held (a) to identify subjects' most likely interpretations of the initial misleading sentence in order to verify that an inference revision was required, and (b) to determine whether the second (revised) interpretation seemed possible. Raters read each misleading sentence, and two sentences would be more likely to follow the first sentence (first interpretation). Next, they were asked to rate how likely it would be that the sentence that they did not choose (second interpretation) could follow the initial sentence (0 = not possible; 1 = possible; 2 = very likely). Determining the viability of the second interpretation was important for establishing whether normally aging adults would be likely to draw the less preferred interpretation of an experimental sentence pair. The rating session was also important for verifying the most likely (first) interpretation of each misleading sentence for members of this age group.

3. Based on the results of this validation session, certain experimental stimuli were chosen for revision. For each sentence written to follow the initial misleading sentence, any rating of "0" by two or more of the eight raters resulted in revision. In addition, stimuli for which the group ratings were inconclusive as to the favored interpretation (e.g., evenly divided "2" ratings for the two possible interpretations) were also revised. This was done to ensure a clear delineation between potential interpretations so that the need for an inference revision was verified. Together, these criteria resulted in the revision of 12 of the original 15 sentence pairs.

4. In another rating task, approximately 6 weeks later, the same judges were presented with two versions of the revised sentence pairs (15 pairs $\times$ 2 versions = 30 pairs total). One version included an initial misleading sentence paired with the sentence intended as its first interpretation; the second version consisted of that same initial sentence paired with the sentence intended as its second interpretation. The two versions of the 15 stimulus pairs were never presented in succession. Subjects rated the

sentence pairs from 0 to 2 (0 = a pair that does not seem to fit together at all; 2 = a pair that fits together well), to indicate their assessment of the differing interpretations. Of the 12 previously revised pairs, 3 were discarded because one or both versions of the sentence pair were rated as "0" by two or more subjects, indicating an unsatisfactory revision. In the remaining nine revised sentence pairs, one version of each was rated as "1" by all but one subject, and the other version was given a rating of "2" by all but one subject.