

Empathy in frontotemporal dementia and Alzheimer's disease

Diego Fernandez-Duque,¹ Sara D. Hodges,² Jodie A. Baird,³
and Sandra E. Black⁴

¹Villanova University, Villanova, PA, USA

²University of Oregon, Eugene, OR, USA

³Swarthmore College, Swarthmore, PA, USA

⁴Department of Medicine, Sunnybrook Hospital, University of Toronto, Toronto, Ontario, Canada

Using naturalistic stimuli, we assessed the ability to infer what other people are feeling in three groups of participants: healthy elderly adults, patients suffering from the behavioral variant of frontotemporal dementia (FTD-b), and patients suffering from Alzheimer's disease (AD). After watching videotaped interviews of everyday people (nonactors) discussing an emotionally relevant event in their lives, participants answered questions regarding the interviewee's feelings. Both patient groups inferred emotions as accurately as the healthy elderly, provided the emotions were displayed unambiguously and consistently across the interview. However, when the displayed emotions became more variable and ambiguous, performance in both patient groups became impaired relative to healthy elderly participants. The similar profile across the two clinical groups despite their differences in social skills suggests that nonsocial cognitive processes affected in dementia may be an important factor in drawing inferences about other people's feelings.

Keywords: Empathy; Dementia; Emotion; Theory of mind; Naturalistic stimulus.

At its most general level, *empathy* refers to the thoughts and feelings that a person experiences when attending to somebody else's experience (S. D. Hodges & Myers, 2007; Ickes, 2003; Preston & de Waal, 2002). Empathy is thought to be abnormal in the behavioral variant of frontotemporal dementia (FTD-b), a disease characterized by inappropriate social behavior, changes in personality, and poor decision making (Fernandez-Duque & Black, 2007; Grossman, 2002; Rankin et al., 2006). In the clinical setting, FTD-b patients are often described by their spouses as unable or unwilling to take other people's feelings into account when deciding how to act (Neary et al., 1998; Snowden et al., 2001).

Empathy skills depend on emotional as well as cognitive processes. Therefore, it is informative to have as a

comparison a group of patients who suffer similar cognitive deficits as patients with FTD-b but, unlike FTD-b, have their social skills intact. Patients at an early stage of Alzheimer's disease (AD) fit such a description. By including an AD group, it is also possible to assess whether cognitive demands contribute to empathy deficits in dementia. A cognitive deficit hypothesis would predict that impairments in empathy will be evident in both patient groups, independent of their social skills. In contrast, a specific deficit in FTD-b and spared performance by the AD group would argue in favor of a mechanism based on emotional dysfunction.

In the pursuit of controlled experimental designs and conceptual clarity, neurological studies have sometimes identified individual components of empathy and have

We would like to thank the patients and their families for their time and effort. Vicki del Rio, Allison Todd, and Ryan Corser helped with data coding. This research was supported in part by a grant from the Center for Consciousness Studies of the University of Arizona to Diego Fernandez-Duque and an operating grant from the Canadian Institute for Health Research to Sandra Black (Grant No. 13129). Diego Fernandez-Duque was also supported by a fellowship from the Heart and Stroke Foundation of Ontario (Grant No. F4866).

Address correspondence to Diego Fernandez-Duque, Psychology Department, Villanova University, 800 Lancaster Ave., Villanova, PA 19085, USA (E-mail: diego.fernandezduque@villanova.edu.).

proceeded to assess those components in isolation. However, social stimuli tend to have ambiguous and convergent features that are often overlooked when examining components in isolation. Take, for example, the inability to recognize static displays of facial anger, a well-established finding in the FTD-b literature (Fernandez-Duque & Black, 2005). Although possible, it is far from clear that this deficit translates into a deficit recognizing anger in natural settings. In real life, facial expressions of anger tend to co-occur with other behavioral cues—body posture, tone of voice, facial movement—that provide convergent information about the person's feelings. Whether that convergent information is sufficient to overcome the face perception deficit remains an open question. Similarly, it is problematic to infer spared capabilities in real life from intact performance in specific processes. For example, in FTD-b, the spared ability to recognize static facial expressions of happiness tells us very little about patients' ability to infer happiness from more ambiguous stimuli typical of natural social interactions. As these examples make clear, the study of FTD-b's difficulties in understanding others' emotions may benefit from the use of more natural, ecologically valid stimuli, particularly when the focus is on abilities related to a multidimensional construct such as empathy (Davis, 1983; S. D. Hodges & Biswas-Diener, 2007). This intermediate-level analysis could serve as a conceptual bridge between clinical descriptions of FTD-b and tightly controlled experiments that use decontextualized stimuli.

In our study, we used videotaped interviews of everyday people (nonactors) discussing a recent emotionally relevant event in their lives. We tested FTD-b patients, early AD patients, and healthy elderly participants in their ability to identify the thoughts and feelings of the interviewees, an approach that has been used successfully in the study of empathy in normal adults (Ickes, 2001; Klein & Hodges, 2001). We supplemented this information with caregivers' reports of patients' empathy.

METHOD

Participants

Patients and healthy volunteers were recruited through the Sunnybrook Dementia Study at Sunnybrook Health Science Centre at the University of Toronto, where the project received approval from the Research Ethics Board. A cognitive neurologist (S.E.B.) assessed all the patients. A history was taken from the patient and from a close relative/caregiver. To rule out contributions from other pathologies, magnetic resonance imaging (MRI) was performed on both patient groups with a 1.5-tesla GE Signa scanner using standard protocol (Callen, Black, Gao, Caldwell, & Szalai, 2001). Apart from atrophy consistent with the patient's dementia, the scans showed no other pathology.

A total of 9 patients with a clinical diagnosis of the behavioral variant of frontotemporal dementia (FTD-b; 7 males), 8 patients with a clinical diagnosis of Alzheimer's

disease (AD; 6 males), and 10 age-matched normal controls (6 males) participated in the study. Besides meeting criterion for behavioral variant of FTD established by the work group on frontotemporal dementia and Pick's disease (McKhann et al., 2001), all the FTD-b patients presented with a corroborated history of progressive decline in social interpersonal conduct. Patients presenting primarily with language complaints (progressive nonfluent aphasia or semantic dementia) were excluded from the study.

All the AD patients met criterion for probable Alzheimer's disease, as established by the workgroup of the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA; McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984). The AD group was slightly older than the FTD-b group, consistent with FTD-b being a pre-senile dementia, but this age difference did not reach statistical significance, $t(15) = 1.7, p = .12$ (see Table 1).

Behavioral symptoms were assessed with the *Frontal Behavioral Inventory* (FBI; Kertesz, Nadkarni, Davidson, & Thomas, 2000). This is a standardized 24-item questionnaire that assesses the major behavioral changes characteristic of FTD-b and has shown excellent reliability in discriminating FTD-b from other dementias (Kertesz et al., 2000). For 8 of the 9 FTD-b patients, the questionnaire was completed with the assistance of their caregiver. Consistent with the clinical diagnosis of FTD-b, all of these patients had abnormal scores even with a conservative cutoff of 30 points (range: 35–48). Signs of neuropsychiatric dysfunction included disinhibition, aberrant motor behavior, apathy, and changes in eating behavior. Assessment of the remaining FTD-b patient was provided by an acquaintance who did not live with the patient. In this case the score was 23, which fell below the cutoff of 30 but nonetheless was significantly higher than the mean score for the AD, which was 8.3 (see Table 1). The AD patients scored within normal range (range: 6–12).

Psychopathology was further evaluated with the *Neuropsychiatric Inventory* (NPI). The NPI is a validated, widely used, semistructured interview by the clinician with the caregiver to assess 12 behavioral domains, including delusions, hallucinations, agitation, depression, anxiety, euphoria, apathy, disinhibition, irritability, aberrant motor behavior, night-time behavior, and appetite disturbance. The inventory takes into account both frequency (on a scale of 0–4) and severity (on a scale of 0–3) of each behavioral domain for a maximum of 12 points in each (Cummings et al., 1994). The amount of distress that each abnormal behavior causes in the caregiver is assessed on a 5-point scale for each domain. The FTD-b group had abnormally high scores on the NPI, particularly in disinhibition, apathy, changes in appetite, and aberrant motor behavior. The amount of stress for the caregivers of FTD-b patients was also larger than that for caregivers of AD patients (see Table 1).

All three groups completed a neuropsychological assessment (see Table 1). All healthy adult participants performed within normal range in this neuropsychological

TABLE 1
Demographic, neuropsychiatric, and neuropsychological information

	Max. score	Healthy elderly	AD	FTD
Age		65.4 (8.5)	67.7 (6.6)	62.3 (6.7)
Sex: male–female ratio		6:4	6:2	7:2
Years of education		16.0 (4.2)	17.0 (3.7)	16.2 (3.1)
FBI	72	n/a	8.3 (3.2)	38.1 (7.5) ^a
NPI total	144	n/a	6.7 (4.2)	27.3 (19.2) ^a
NPI caregiver distress	60	n/a	5.9 (3.7)	13.4 (9.6) ^a
NART-R FS IQ	115.2 (2.2)	114.6 (5.4)	107.6 (10.4)	
MMSE	30	29.0 (0.7) ^{b,c}	25.6 (1.6)	27.0 (1.4)
DRS (total)	144	140.6 (1.1) ^{b,c}	129.6 (7.7)	126.0 (7.7)
Boston Naming	30	28.1 (1.2) ^b	27.1 (2.7)	23.3 (5.8)
WAB total	100	99.2 (0.7) ^{b,c}	95.5 (1.4)	93.0 (3.4)
WAB comprehension	10	10.0 (0.0)	10.0 (0.0)	9.9 (0.1)
WAB fluency	10	10.0 (0.0) ^{b,c}	9.2 (0.5)	9.1 (0.4)
Verbal fluency	48.7 (14) ^{b,c}	30.4 (15.0)	21.6 (10.8)	
Semantic fluency	19.1 (6) ^b	15.8 (3.0)	12.0 (5.0) ^a	
Forward digit span	12	9.4 (1.8)	8.6 (2.2)	8.3 (2.2)
Backward digit span	12	7.9 (1.8) ^{b,c}	6.1 (2.4)	4.8 (2.9)
Trails A	240	36.4 (9.0)	41.4 (13.0)	45.8 (19.0)
Trails B	240	75.1 (23.0) ^{b,c}	148 (70.0)	139 (64.0)
Trails B error		0.0 (0.0)	1.0 (1.4)	0.9 (1.1)
CVLT Acquisition ^d	80	46.4 (7.7) ^{b,c}	28.8 (6.3)	34.8 (7.3)
CVLT Delayed Free Recall ^d	16	9.2 (3.2) ^{b,c}	1.9 (2.5)	5.9 (2.6) ^a
Ravens ^d	36	32.8 (1.1) ^b	29.1 (4.0)	25.0 (8.0)
Line Orientation task ^d	30	25.6 (6.0) ^b	24.5 (4.0)	20.2 (5.0)
Visual memory immediate ^d	41	32.0 (3.0) ^{b,c}	24.0 (7.0)	18.9 (5.0)
Visual memory delayed ^d	41	23.8 (4.0) ^{b,c}	6.9 (7.0)	9.5 (10.0)
Rey Copy ^d	36	33.4 (3.0) ^b	30.9 (5.0)	26.0 (6.0)
WCST Categories	6	2.2 (1.3)	2.3 (1.4)	2.6 (1.4)
WCST Correct	64	44.0 (8.7)	40.6 (10.0)	42.0 (13.0)

Note. AD, Alzheimer's disease; FTD-b, behavioral variant of frontotemporal dementia; FBI, Frontal Behavioral Inventory; NPI, Neuropsychiatric Inventory; NART-R, National Adult Reading Test–Revised; FS IQ, full scale IQ; MMSE, Mini-Mental State Examination; DRS, Dementia Rating Scale; WAB, Western Aphasia Battery; CVLT, California Verbal Learning Test; WCST, Wisconsin Card Sorting Test. All group comparisons were made using Mann–Whitney *U*, significance at $p < .05$.

^aFTD-b significantly different from AD. ^bHealthy elderly significantly different from FTD-b. ^cHealthy elderly significantly different from AD. ^dData from one FTD-b patient were unavailable for this task.

assessment. As expected, both patient groups were impaired relative to healthy elderly adults in most domains. More importantly, the patient groups were well matched in most domains. Exceptions included the semantic fluency task, for which the FTD-b group performed significantly worse than AD group, and the delayed free recall of the California Verbal Learning Test in which the FTD-b group performed better than AD group. These results are consistent with previous findings from the literature (J. R. Hodges et al., 1999).

Stimuli

The stimuli for this study included three videotapes, each depicting a different woman (nonactor) describing a recent personal experience (all experiences were real). The first video showed a woman who had recently (in the last 4 months) given birth to her first child, and who was very happy in her new role as a mother. The second

video showed a senior citizen explaining some of the difficulties she had recently had with her housemate. The mood of the interviewee in this videotape was clearly negative and included complaints about her housemate's behavior and the lack of options available to the interviewee. The third—and most critical—video showed another woman who had also recently become a mother, but who expressed more ambivalence about her experience. Each video lasted approximately 5 minutes.

At the time these videos were made, immediately following the taping, each interviewee watched her own videotape, stopping the tape at each moment she remembered having had a specific thought or feeling. The interviewee then wrote down the thoughts or feelings that she remembered having had during the interview at each of these stop points. We used these self-reported thoughts and feelings to guide our selection of candidate words for the *Empathic Inference Questionnaire* (described below) and to select the *stop points* at which we would

ask patients to infer the interviewee's feelings. A similar protocol has been used successfully in other studies of empathy (S. D. Hodges, Kiel, Kramer, Veach, & Villanueva, 2009; Klein & Hodges, 2001).

The videos were presented in fixed order across two testing sessions that occurred between a week and a month apart. In Session 1, participants saw the video of the happy new mother. In Session 2, participants watched the video of the unhappy elderly woman followed by the video of the ambivalent mother. Participants completed unrelated tasks for approximately 30 minutes between the last two videos.

Empathic inference questionnaire

Studies of empathy usually define "empathic accuracy" as the match between an interviewee's self-report and a participant's inference of that interviewee's feelings (Ickes, 2001). Although this approach is very useful, it does depend heavily on the participant's ability and willingness to voice those inferences. There are a variety of reasons why FTD-b patients might not fit this criterion, including reduced verbal output, impaired semantic ability, and/or lack of motivation. For example, reduced speech is a main symptom of FTD in its "progressive nonfluent aphasia" variant, and it is also an occasional symptom in its behavioral variant. We worried that if asked for open-ended responses, FTD-b patients would be less talkative than the AD patients, and their responses would be incorrectly coded as less empathically accurate.

To overcome these limitations, we developed a simplified empathic accuracy measure. As in most studies of empathy, we first asked interviewees to report their thoughts and feelings at each stop point. Next, from each of those statements we selected a word or two that best captured the statement's gist; we also included words that were not related to the statement feelings; and occasionally we also added yet another word that was only moderately related to the statement. For example, at the first stop point of the interview, the positive mom reported "thinking that I am tired all the time but it is worth it," the ambivalent mother reported "I feel overwhelmed at times of being a mom," and the senior citizen reported "I feel I am pleading and not being heard." Thus, for that first stop, we listed the following adjectives: "tired," "fulfilled," "frustrated," "overwhelmed," "ignored," "joy," "mixed feelings," and "She thinks that having a baby is worth it" (this last one for the mother videos only). This approach guaranteed that at each interviewee's stop point there would at least one item that captured the interviewee's thoughts and feelings and also items that were unrelated or only moderately related to such feelings. For example, for the statement "I'm tired all the time but it is worth it" there were three well-matched items ("tired," "fulfilled," "She thinks that having a baby is worth it"), four unrelated items ("frustrated," "overwhelmed," "ignored," "mixed feelings"), and a moderately related item ("joy"). Occasionally, none of the self-report words was able to capture, in isolation, the sentiment of the statement. In

those cases, we drafted words that in our opinion did capture the gist of the sentiment. For example, at one point the ambivalent mother reported that she was "thinking that I am not a very good mom when I left him by himself." This led us to include the words "unfit" and "guilt" to our adjective list.

To confirm that our choices did indeed capture the feelings of the interviewees' self-reports, we recruited three independent coders who had never watched the videos. We asked these coders to read each of the interviewees' self-reports and select the adjectives that best captured (in part or in full) the feeling of the statement. For each statement, the list of adjectives that coders chose from was the same list that participants chose from in the actual study. Coders almost always endorsed the closely related items (92%), almost never endorsed unrelated items (6%), and sometimes endorsed moderately related items (56%).¹ These results support the contention that the list of adjectives we created could serve as an effective proxy to assess empathic accuracy (i.e., a matching between the interviewee's feelings and thoughts and the participant's inference about those thoughts and feelings).

Procedure

At the beginning of each video, participants were reminded that they would be asked about the thoughts and feelings expressed by the interviewee. After watching each interview uninterrupted once, the participant watched it a second time. This second time the videotape was paused four times (at the *stop* points). At each stop point, the participant was presented with a list of 6 to 8 adjectives/statements. For each item, the participant had to decide whether the adjective/statement resembled the feelings and thoughts the interviewee had previously reported having at that time. The participant answered "yes" or "no" to each item.

There were a total of 29 adjectives/statements distributed across the four stop points (see the Appendix). A total of 11 had a positive connotation, and 18 had a negative connotation. To reduce semantic knowledge demands, we used everyday words. The same set of 29 items was used for each of the three videos, with the exception of a statement regarding motherhood, which was not applicable to the video of the elderly woman.

¹Not only did coders agree with our decision of which items matched the interviewee's statements, but also the intercoder agreement was very high, ranging from 82% agreement between Coders 1 and 2 to 88% agreement between Coders 2 and 3. This agreement remained very high independent of which interviewee's feelings were being assessed. There was an intercoder agreement of 88.1% for emotions attributed to the positive mother, 83.9% for emotions attributed to the ambivalent mother, and 82.7% for emotions attributed to the senior citizen. A chi-square analysis revealed no statistically significant differences, $\chi^2(2) = 1.19, p > .50$.

Adjustment questionnaire and maternal attitudes questionnaire

Two additional measures of empathy were included for the videotapes of the new mothers. The *Adjustment Questionnaire* included four questions that asked about the mother's adjustment to motherhood (e.g., "she is coping with new motherhood . . .") as well as her beliefs about it (e.g., "she thinks she is coping . . ."). Questions were answered on a 5-point scale. The *Maternal Attitudes Questionnaire*, based on Warner, Appleby, Whitton, and Faragher's (1997) measure of the same name, asked participants to guess how the new mother targets would respond to various statements about motherhood. To accommodate our patient sample, we converted the original Likert scale questions to two-alternative (yes/no) items. This simplification was aimed to facilitate comprehension. We omitted two of the original items that did not work in this format. This left 12 statements, 2 of which had a positive connotation (e.g., she is proud of being a mother), and 10 had a negative connotation (e.g., she regrets having the baby).

Liking Questionnaire

This was a five-question questionnaire asking participants how much they were interested in the target, how likable and similar to them she was, whether she seemed like someone who would be their close friend, how well they understood her, and how well they understood her situation. Participants answered on a 5-point scale. The *Liking Questionnaire* differed from the other questionnaires in that it asked participants to report their own feelings rather than asking them to infer the feelings of the interviewee.

Caregivers' report of patients' empathy

To assess caregivers' opinions about the empathy skills of the patient under their care, we asked caregivers to complete a third-person version of the *Interpersonal Reactivity Index* (IRI). The IRI is a questionnaire used in social psychology research as a self-report, first-person measure of empathic tendencies (Davis, 1983). However, clinical researchers have successfully adapted the scale to allow third-person reports, so that patients' primary caregivers could complete the measure about the patients (Rankin et al., 2006). Used this way, the scale allows an assessment of patients' empathy as communicated in daily social interactions with their caregivers. In social interactions, how effectively the person communicates his or her feelings will be an important part of whether he or she is perceived as empathic. The third-person version of the IRI questionnaire provides a window into those communicative skills.

Primary caregivers (i.e., spouses) were asked to rate the patient on a 5-point Likert scale ranging from 1 (*does not describe him/her well*) to 5 (*describes him/her very well*). In the case of one FTD-b patient the questionnaire was completed by a friend who did not live with the patient nor was his primary caregiver. Data were not available for one of the AD patients. We used three of

the IRI's subscales: *perspective taking*, *empathic concern*, and *personal distress*.² *Perspective taking* measured the tendency to spontaneously adopt another person's point of view (e.g., "He believes there are two sides to every question and tries to look at them both"). *Empathic concern* is the feeling of sympathy and compassion for other people (e.g., "He often has tender, concerned feelings for people less fortunate than him"). *Personal distress* is a "self-oriented" feeling of anxiety and discomfort (e.g., "In emergency situations, he feels apprehensive and ill-at-ease"). Each subscale had seven items. Reliability was acceptable for each subscale in each group (Cronbach's alphas > .75).

Data analysis

For each participant and each of the 3 videos, the *Empathic Inference Questionnaire* collected 29 yes/no judgments describing the interviewee's thoughts and feelings. Some of these adjectives were positive (e.g., glad, loving), and others were negative (e.g., hopeless, sad). In order to make all of the questions yield "yes" answers when participants selected the more negative option (e.g., when participants endorsed that the woman in the video was hopeless, or when they did not endorse that she was happy), we reversed the scoring for all of positive items. Thus, for each of the items, a score of "1" indicated that the participant endorsed the negative option, and a score of "0" indicated that the participant endorsed the positive option. Next, we calculated the percentage of items endorsed by each participant for each video. High percentage numbers meant that the participant attributed to the interviewee many negative thoughts and feelings and only a few positive ones.

For data that were not normally distributed, we report nonparametric statistics (for such data, the pattern of result did not change when using parametric statistics). Related samples were compared using Wilcoxon rank tests, and independent samples were compared using Mann-Whitney *U*. All reported differences were significant at $p < .05$.

RESULTS

Figure 1 shows the results for each of the three videos. There were no group differences regarding the thoughts and feelings of the happy mother: FTD-b and AD patients resembled healthy adults in recognizing her positive feelings and thoughts and thus endorsing a very low percentage of negative attributions. Given that recognition of positive emotions is thought to be spared in FTD-b, this result should not be surprising. Nonetheless, it provides an important baseline for performance: It suggests that patients understood the instructions, followed the video, and understood the meaning of the words presented in the questionnaire. It suggests that

²The IRI also has a "fantasy" subscale. We did not include this scale because it was less relevant to the concept of empathy as discussed in this study.

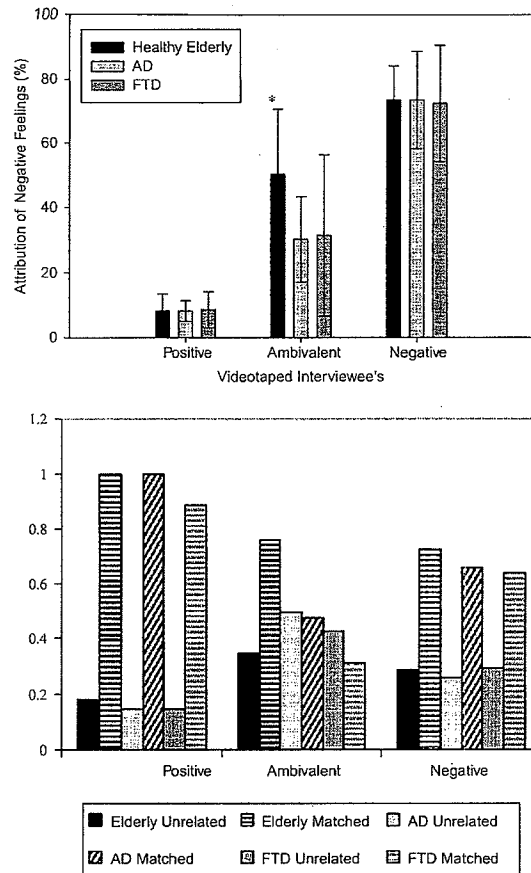


Figure 1. Empathic Inference Questionnaire. (A) The graph shows the percentage of negative feelings attributed to each videotaped interviewee by each group of participants (AD, Alzheimer's disease; FTD, frontotemporal dementia). Both patient groups were accurate in their assessment of the unambiguous videos, whether it was the positive mother or the negative elderly woman. However, both patient groups were overly optimistic when assessing the feelings of the interviewee who expressed ambivalent feelings relative to the judgment by the healthy elderly group. Vertical bars depict standard deviations for each group in each videotape. The asterisk denotes that, relative to the Healthy Elderly group, the patient groups (AD, FTD) endorsed a more positive description of the ambivalent interviewee, $p < .05$. (B) This graph shows participants' endorsement of adjectives that were either well matched or unrelated to the feelings and thoughts expressed by the interviewee in the self-report statement. For the unambiguous videos, both patient groups endorsed primarily those items that were well matched with the interviewee's report (i.e., empathic accuracy). However, for the interviewee who was ambivalent in her feelings, patients were unable to discriminate well-matched items from unrelated items. Healthy adults, on the other hand, were capable of such discrimination.

patients in both groups remembered the information provided in the videos well enough to respond as intended. The ceiling performance also rules out response biases, such as a perseveration in responding "yes," as both positive and negative adjectives were included.

More importantly, there were no group differences in the attribution of negative thoughts and feelings to the senior citizen interviewee. Performance on this interviewee was almost identical in all three groups, minimizing the concern that the null effect may have stemmed from reduced statistical power. Thus, we found no evidence for the hypothesis that in this task empathic accuracy about negative emotions was specifically impaired in FTD-b. However, using a target interviewee whose feelings were unambiguously negative may result in an overestimate of patients' ability to make emotion attributions. That is, patients' relatively good performance could have stemmed from a global judgment on the interviewee's feelings, rather than from a sophisticated understanding of her particular thoughts and emotions. The same could be argued about patients' excellent performance attributing mental states to the mother who was unambiguously happy.

A more subtle display of negative and positive emotions should provide a more sensitive assessment of empathy skills. This assessment was obtained with the video of the ambivalent mother whose feelings and thoughts were a mix of positive and negative. Patients in both clinical groups endorsed a more positive description (e.g., a lower percentage of negatively valenced attributions) of this interviewee than did healthy adults. This point is illustrated in the center part of Figure 1, which shows the percentage of negative feelings attributed to the ambivalent mom by each participant group. The overoptimistic bias of both patient groups was revealed by a Mann-Whitney U test that compared the healthy adult group against the two patient groups combined, $U = 39$, $Z = -2.3$, $p < .05$. Importantly, this overoptimistic bias was not specific to FTD-b patients. In other words, there was no significant difference between FTD-b and AD in their judgment of the ambivalent mom.

As one would expect, judgments by healthy adults were more variable for the ambivalent mother than for the other two interviewees (see standard deviations in Figure 1). This raises a question of what counts as a "correct" response. In cognitive and perceptual tasks there is usually an objective criterion against which performance can be assessed. In much social perception research, including some empathy research, the "correct" response is often determined by a consensus criterion: What goes on in the interviewee's mind is defined as what most healthy adults agree goes on in the interviewee's mind (Robins & John, 1997). If healthy adults cannot agree, it is hard to define how patients' assessments depart from normal.

To address this problem, we reanalyzed the data including only adjectives for which there was consensus. These were adjectives that at least 8 out of 10 healthy elderly agreed applied to the interviewee or agreed did not apply to the interviewee. A total of 13 items that did not meet this criterion were excluded (6 positive, 7 negative). Data from the remaining 16 items (5 positive, 11 negative) about which there was consensus replicated the findings of the initial analysis: Patients showed an overoptimistic bias even against the consensus of healthy adults, $U = 21$, $Z = -3.2$, $p < .001$; healthy elderly, 50.6%

(14); AD, 70.3% (10); FTD-b, 68.0% (24). Once again, the deficit was not specific to the FTD-b group, in that AD patients were as optimistic in their judgment of the ambivalent mom as FTD-b patients were.

Yet another way we measured empathic accuracy was by assessing whether participants' responses to the list of adjectives fit a priori categories used in developing our list of adjectives. As described in the Method section, our list of adjectives included many words used by the interviewees to express their feelings at the stop points (i.e., during self-report, rather than in the videotaped interview). The list also included many words unrelated to those feelings. We asked whether participants would be able to discriminate words between these two categories. If so, it would suggest that participants had some access to the feelings and thoughts of the interviewee. In other words, it would suggest some degree of empathic accuracy.

Data from each video stimuli were submitted to a 3×2 analysis of variance (ANOVA) that had group (healthy elderly, AD, FTD-b) as a between-subjects factor and adjective-target relation (well matched, unrelated) as a within-subject factor.³ The proportion of endorsed adjectives served as the dependent variable. For the unambiguous videos, the healthy elderly as well as both patient groups endorsed primarily items that were well matched to the interviewee's feelings and thoughts (i.e., empathic accuracy; see Figure 1B). This result is shown in the lack of interaction between group and adjective-target relation, both in the positive mom video, $F(2, 24) = 1.88, p = .17$, and in the negative senior citizen video, $F(2, 24) = 0.32, p = .72$. For the video of the ambivalent mom the pattern looks quite different, with a significant interaction between group and adjective-target relation, $F(2, 24) = 7.9, p = .002$. Follow-up t tests reveal that, as in the other videos, the healthy elderly group remains capable of limiting their endorsement to items that are well matched to the interviewee's feelings and thoughts, $t(1, 9) = 4.4, p = .002$. In contrast, AD and FTD-b were unable to discriminate adjectives that match from adjectives that did not match the interviewee's feelings and thoughts: AD, $t(1, 7) = 0.18, p = .86$; FTD-b, $t(1, 7) = -1.0, p = .34$ (see Figure 1B). In other words, for ambiguous videos, both patient groups exhibited a deficit in empathic accuracy.

For the analysis of the Adjustment Questionnaire and the Maternal Attitude Questionnaire we reversed the score of the positive items so that we could compute a total score indicating the number of items for which participants attributed a more negative response to the interviewee (see Table 2).

In the Adjustment Questionnaire, all groups perceived the ambivalent mother as having a harder time adjusting to her new role than the happy mother. Perceived difficulty adjusting to motherhood was larger for the ambivalent

³In an initial analysis, we included interviewee type (positive, ambivalent, negative) as a third factor. This analysis revealed a significant three-way interaction, $F(4, 48) = 3.45, p = .01$, which we explored by running 3×2 analyses at each level of interviewee type, as described in the main text.

TABLE 2
Perceived difficulty adjusting to motherhood and perceived maternal attitude

Questionnaire	Videotaped interviewee		Difference
	Positive mom	Ambivalent mom	
Adjustment ^a			
Healthy elderly	0.3 (0.3)	2.1 (0.6)*	1.8 [^]
AD	0.6 (0.4)	1.8 (0.5)*	1.2
FTD-b	0.4 (0.3)	1.7 (1.2)*	1.3
Maternal attitude ^b			
Healthy elderly	19 (18)	48 (26)*	29
AD	16 (10)	35 (13)*	19
FTD-b	24 (14)	45 (25)*	21

Note. The values shown are the scores for the two videos of new mothers (the positive mother and the ambivalent mother) and the difference between each of these two scores. Higher scores mean more negative attributions. Standard deviations are in parentheses. AD, Alzheimer's disease; FTD-b, behavioral variant of frontotemporal dementia.

^aMean responses using a 5-point scale (1 = very easily; 5 = very poorly). ^bPercentage of negative maternal attitude statements endorsed by each group of participants.

*Wilcoxon paired test, $p < .05$. [^]Mann-Whitney U on the difference scores (healthy elderly vs. clinical groups combined), $p < .05$.

mother than for the positive mother in each of the three groups, as revealed by Wilcoxon tests for related samples run in each group, $Z_s > 2.5, p_s < .05$ (see Table 2).

More importantly, the clinical groups did not notice as much of a difference between the two moms as the healthy elderly did, a result revealed in a Mann-Whitney U test comparing both clinical groups against the healthy group on their difference scores, $U = 44, Z = -1.9, p < .05$.

In the Maternal Attitude Questionnaire, once again each of the groups perceived the ambivalent mom as having a harder time than the happy mom, $Z_s > 2.5, p_s < .05$. As before, healthy adults made a stronger discrimination between the two moms than did the clinical groups, although in this case the group difference failed to reach statistical significance ($Z = 1.4, p = .15$).

In the Liking Questionnaire (see Table 3), all three groups found the positive mom to be most likable while finding the senior citizen least likable. Wilcoxon tests were significant for each of the participant groups, $Z_s > -3, p_s < .01$. All three groups of participants found the mom with ambivalent feelings moderately likable. There were no group differences among participants in their overall liking and perceived similarity to the interviewees, Mann-Whitney $Z_s < 1, p_s > .30$.

As assessed by the Interpersonal Reactivity Index, FTD-b patients were seen by their caregivers as having a larger deficit in perspective taking than the AD patients, $U = 9, Z = 2.4, p < .05$ (see Figure 2). FTD-b patients were also seen as having less empathic concern than AD patients, $U = 14, Z = 1.9, p < .06$. There was no statistically significant difference between groups in the amount

TABLE 3

Liking Questionnaire percentage scores for each participant group, as a function of interviewee's feelings

	Videotaped interviewee's feelings		
	Positive	Ambivalent	Negative
Healthy elderly	4.3 (0.8) ^a	3.8 (0.7)	3.5 (0.6) ^b
AD	4.3 (0.7) ^a	3.8 (0.8)	3.3(0.8) ^b
FTD	4.1 (0.6)	3.7 (1.1) ^c	2.9 (0.7) ^b

Note. Table shows average across five questions, in a 5-point scale (5 = greater liking). Standard deviations are in parentheses. AD, Alzheimer's disease; FTD-b, behavioral variant of frontotemporal dementia.

Wilcoxon rank tests ($p < .05$) comparing: ^apositive/ambivalent, ^bpositive/negative, and ^cambivalent/negative.

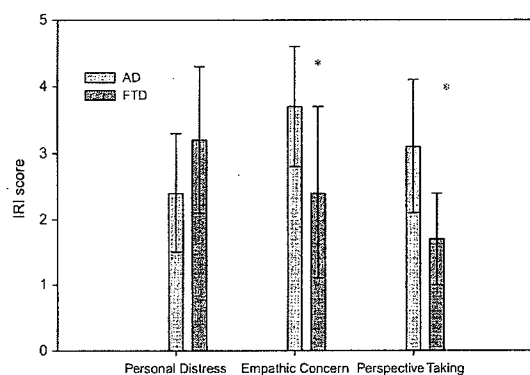


Figure 2. Results on a third-person adaptation of the Interpersonal Reactivity Index (IRI), a questionnaire measure of communicative empathy. FTD-b (behavioral variant of frontotemporal dementia) patients were perceived as being less sympathetic and less amenable to perspective taking than AD (Alzheimer's disease) patients. Vertical bars depict standard deviations for each group. The asterisks denote the subscales in which there was a significant difference between AD and FTD.

of personal distress, $U = 48$, $Z = 1.1$, ns . These results are consistent with previous reports and suggest that FTD-b patients are seen by their caregivers as lacking in perspective taking and empathic concern (Rankin et al., 2006).

GENERAL DISCUSSION

This study used naturalistic stimuli in an experimental setting to assess the empathic abilities of patients with FTD-b. The results revealed both similarities and differences with past studies in which tightly controlled but decontextualized stimuli were used for investigating empathy and its components. For example, similar to studies of emotion recognition from static facial displays, the current study showed spared ability by FTD-b and AD patients to attribute positive feelings when appropriate. Unlike those previous studies, however, FTD-b patients were also capable of attributing negative

emotions provided that they were displayed unambiguously and consistently. This finding should serve as a reminder that deficits in the processing of impoverished stimuli (e.g., static facial displays) do not necessarily imply deficits in the processing of richer displays (e.g., live action videos).

Both FTD-b and AD patients appear to have relied on a global judgment about the interviewee's overall feelings and situation, rather than trying to decode each particular thought or emotion. Consistent with this *global discrimination* hypothesis, patients' performance suffered when the displayed emotions became more variable and ambiguous. For example, both groups of patients were too optimistic when inferring the feelings of the interviewee who was talking about what is usually considered a happy event (new motherhood) but whose own emotions about this event were mixed. The exact mechanism of this bias is unclear and should be the topic of future studies. One possibility is that patients were overreliant on a "motherhood" schema (i.e., a framing of motherhood as a happy event). Other deficits in social cognition might have further contributed to the difficulties in empathic accuracy. Alternatively, the deficit might have had an independent cause in each disease. For example, FTD-b may selectively impair the encoding of facial and prosodic information, while AD might selectively impair the encoding of verbal information. Future studies should test these competing hypotheses by systematically manipulating visual, prosodic, and linguistic information.

Interestingly, both patient groups performed normally on the Liking Questionnaire. In other words, patients' reports were indistinguishable from healthy adults when judging how much they liked each interviewee. The Liking Questionnaire differed from the other questionnaires in that it asked participants to report their own feelings rather than asking them to infer the feelings of the interviewee. Liking or disliking somebody is a simpler process than inferring the person's feelings and thoughts. Such simplicity may help explain the relatively good performance by patients on this questionnaire.

An unexpected finding of our study was that the deficit in empathic accuracy was not limited to FTD-b but rather it generalized to patients with AD. This finding argues that cognitive deficits can lead to impaired empathy even in patients who do not display obvious social deficits. The result is consistent with findings from studies of mental state reasoning, which show that FTD-b and AD patients have similar difficulty in false-belief tasks, particularly under increased working-memory demands (Fernandez-Duque, Baird, & Black, 2007; Zaitchik, Koff, Brownell, Winner, & Albert, 2006). Together with the findings from the current study, these theory-of-mind studies highlight the importance of cognitive deficits as a possible contributor to empathy problems (Fernandez-Duque & Baird, 2005; Porter, Coltheart, & Langdon, 2007).

One of the contributions of the current study is methodological. It shows that it is possible to use naturalistic stimulus and still retain a fair amount of control on the quality of the data. Our measure of empathic accuracy

was easy for patients to understand and respond to, as well as easy for researchers to quantify. Rather than asking for open-ended responses, which might be distorted by reduced verbal output in FTD-b patients, the measure in our study asked for a yes/no response to a set of familiar adjectives. Its low demand on semantic knowledge is well suited for clinical studies. Furthermore, the same list of words was used in all three videos. This ensured that the options were constant across the three videos. Thus, when differences between groups emerged as a function of type of video (e.g., positive vs. ambiguous emotion) we could be sure that the difference had not stemmed from differences in semantic or verbal output problems. Nevertheless, our new method is not without limitations, such as the fact that it makes it more difficult to match the participants' inferred emotions to the target's reported emotions.

In the questionnaire measure of communicative empathy, caregivers of FTD-b patients saw them less capable—or less willing—to take the perspective of somebody else. This perception contrasted with FTD-b patients' empathic accuracy performance, which showed them to be as accurate as AD patients in reporting the feelings of the videotaped interviewees. This raises the possibility that empathy problems in FTD-b are exacerbated by expressive deficits. For example, studies of facial muscle movements have revealed that FTD patients are impaired in generating a spontaneous smile (Perry et al., 2001). Other expressive abnormalities, such as a reduction in goal-directed behavior and a tendency to avert gaze, may further compromise FTDs' ability to communicate their emotions (Brown & Pluck, 2000). The result is also consistent with studies of empathy using healthy, younger participants, where performance on accuracy tasks is independent of other factors that influence how empathic these participants are perceived to be (S. D. Hodges, 2005).

One of the biggest strengths of the current study is its reliance on naturalistic stimuli. However, this is also the study's biggest weakness: Naturalistic stimuli are highly idiosyncratic. This limitation becomes even more serious when only a few participants are tested. This is often the case in frontotemporal dementia research, and ours is no exception. Thus, it remains to be seen whether the findings reported here generalize to other stimuli and methodologies. Some of our results, in particular, go against the established view and therefore need to be interpreted very carefully. One such example is the lack of group differences in empathic accuracy between FTD-b and AD. Another example is the relatively spared performance by patients with FTD-b when judging a consistently negative video. Other findings are probably less controversial, such as the evidence that both patient groups were impaired when presented with an ambiguous display of emotions. Nonetheless, it is worth stressing that such finding is based on performance in a single videotape and thus ought to be replicated with other stimuli before strong conclusions can be drawn.

These limitations notwithstanding, the findings from the current study highlight the important role that naturalistic stimuli can have in the study of empathy in

clinical populations. The videotaped interviews we used approximated the kind of information one might encounter in real life. It is also true that these videotapes are not a perfect simulation of the social world, as they don't afford the true give-and-take of natural interaction. New methodologies are being developed that would make such testing possible (Beer, John, Scabini, & Knight, 2006; Sturm, Rosen, Allison, Miller, & Levenson, 2006; Zaki, Bolger, & Ochsner, 2008). Such new methodologies, together with the paradigms presented here, should prove fruitful to researchers seeking a naturalistic exploration of empathy in the emerging field of social neuroscience (Decety & Jackson, 2004).

Original manuscript received 25 July 2008

Revised manuscript accepted 26 April 2009

First published online 18 June 2009

REFERENCES

- Beer, J. S., John, O. P., Scabini, D., & Knight, R. T. (2006). Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions. *Journal of Cognitive Neuroscience*, *18*, 871-879.
- Brown, R. G., & Pluck, G. (2000). Negative symptoms: The "pathology" of motivation and goal-directed behaviour. *Trends in Neurosciences*, *23*, 412-417.
- Callen, D. J., Black, S. E., Gao, F., Caldwell, C., & Szalai, J. P. (2001). Beyond the hippocampus: MRI volumetry confirms widespread limbic atrophy in AD. *Neurology*, *57*(9), 1669-1674.
- Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, *44*, 2308-2314.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113-126.
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71-100.
- Fernandez-Duque, D., & Baird, J. A. (2005). Is there a "social brain"? Lessons from eye-gaze following, joint attention, and autism. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 75-90). New York: Guilford Press.
- Fernandez-Duque, D., Baird, J. A., & Black, S. E. (2007). False-belief understanding in fronto-temporal dementia and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *31*(4), 489-497.
- Fernandez-Duque, D., & Black, S. E. (2005). Impaired recognition of negative facial emotions in patients with frontotemporal dementia. *Neuropsychologia*, *43*, 1673-1687.
- Fernandez-Duque, D., & Black, S. E. (2007). Metacognitive judgment and denial of deficit: Evidence from frontotemporal dementia. *Judgment and Decision Making*, *2*, 359-370.
- Grossman, M. (2002). Frontotemporal dementia: A review. *Journal of the International Neuropsychological Society*, *8*, 566-583.
- Hodges, J. R., Patterson, K., Ward, R., Garrard, P., Bak, T., Perry, R., et al. (1999). The differentiation of semantic dementia and frontal lobe dementia (temporal and frontal variants of frontotemporal dementia) from early Alzheimer's disease: A comparative neuropsychological study. *Neuropsychology*, *13*, 31-40.
- Hodges, S. D. (2005). Is how much you understand me in your head or mine? In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 298-309). New York: Guilford Press.

- Hodges, S. D., & Biswas-Diener, R. (2007). Balancing the empathy expense account: Strategies for regulating empathic response. In T. F. D. Farrow & P. W. R. Woodruff (Eds.), *Empathy in mental illness* (pp. 389–407). Cambridge, UK: Cambridge University Press.
- Hodges, S. D., Kiel, K. J., Kramer, A. D. I., Veach, D., & Villanueva, B. R. (2009). *Giving birth to empathy: The effects of similar experience on empathic accuracy, empathic concern, and perceived empathy*. Manuscript in preparation, University of Oregon.
- Hodges, S. D., & Myers, M. W. (2007). Empathy. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 296–298). Thousand Oaks, CA: Sage.
- Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 219–241). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ickes, W. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus.
- Kertesz, A., Nadkarni, N., Davidson, W., & Thomas, A. W. (2000). The Frontal Behavioral Inventory in the differential diagnosis of frontotemporal dementia. *Journal of the International Neuropsychological Society*, 6(4), 460–468.
- Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, 27, 720–730.
- McKhann, G. M., Albert, M. S., Grossman, M., Miller, B., Dickson, D., & Trojanowski, J. Q. (2001). Clinical and pathological diagnosis of frontotemporal dementia: Report of the work group on frontotemporal dementia and Pick's disease. *Archives of Neurology*, 58, 1803–1809.
- McKhann, G., Drachman, D., Folstein, M. F., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology*, 34, 939–944.
- Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., et al. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, 51, 1546–1554.
- Perry, R. J., Rosen, H. R., Kramer, J. H., Beer, J. S., Levenson, R. L., & Miller, B. L. (2001). Hemispheric dominance for emotions, empathy and social behaviour: Evidence from right and left handers with frontotemporal dementia. *Neurocase*, 7, 145–160.
- Porter, M. A., Coltheart, M., & Langdon, R. (2007). The neuropsychological basis of hypersociability in Williams and Down syndrome. *Neuropsychologia*, 45, 2839–2849.
- Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *The Behavioral and Brain Sciences*, 25, 1–20; discussion 20–71.
- Rankin, K. P., Gorno-Tempini, M. L., Allison, S. C., Stanley, C. M., Glenn, S., Weiner, M. W., et al. (2006). Structural anatomy of empathy in neurodegenerative disease. *Brain: A Journal of Neurology*, 129, 2945–2956.
- Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 649–679). San Diego, CA: Academic Press.
- Snowden, J. S., Bathgate, D., Varma, A., Blackshaw, A., Gibbons, Z. C., & Neary, D. (2001). Distinct behavioural profiles in frontotemporal dementia and semantic dementia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 70, 323–332.
- Sturm, V. E., Rosen, H. J., Allison, S., Miller, B. L., & Levenson, R. W. (2006). Self-conscious emotion deficits in frontotemporal lobar degeneration. *Brain: A Journal of Neurology*, 129, 2508–2516.
- Warner, R., Appleby, L., Whitton, A., & Faragher, B. (1997). Attitudes toward motherhood in postnatal depression: Development of the maternal attitudes questionnaire. *Journal of Psychosomatic Research*, 43(4), 351–358.
- Zaitchik, D., Koff, E., Brownell, H., Winner, E., & Albert, M. (2006). Inference of beliefs and emotions in patients with Alzheimer's disease. *Neuropsychology*, 20, 11–20.
- Zaki, J., Bolger, N., & Ochsner, K. N. (2008). It takes two: The interpersonal basis of empathic accuracy. *Psychological Science*, 19, 399–404.

APPENDIX

Alphabetical list of items used to construct the Empathic Inference Questionnaire.

A (2) appears next to items that were repeated.

Alarmed	Outraged (2)
Amused	Overwhelmed
Ashamed	Relief
Frustrated	Sad
Fulfilled	Satisfied (2)
Glad	Scared
Guilty (2)	Tired
Happy	Unfit
Hopeless	Unprepared
Ignored	Upset
Joy	She thinks that having a baby/roommate is worth it
Loving	She thinks that motherhood comes naturally to her
Mixed Feelings (2)	