Metacognitive judgment and denial of deficit: Evidence from frontotemporal dementia

Diego Fernandez-Duque*
Department of Psychology
Villanova University

Sandra E. Black
Department of Medicine
University of Toronto

Abstract

Patients suffering from the behavioral variant of Frontotemporal Dementia (FTD-b) often exaggerate their abilities. Are those errors in judgment limited to domains in which patients under-perform, or do FTD-b patients overestimate their abilities in other domains? Is overconfidence in FTD-b patients domain-specific or domain-general? To address this question, we asked patients at early stages of FTD-b to judge their performance in two domains (attention, perception) in which they exhibit relatively spared abilities. In both domains, FTD-b patients overestimated their performance relative to patients with Dementia of Alzheimer Type (DAT) and healthy elderly subjects. Results are consistent with a domain-general deficit in metacognitive judgment. We discuss these findings in relation to "regression to the mean" accounts of overconfidence and the role of emotions in metacognitive judgments.

Keywords: frontotemporal dementia, metacognition, overconfidence, generality.

1 Introduction

How good are you at multitasking? How good are you at remembering people's names? Judgments people make about their own cognitive abilities vary greatly from one person to another. These judgments about the self are also systematically biased relative to the judgments made by others about us (Gilovich, Griffin, & Kahneman, 2002). We often regard ourselves as more competent than others (Alicke & Govorun, 2005) and better than what is justified by our actual performance (Lichtenstein & Fischhoff, 1977). Discouraging as these observations might be, people can learn and improve at most cognitive tasks. The learning occurs not only at the task level — we *can* get better at playing the piano — but also at the meta-level — we become more attuned to judging just how good we are at playing the piano. These examples illustrate several aspects of what is known in the cognitive and developmental literatures as *metacognition* or the knowledge people possess about their own cognitive abilities, including beliefs about their own performance in specific tasks.

Some psychologists have argued that metacognitive judgment is secondary to actual performance (Fischhoff, Slovic, & Lichtenstein, 1977; Kruger & Dunning, 1999). According to this view, the knowledge and skills required to *perform* tasks in a certain domain are also required for *judging* one's ability in that domain. One corollary of this view is that metacognitive judgment is domain-specific: a person who is good at remembering things and bad at understanding how others feel will be relatively aware of his good memory skills but relatively unaware of his poor empathy skills. Despite its intuitive appeal, this view has been challenged as confounding metacognitive judgment with level of performance. Because subjective estimates of performance are never perfectly correlated with actual performance, judgments of performance will regress toward the mean: participants who perform worst in a given domain will tend to overestimate their ability in that domain, while those performing best will tend to underestimate. In other words, the relation between actual performance and subjective judgment may reflect a statistical artifact rather than a genuine relation between cognitive and meta-cognitive levels (Burson, Larrick, & Klayman, 2006; Krueger & Mueller, 2002).

Although biased beliefs about one's own skills are the norm, such biases are often exacerbated in psychiatric and neurological disorders (Vuilleumier, 2004). Denial of deficit, reduced self-awareness, loss of insight, and anosognosia are all clinical terms that broadly refer to the same phenomenon: a patient's distorted assessment of her own skills. These patient populations provide

an excellent opportunity to test the domain-specificity of metacognitive judgment, as certain abilities become severely affected by the disease while others remain relatively unaffected. Thus, it is possible to ask whether metacognitive judgments become distorted only in the affected domains or the distortion generalizes to judgments in spared domains. The answer to this question so far has been mixed.

Consistent with the domain-specificity hypothesis, some studies have reported cases of patients with hemiplegia who deny their paralysis but acknowledge limitations in other domains (Marcel, Tegner, & Nimmo-Smith, 2004). There is also evidence that failures of awareness cluster around symptoms. For example, in a study of denial in Dementia of Alzheimer's Type (DAT), a factor analysis of questionnaire data revealed two independent factors: a "cognitive denial" related to length of disease and severity of cognitive deficits, and a "behavioral denial" related to behavioral disinhibition and inappropriate emotional displays (Starkstein, Sabe, Chemerinski, Jason, & Leiguarda, 1996). In contrast, other studies have reported findings consistent with a domain-general deficit, showing that metacognitive errors extend beyond the specific domain of impairment. For example, patients with probable DAT have been reported to overestimate their performance not only in memory but also in visuospatial tasks (Barrett, Eslinger, Ballentine, & Heilman, 2005). Furthermore, denial in DAT seems to correlate more strongly with frontal lobe deficit than with memory problems, despite the latter being more typical of DAT (Michon, Deweer, Pillon, Agid, & Dubois, 1994).

Although *denial of deficit* is prevalent in the clinical setting and may help us understand over-optimism in healthy adults, until not long ago research on denial of deficit had proceeded independently from the experimentally based literature on metacognitive judgment. Recently, the clinical and experimental traditions have become better integrated and clinical researchers have begun to rely more heavily on experimental paradigms (Cosentino & Stern, 2005; Moulin, Perfect, & Jones, 2000; O'Keeffe, Dockree, Moloney, Carton, & Robertson, 2007). Such experimental designs are useful for comparing patients' judgments of their performance to their actual performance on the task. Furthermore, by asking the patient to predict performance *before* the task and to estimate it *after* the task, it is possible to assess metacognitive monitoring (i.e., whether metacognitive judgment improves after experiencing the task). Experimental designs can be applied to domains other than the one of primary clinical concern. This feature becomes important when testing the domain generality of metacognitive judgment.

The patient populations in our study consisted of patients with a behavioral variant of Frontotemporal De-

mentia (FTD-b) and patients with early Dementia of Alzheimer Type (DAT). Although FTD-b has an insidious onset and a gradual progression, its clinical presentation bears close resemblance to cases of orbitofrontal damage caused by traumatic brain injury, such as the famous case of Phineas Gage (Grossman, 2002). 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. They often exhibit inappropriate social behavior, changes in personality and poor decision making (Rankin, Baldwin, Pace-Savitsky, Kramer, & Miller, 2005). At early stages of the disease, many FTD-b patients deny having difficulties or seem unconcerned about them (Eslinger et al., 2005). Such a denial has been documented in the social and emotional domains but needs to be explored further in domains of relatively spared performance. In contrast to FTD-b, patients in early stages of DAT have their social skills relatively spared. The deficit in DAT is primarily of episodic memory (Petersen et al., 1999). At early stages, DAT patients tend to be aware of their deficit, but this awareness declines as the disease progresses (Starkstein, Jorge, Mizrahi, & Robinson, 2006).

After describing FTD-b's *denial of deficit* in semi-structured interviews, we report two experiments that examined whether FTD-b patients overestimate their performance relative to DAT patients and healthy adults. Pre- and post-test judgments of performance were obtained for an attention task (Experiment 1: Stroop Task) and a perception task (Experiment 2: Change Blindness Task). By testing metacognition in two different domains (attention, perception), neither of which is prototypically impaired in early dementia, we were able to test the domain-generality of metacognitive judgments.

2   General procedure

2.1   Participants

Participants 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. Only patients with mild dementia were selected, based on a cut-off score of 24 in the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975).

Ten patients with clinical diagnosis of Frontotemporal Dementia (FTD-b), 12 patients with clinical diagnosis of early Dementia of Alzheimer's Type (DAT), and 14 age-matched normal controls participated in the study. All the FTD-b patients met Lund-Manchester criteria (Neary et al., 1998) and the criterion for behavioral variant of FTD established by the work group on frontotemporal dementia and Pick's disease (McKhann

Table 1: Demographic, neuropsychiatric, and neuropsychological information.

	Max Score	Healthy Elderly	DAT	FTD
Age		$63.2 (9.8)^b$	72.3 (6.3)	$61.0 (7.5)^c$
Years of education		16.2(3.9)	16.0 (3.9)	16.3 (3.3)
Sex: male-female ratio		7/7	9/3	7/3
Handedness (right/left)		11/1	10/2	6/4
FBI		n/a	19.3 (6.7)	$39.6(7.4)^c$
MMSE	30	$28.9 (0.8)^{a,b}$	26.7 (2.1)	26 (1.7)
WAB comprehension	10	$10.0 (0.0)^a$	10.0 (0.0)	$9.8 (0.4)^c$
WAB fluency	10	$10.0 (0.0)^{a,b}$	9.5(.5)	9.1(.4)
Boston Naming	30	$27.7 (1.8)^a$	25.4 (3.9)	23.3 (6.0)
Semantic Fluency		$20.1 (5)^{a,b}$	15.9 (4)	$10.9 (5)^c$
Verbal Fluency (FAS)		$50.5 (13)^{a,b}$	35.3 (14)	$21.3 (11)^c$
Forward digit span	12	$9.8 (1.7)^a$	10.2 (1.4)	$7.8 (2.0)^c$
Backward digit span	12	$8.6 (1.9)^a$	7.3 (2.3)	$4.9 (2.7)^c$
Trails A	240	38.4 (16)	41.5 (13)	53.4 (28)
Trails B	240	$74.1 (24)^{a,b}$	116 (60)	138 (65)
CVLT Acquisition[1]	80	$47.4 (7)^{a,b}$	34.4(10)	32.3(11)
CVLT Delayed Free Recall[1]	16	$10.4 (3)^{a,b}$	3.4 (4)	$5.1 (3)^c$

Note: MMSE, Mini-Mental State Examination; WAB, Western Aphasia Battery; CVLT, California Verbal Learning Test. All group comparisons were made using Mann-Whitney U, significance at p < .05.

[a] Healthy elderly significantly different from FTD-b.

[b] Healthy elderly significantly different from DAT.

[c] DAT significantly different from FTD-b.

[1] Data from one other FTD patient were unavailable for this task.

et al., 2001). Six of the patients in the DAT group met criterion for probable early 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 (NICNCDS-ADRDA) (McKhann et al., 1984). The other six patients in this group met criterion for Amnestic Mild Cognitive Impairment (MCI-a), the prodromal phase of Alzheimer's disease in which cognitive deficits are limited to episodic memory (Petersen et al., 1999). The DAT group was slightly older than the FTD-b group, consistent with FTD-b being a pre-senile dementia, $t(20) = 3.8$, $p = .001$ (see Table 1). To rule out contributions from other pathologies, MRI was performed with a 1.5 Tesla GE Signa scanner using a standard protocol (Callen, Black, Gao, Caldwell, & Szalai, 2001). Apart from atrophy consistent with dementia, the scans showed no other pathology.

2.2 Neuropsychiatric and neuropsychological assessment

Behavioral symptoms were assessed in the FTD-b group with the *Frontal Behavioral Inventory* (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 good reliability in discriminating FTD-b from other dementias. For all but one of the FTD-b patients, the questionnaire was completed with the assistance of the patient's caregiver. Consistent with the clinical diagnosis of FTD-b, all of these patients had abnormal scores (Cutoff: 30; 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
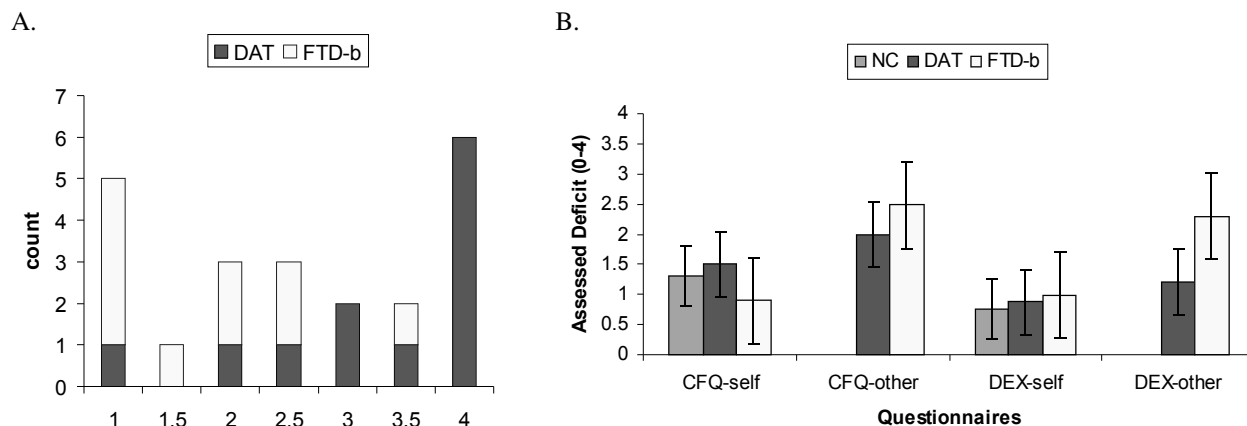
A.

B.

Figure 1: A. Histogram showing the number of patients in each group for each Clinical Interview score (1 = completely unaware, 2= somewhat unaware; 3 = somewhat aware; 4 = completely aware). B. Total scores for Cognitive Failure Questionnaire (CFQ) and Dysexecutive Questionnaire (DEX) for each group, as a function of respondent type (self, caregiver) ($\pm$ 1 SD).

was elevated at 23 but below the cutoff. All of the DAT patients scored within normal range (range: 7–25).[1]

All three groups (FTD-b, DAT, healthy elderly) completed a neuropsychological assessment that probed language, memory, executive function, and visuo-motor skills (Benton, Hemsher, Varney, & Spreen, 1983; Delis, Kramer, Kaplan, & Ober, 1987; Kaplan & Weintraub, 1982; Kertesz, 1982; Weschler, 1987). Both patient groups were impaired relative to the normal controls in most domains. As expected, DAT patients were most impaired in episodic memory while FTD-b patients were most impaired in executive functions (see Table 1).

2.3   Clinical interview on perceived deficits

Patients' awareness of deficit was documented in a semistructured interview, which usually took place at the hospital before a clinical appointment the patient had with his/her neurologist. The interviews were broadly based on the self-awareness of deficit interview, a semistructured interview for testing awareness in Traumatic Brain Injury (Simmond & Fleming, 2003).

After commenting to the patient that Dr. Z was his/her doctor, the interviewer asked about the reasons that had brought the patient to Dr. Z's office and to have Dr. Z as a doctor. This question offered the patient a first opportunity to acknowledge his/her disease. Next, the interviewer asked the patient whether s/he had noticed any changes in his/her abilities over the last few years. If the patient acknowledged any deficit the interviewer asked whether the patient himself/herself had noticed the

deficit, or other people had mentioned it to him/her. If the patient did not acknowledge any deficit the interviewer would ask the patient whether friends and relatives had voiced to him/her any concerns about it. Following, the interviewer probed more specifically for possible problems in memory, concentration, personality, decision making, and social interactions. Finally, patients were encouraged to mention any other difficulties.

At a later time, the interviews were transcribed verbatim except for information that would have compromised patients' anonymity (e.g., names). To assess how readily FTD-b and DAT patients acknowledged their deficit, we asked two coders to rate the transcribed interviews for whether the patient "seemed aware of *currently* having a deficit in his/her mental capacities." Coders were instructed that "mental deficits included cognitive deficits (e.g., forgetting appointments) as well as socio-emotional deficits (e.g., depression, increased irritability for no apparent reason)." Coders were instructed to ignore physical problems (e.g., back pain, appendicitis) and practical problems (e.g., lack of mobility due to not having a driver's license) when judging whether patients acknowledged a deficit. The interviews were organized in alphabetical order (one coder read 1 to 22, the other read 22 to 1). Coders first read all 22 interviews one time through. Next they score each interview in a 4-point Likert scale (completely unaware: 1, completely aware: 4). Coders worked independently. Coders were undergraduate students majoring in psychology and were blind to the purpose of the study and to the specific diagnoses. The agreement between the coders was excellent, as the intraclass correlation was $r = .79$ (Cicchetti, 1994).

On a 4 point-scale, FTD-b patients were coded as having significantly more denial than DAT patients as re-

---

[1]We did not collect data on the FBI for the five MCI-a patients in the DAT group who were independent and came to the clinic visit alone.

vealed by a non parametric Mann-Whitney test, $U = 16.5$, $Z = 2.9$, $p = .003$, ($M_{FTD-b} = 1.8$; $M_{DAT} = 3.2$) (see distribution in Figure 1A).[2] Most FTD-b patients were reluctant to acknowledge their current deficits while the vast majority of DAT patients readily acknowledged their deficit. The assessment of the coders was also consistent with the discourse characteristics of the interviews. For example, when asked about their reason for seeing Dr. Z, it was common for FTD-b patients to answer by referring to the opinion of somebody else (e.g., "my wife thinks that I have problems"). This type of utterances indicates that the speaker does not agree with the assessment (Malle & Pearce, 2001).

2.4 Questionnaire assessment about perceived deficit in everyday performance

Patients' awareness of cognitive deficit was further assessed with self- and caregiver-reports on the Cognitive Failure Questionnaire (CFQ) and the Dysexecutive Questionnaire (DEX) (Broadbent, Cooper, FitzGerald, & Parkes, 1982; Burgess, Alderman, Evans, Emslie, & Wilson, 1998). The DEX has 20 items that probe deficits typical of frontal lobe damage, and can be divided into 5 subscales: emotional bluntness, emotional instability, insight, impulsivity, and cognition. The informant version has the same 20 items as the self-report form. Answers ranged from 0 (never) to 4 (very often) in both scales. The CFQ is a 25-item self-report of how frequently certain lapses in everyday tasks had occurred over the last 6 months. The items can be grouped into five subscales: general memory, memory for specific names, distractability, social blunders, and motor lapses (Wallace, Kass, & Stanny, 2002). The informant version of the CFQ has 8 items probing aspects of cognitive failure that are apparent from third-person's view (e.g., "forgetful, such as forgetting where she put things"). We did not collect caregiver reports for the five MCI-a patients who were independent and came to the clinic visit alone.

Data were not normally distributed; therefore we ran non-parametric Mann-Whitney tests for independent groups, comparing FTD-b to DAT (see Figure 1b). Total scores of the CFQ self-reports revealed less awareness of deficit in FTD-b than in DAT, $U = 22$, $Z = 2$, $p = .04$. Follow-up comparisons for each of the five CFQ subscales revealed that FTD-b patients significantly downplayed their social blunders relative to DAT patients, $U = 19$, $Z = 2.3$, $p = .02$. FTD-b patients also reported fewer

memory problems than DAT patients, $U = 21$, $Z = 2.1$, $p = .04$, but this difference in metamemory cannot be interpreted as a metacognitive error as it might have stemmed from group differences in actual memory performance.

Caregivers for FTD-b patients reported more deficits for these patients than caregivers for DAT in the DEX questionnaire, $U = 8.5$, $Z = 2.6$, $p < .01$. Follow-up comparisons for each of the five subscales revealed that relative to caregivers of DAT patients, caregivers of FTD-b patients reported FTD-b patients as having reduced insight, increased emotional bluntness and increased impulsivity, $Zs > 2.2$, $ps < .05$.

3 Experiment 1. Stroop task

In Experiment 1, participants were asked to make metacognitive judgments before and after participating in a computerized version of the Stroop task. Participants had to report the color in which a word was displayed while ignoring the meaning of the word. For trials in which meaning and color were in conflict (e.g., the word RED in green ink), answering correctly ("green") required the inhibition of the prepotent response ("red").

3.1 Method

Equipment. Stimuli were displayed on a 14-inch laptop monitor set to a screen resolution of 1024 x 768 pixels, with true color (32 bits). The laptop was a Dell Inspiron 3800 equipped with a Pentium III processor and Windows 98. The timing of the stimulus display and data collection were managed using E-prime, a commercial experiment application. Data were collected via a microphone and response box.

Stimuli. Congruent and incongruent stimuli were created by the combination of four words ("red" "blue" "green" and "yellow") and font colors (red, blue, green, yellow). Neutral stimuli were created by the combination of four words (bad, poor, deep, legal) and four colors (red, blue, green, yellow). Words were displayed in Courier New 36 point against a black background.

Procedure. Immediately after receiving the task instructions, participants completed 8 trials, each one displaying a string of Xs in one of four possible colors. The participant reported the color ink verbally. Next, three trials were shown in fixed order to illustrate the three conditions: first a neutral trial (word "poor" in green ink), next an incongruent trial (word "blue" in red ink), and finally a congruent trial (word "yellow" in yellow ink). For these three illustration trials, participants were instructed to *not* make an overt response and instead wait

---

[2]Interestingly, the only FTD-b patient who was aware of his deficit was also the only patient with signs of non-fluent aphasia. That patient was also the only FTD-b patient who correctly predicted his performance in attention and perceptual task to be below average. Futures studies should explore whether speech difficulty plays a preferential role in informing the self.

for the experimenter, who provided the correct response. These instructions aimed to minimize the use of experiential cues in these illustration trials for predicting future performance.

Pre-test metacognitive judgment. After illustration, participants were asked to predict their performance (accuracy, speed) relative to other people of their age, and also relative to young adults. For example, one question asked: "of 100 people of your age who perform this task, you think you will be MORE ACCURATE than:
nobody   half   everybody
 10   20   30   40   50   60   70   80   90   100"
Questions were presented in a fixed order, with accuracy questions first, and the age matched comparison group preceding the comparison to young adults.

Actual task. After predicting their performance, participants completed a practice block of 14 practice trials and a test block of 104 test trials. The test block lasted approximately 8 minutes with equal number of congruent, incongruent, and neutral trials displayed in random order. Each word was displayed until a verbal response was made or for up to 10 seconds. The response-target interval was approximately 2.700 ms.

Post-test metacognitive judgment. Participants completed a post-test questionnaire to provide judgment of their past performance (as already described for the pretest).

3.2 Results and discussion

3.2.1 Metacognitive judgment

An analysis of variance included Group (healthy elderly, FTD-b, DAT) as a between-subjects factor, and three within-subject factors: Assessment Time (pre, post-task), Type of Performance Assessed (accuracy, speed), and Hypothetical Comparison Population (same age, young). The dependent variable was each participant's judgment of where s/he would rank in a pool of 100 participants (higher numbers meaning better/faster performance). The results are shown in Table 2.

The analysis revealed several findings of interest. First, self-assessment of performance differed across groups (FTD-b: 73.1; Healthy Elderly: 53.9; DAT: 45.5) as revealed by a Group main effect $F(2,33) = 10.2$, $p = .001$, $MSE = 1681$. Post-hoc pair-wise comparisons using Tukey HSD tests revealed that the FTD-b group assessed their performance higher than the DAT and the healthy elderly groups. Second, experience with the task did not change participants' judgment of their performance (Pretest: 58.2; Post-test: 56.7) as time of assessment did not

Table 2: Self-assessed performance in the Stroop task as a function of comparison group and experience with the task. Standard deviations appear in parentheses.

	Assessed performance relative to:			
	Age peers		Young adults	
	before task	after task	before task	after task
Accuracy				
NC	75 (14)	72 (11)	41 (23)	41 (21)
DAT	56 (17)	59 (12)	43 (25)	38 (22)
FTD-b	81 (15)	77 (17)	66(31)	71 (21)
Speed				
NC	68 (14)	64 (15)	36(18)	34 (17)
DAT	55 (12)	54 (11)	31 (20)	28 (14)
FTD-b	77 (20)	75 (18)	70 (26)	68 (27)

Note. NC, normal control; DAT, Dementia of Alzheimer's type; FTD-b: behavioral variant of Frontotemporal Dementia

have a significant main effect, $F(1,33) = .8$, *ns*, $MSE = 170$, nor did it interact with any other variable.

Although participants did not change their assessment after experiencing the task, participants did modify their assessment based on other contextual cues. For example, participants were more modest when assessing their speed than when assessing their accuracy, $F(1,33) = 9.8$, $p = .005$, $MSE = 168$, consistent with elderly participants' emphasis on avoiding errors during speeded tasks. Also as expected, participants lowered their expectations when comparing their performance to young people (in comparison to age peers: 67.8; in comparison to young adults: 47.2), $F(1,33) = 58.8$, $p < .001$, $MSE = 510$. This flexibility in judgment differed across groups, as revealed by a Group by Hypothetical Comparison interaction, $F(2,33) = 6.0$, $p < .01$, $MSE = 510$. Follow-up analyses contrasting each group pair revealed that when comparing their performance to a population of young adults, FTD-b patients lowered their assessment less than either healthy elderly or DAT patients, $F(1,22) = 11.4$, $p = .005$, $MSE = 533$; $F(1,20) = 4.2$, $p = .05$, $MSE = 397$.

3.2.2 Accuracy

Error percentages for each participant in each condition were submitted to a 3 x 3 mixed analysis of variance that had Group (DAT, FTD-b, Healthy Elderly) as a between-subjects factor and Trial Type (congruent, incongruent, neutral) as a within-subject factor (see Table 3).

There was a main effect of Group, $F(2,33) = 3.6$, $p = .04$. Post-hoc pair-wise comparisons using Tukey HSD

Table 3: Error Percentages and Median Reaction Times in the Stroop Task as a function of group and trial type. Standard deviations appear in parentheses.

Group	Congruent	Neutral	Incongruent
Errors (%)			
NC	0.2 (0.8)	0.5 (1.1)	2.3 (2.7)
DAT	0.5 (1.1)	0.5 (1.2)	8.3 (7.7)
FTD-b	0.3 (1.0)	1.5 (2.9)	5.3 (4.1)
RT (ms)			
NC	674 (71)	717(67)	839 (74)
DAT	745 (99)	848 (83)	1044 (171)
FTD-b	924 (253)	1136 (424)	1387 (526)

Note: NC, normal control; DAT, Dementia of Alzheimer's type; FTD-b: behavioral variant of Frontotemporal Dementia.

tests revealed that DAT patients made more errors than healthy elderly participants. Importantly, FTD-b were no better than healthy elderly or DAT patients, despite their elevated assessment of performance.

There was also a main effect of Trial Type, $F(2, 66) = 27.8$, $p = .001$, which was qualified by an interaction with Group, $F(4, 66) = 4.0$, $p = .02$. To explore this interaction, we ran follow-up analyses for each level of Trial Type. These analyses revealed that DAT patients made more errors than healthy adults only in incongruent trials, $F(2, 33) = 4.2$, $p = .05$ (Tukey HSD p < .05) (see Table 3). Error rates did not differ across groups for neutral and congruent trials but this was not surprising because accuracy for those trials was at ceiling.

3.2.3   Reaction time

Error trials and trials immediately following an error were filtered out and from the remaining trials median reaction times in each condition for each participant were calculated (see Table 3). A 3 (Group) by 3 (Trial Type) ANOVA revealed a main effect of Group, $F(2,33) = 10.1$, $p = .001$, $MSE = 143387$. FTD-b patients were significantly slower than healthy elderly and DAT patients, as revealed by Tukey HSD post-hoc comparisons. There was also a main effect of Trial Type $F(2, 66) = 86.1$, $p = .001$. As expected, incongruent trials led to slower responses than congruent and neutral ones. Finally, there was a Group by Trial Type interaction, $F(4, 66) = 6.6$, $p = .001$. To further explore this interaction while at the same time controlling for group differences in overall speed of response, we calculated proportional conflict costs [(incongruent − congruent)/congruent]. We submitted data

on proportional scores to an analysis of variance with Group as the between-subjects factor. This analysis revealed a main effect of Group, $F(2, 33) = 6.2$, $p = .005$ $MSE = 262$. Posthoc comparisons using Tukey HSD tests revealed that DAT and FTD-b patients were impaired relative to the healthy elderly group, while the difference between patient groups was non-significant [NC: 25 (10), DAT: 41 (18), FTD-b: 48 (20)].

In sum, Experiment 1 found that FTD-b patients overestimated both accuracy and speed of performance relative to DAT and healthy elderly groups, despite performing more slowly than the other groups and with as many errors. FTD-b patients were also less likely to lower their assessment when comparing their performance to a hypothetical group of young adults. These findings reveal serious metacognitive errors by FTD-b patients in this executive attention task. Finally, assessed performance after task participation was no different than the initial prediction for any of the groups. However, near ceiling accuracy performance makes this result difficult to interpret. To overcome this limitation, the next experiment used a design in which failures are common.

4   Experiment 2: Change blindness

Experiment 2 assessed judgments of performance in the visual domain with the use of a flicker paradigm. In this paradigm, two versions of a complex scene are presented in alternating sequence, separated by a blank field. The two versions of the scene differ from one another only with respect to a single changing item, and the participant is instructed to look for the change. The change is well above threshold and once it has been detected it is clearly visible, often appearing very "obvious" (see Figure 2). However, it usually takes several seconds for participants to first notice the change, a phenomenon that has been labeled *change blindness* (Simons & Rensink, 2005).

In many ways, the flicker paradigm is ideal for studying metacognitive judgment. First, even healthy adults are known to over-predict performance in this task (Beck, Levin, & Angelone, 2007). Thus, their judgment would allow for a stringent criterion against which to test FTD-b's over-optimism bias. Second, although finding the change is difficult, the change stands out once detected. The belief that finding the change should be easy combined with the feeling of having to struggle to find such a change should lead to reduced assessed performance after participating in the task. Thus, this paradigm is well suited to examine whether FTD-b patients can update their beliefs based on past performance.
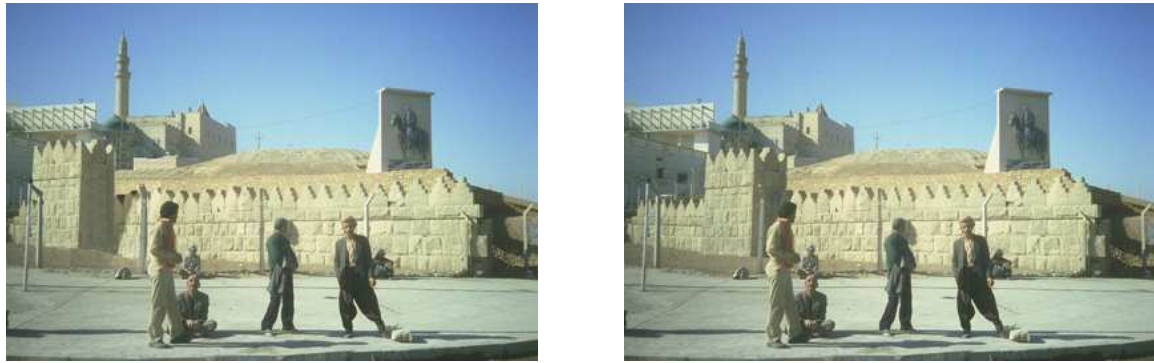
Figure 2: An example of a complex scene and its modified version. The tower in the left changes position. The pictures were displayed in color.

4.1   Method

Participants.   All of the participants from Experiment 1 participated in Experiment 2.

Stimuli.   Twelve complex scenes and their modified versions were borrowed from the Cambridge Basic Research image database. Images 4.5 inches wide by 3 inches tall were displayed in the center of the screen. Half of the modified scenes consisted of a position change and the other half consisted of a color change.

Procedure.   The first two trials served as examples. In these trials, the experimenter pointed to the location where the change would occur and asked the participant to report the change. By telling participants where the change would occur we aimed to minimize the amount of unsuccessful search. Consistent with these instructions, all participants were able to detect the changes in less than 8 flickers. One of the example trials depicted a change in color and the other a change in location. Participants were instructed that in the actual test they would not be told where the change would occur.

Pre-test metacognitive judgment.   After this illustration, participants were asked to predict their performance relative to other people of their age, and relative to young adults. For example, one question asked "of 100 people of your age who perform this task, you think you will find the changes FASTER than:

nobody   half   everybody
 10   20   30   40   50   60   70   80   90   100"
Questions were presented in a fixed order, with the age matched comparison group preceding the comparison to young adults.

Actual task.   Ten trials were presented in a fixed order. In each trial, the original and modified version of the scene alternated. Each version was displayed for 300 msec separated by a 150 msec blank field. The scene flickered for up to 15 seconds or until the participant pressed a key in the keyboard to report the change. After reporting the change, the participants had to confirm that their identification was veridical by pointing to the location of change, something participants did with no problem. If 15 seconds went by without the participant reporting a change, the flickering stopped and the experimenter pointed the location and identity of the change. Next, the flickering was restarted for the participant to observe the change at the indicated location.

The sixth trial displayed a large change to the central object in the scene. Pilot data had revealed that this change was relatively easy to detect. Thus, it served to probe whether participants were still engaged in the task and performing as instructed. We also thought it would help to keep morale up and discourage participants from stopping the experiment.

Post-test metacognitive judgment.   Finally, participants completed a post-test questionnaire to provide judgment of their past performance, as already described for the pre-test questionnaire.

4.2   Results and discussion

4.2.1   Metacognitive Judgment

An analysis of variance included Group (healthy elderly, FTD-b, DAT) as a between-subjects factor, and Assessment Time (pre, post-task) and Hypothetical Comparison Population (same age, young) as within-subject factors. The dependent variable was participant's judgment of where s/he would rank in a pool of 100 participants (higher numbers meaning faster detection of change). The results are shown in Table 4.

Self-assessment of performance differed across

Table 4: Self-assessed Performance in the Change Blindness Task as a Function of Comparison Group and Experience with the Task. Standard deviations appear in parentheses.

	Assessed performance relative to:			
	Age peers		Young adults	
	before task	after task	before task	after task
Detection speed				
NC	52 (20)	41 (25)	29 (19)	15 (14)
DAT	51 (7)	31 (17)	29 (15)	18 (19)
FTD-b	69 (19)	54 (27)	61 (28)	42 (27)

Note: NC, normal control; DAT, Dementia of Alzheimer's type; FTD-b: behavioral variant of Frontotemporal Dementia.

Figure 3: Participants' actual performance in the Change blindness task (+ 1 SD). Change detection was well below ceiling levels for all groups.

groups, as revealed by a Group main effect, $F(2,33) = 6.6$, $p = .004$, $MSE = 1182$. Post-hoc pair-wise comparisons using Tukey HSD tests revealed that the FTD-b group assessed performance higher than the DAT and the healthy elderly groups (FTD-b: 56.5 ; DAT = 32.2; Healthy Elderly = 34.5).

Experience with the change blindness task did change participants' judgment of their performance, as revealed by a main effect of Assessment Time (Pre-test: 48.6; Post test: 33.6), $F(1,33) = 32.8$, $p = .001$, $MSE = 243$. Participants also modified their assessment based on the Comparison Population, $F(1,33) = 60.1$, $p = .001$, $MSE = 176$, (in comparison to age peers = 49.8; in comparison to young adults = 32.3) although this flexibility differed across groups [Group by Hypothetical Comparison interaction: $F(2,33) = 3.6$, $p = .04$, $MSE = 176$]. Follow-up analyses contrasting each group pair revealed that FTD-b patients lowered their assessment less than either healthy elderly, $F(1,22) = 7.8$, $p = .01$, $MSE = 1455$, or the DAT patients, $F(1,20) = 9.8$, $p = .005$, $MSE = 1314$.

4.2.2 Actual performance

If 15 seconds went by without the subject detecting the change, the trial ended and the change was coded as undetected. Percentage of changes detected served as a proxy for how fast participants were at detecting changes (i.e., how many changes they detected before the 15 sec cutoff). All participants except one detected at least one change (10%) and no one detected more than 8 changes (80%). Healthy elderly participants detected 48% of changes (Standard deviation: 20), DAT detected 31% (14), and FTD-b patients detected 20% (14). A one-way ANOVA on these data revealed a main effect of Group,
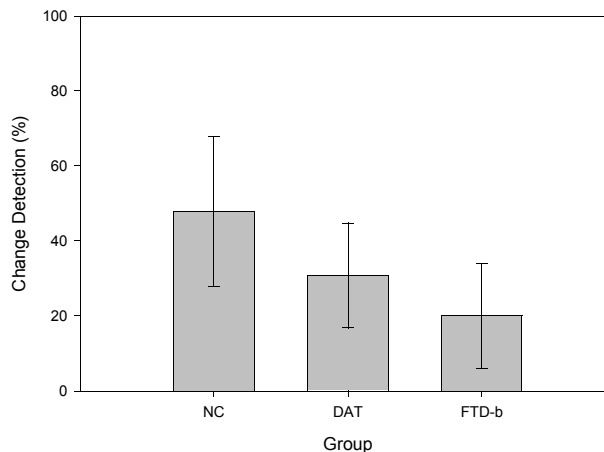
$F(2,33) = 8.4$, $p = .001$, MSE= 2.8. Post-hoc pair-wise comparisons using Tukey HSD tests revealed that the healthy elderly group detected more changes than either the FTD-b group or the DAT group. Finally, detection rates were higher for the 6th trial than for the other trials (healthy: 93%; DAT: 66%; FTD-b: 70%), as revealed by one-sample t-tests against these values (ps < .001). This suggests that participants were engaged in the task and willing to comply with the instructions, at least up to that trial.[3]

In sum, the findings from the change blindness task replicated the major findings of the Stroop task. FTD-b patients assessed their performance more optimistically than both healthy adults and DAT patients despite performing worse than healthy adults and no better than DAT patients. Performance in the change blindness task led all groups to adopt more humble assessments of their performance suggesting that strong and unambiguous feedback may be used by patients with FTD-b to update their beliefs.

5   General discussion

A variety of measures were used in this study to examine metacognitive function and impairment in FTD-b and DAT. The questionnaires and interviews were indicative of denial of deficit in FTD-b and relatively spared awareness in early stages of DAT: While caregivers of patients with FTD-b reported increased emotional bluntness and

---

[3]We also explored how many flickers it took for participants to detect the change in this "easy" trial. The healthy elderly detected the change in fewer cycles than FTD-b patients, as revealed by an independent sample t-test, t(18) = 2.2, p = .04 [Healthy elderly: 9,6 (Standard deviation: 3.6); DAT: 11.8 (4.4); FTD-b: 13.7 (4.8)].

impulsivity in these patients, the patients had a tendency to downplay those problems. These results are consistent with a previous study which measured patient- and caregiver-reports of everyday cognitive, social, and emotional behaviors and found large denial of deficit in the FTD-b patients relative to DAT patients (Eslinger et al., 2005).

The two experiments showed a presence of metacognitive errors in two domains (attention, perception) different from the ones most impaired in FTD-b (social, emotional), a result that is consistent with a domain-general metacognitive bias. In other words, we found no evidence that FTD-b's errors were limited to the social domain. Although we aimed to equate actual performance across the two patient groups, slight differences in actual performance remained (e.g., FTD-b's slower reaction time in the Stroop task). Thus, we cannot rule out the possibility that FTD-b's poor cognitive and meta-cognitive performance had a common substrate (Kruger & Dunning, 1999), as a strict rejection of the domain-specific hypothesis would require FTD-b patients to have performed as well as DAT patients in all measures of attention and perception. On the other hand, it seems unlikely that the relatively subtle differences in actual performance across clinical groups could account for the quite large differences in metacognitive judgment.

The finding that metacognitive deficits in FTD-b extend beyond the socio-emotional domain is also consistent with the findings from two studies of metamemory in FTD-b. The first one compared a group of 6 FTD-b patients and a group of DAT patients (Souchay, Isingrini, Pillon, & Gil, 2003). Participants studied 20 word pairs in preparation for a cued recall task. Before and after the learning phase, subjects predicted how many words they would recall in the test phase. Both patient groups were overoptimistic in their predictions relative to the number of words they actually recalled but the discrepancy between prediction and actual performance was largest for the FTD-b group. The second study explored a more varied group of FTD patients that included not only patients with behavioral disorders (FTD-b) but also patients with aphasia (primary progressive aphasia and semantic dementia patients) (Eslinger et al., 2005). Patients estimated their past performance in three standardized tests (verbal fluency, word list learning, and verbal memory). Those estimates were positively related to patients' actual performance. However, the correlations were weaker for the FTD-b subgroup, hinting at possible metamemory deficits in the behavioral variant of the disease.

The results from our experiments also confirm that group differences in metacognitive judgment cannot be discounted as merely statistical artifact. Certainly, a regression-to-the-mean account can help explain gaps in judgment relative to actual performance: a patient whose performance is the worst in the group could only show a bias in judgment away from her extremely poor performance. Therefore, the gap between judged and actual performance will be greater for those who perform worst. But the regression-to-the-mean account does not predict that those who perform worst will rank their performance higher than those who perform better. In other words, regression to the mean may explain why an FTD-b patient thinks she performed better than her true performance, but cannot explain why that FTD-b patient will rank her performance higher than a DAT patient whose performance was as good, if not better than hers. Such a result, as our study indicates, reveals a true impairment in the metacognitive judgment of FTD-b patients.

Having established that FTD-b patients truly overestimate their abilities, we must explore its possible root mechanisms. At this time, we can provide only some tentative ideas. The first one has to do with which domains are affected in the disease. Early stages of DAT are characterized by memory deficits while early stages of FTD-b show mostly a socio-emotional deficit. If memory errors are more salient to the self than socio-emotional errors, DAT patients should indeed be more aware of deficits than FTD-b. According to this argument, FTD-b patients are not in denial *per se* but rather they suffer a type of deficit that often goes undetected by the self (Allison, Messick, & Goethals, 1989).

The situation for informant-reports is likely the reverse, with socio-emotional deficits outweighing memory deficits in the caregiver's assessment of a patient's deficit. Socio-emotional deficits are more salient and more unsettling for caregivers than cognitive deficits, leading to increased caregiver stress. This hypothesis is consistent with the patterns of self-report and informant-based questionnaires as well as with clinical interviews. On the other hand, the results from our experiments clearly show that judgment errors in FTD-b extend well beyond the socio-emotional domain. They extend to areas of perception and cognition.

Of course, this not to say that emotion plays no role in denial of deficit. It remains a likely possibility that deficits in emotion will contribute to denial of deficit in cognitive tasks. Error detection, which acts as an important cue for judgments of performance, is as much an emotional process as a cognitive one (Bush, Luu, & Posner, 2000).[4] Emotions may also play a role in monitoring judgments of performance. In normal subjects, the tendency to overestimate one's own abilities is often tempered by the prospect of making an embarrassingly immodest prediction. Given the emotional bluntness characteristic of FTD-b patients and their inability to

---

[4]Indeed, we had hoped to assess group differences in reaction times following errors in the Stroop task in Experiment 1, but unfortunately patients did not make enough errors to yield meaningful comparisons.

feel embarrassment, it seems likely that those checks and balances would be absent when self-judging their performance. Although the neuroanatomy of denial of deficit is not yet well understood, some studies suggest involvement of areas that are also involved in emotional regulation, such as the orbitofrontal cortex and the right frontal lobe (Mendez & Shapira, 2005; Salmon et al., 2006).

In the current study, participants were asked to provide judgments of performance only at the beginning and at the end of the task. We chose this design instead of one with a judgment after every trial because we worried about perseveration tendencies in FTD-b. We were also concerned that trial-by-trial judgments might introduce spurious group differences in actual task performance. Nonetheless, future studies should include trial-by-trial judgments to assess whether FTD-b patients are sensitive to item difficulty. In this regard, metamemory studies in DAT may prove informative: DAT patients tend to overestimate their memory performance (i.e., a calibration bias) but their sensitivity to item difficulty seems relatively unimpaired: when presented with words of varying recallability, DAT patients correctly rate the highly recallable words as being more likely to be recalled and allocate less study time to them (Moulin, Perfect, & Jones, 2000). These results suggest that metamemory monitoring is relatively intact in AD. It would be interesting to assess whether the same is true in FTD-b.

References

Alicke, M. D., & Govorun, O. (2005). The Better-Than-Average Effect. In M. D. Alicke, D. A. Dunning & J. I. Krueger (Eds.), *The Self in Social Judgment* (pp. 85–106). New York, NY: Psychology Press.

Allison, S. T., Messick, G. M., & Goethals, D. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition, 7*, 275–295.

Barrett, A. M., Eslinger, P. J., Ballentine, N. H., & Heilman, K. M. (2005). Unawareness of cognitive deficit (cognitive anosognosia) in probable AD and control subjects. *Neurology, 64*, 693–699.

Beck, M. R., Levin, D. T., & Angelone, B. (2007). Change blindness blindness: Beliefs about the roles of intention and scene complexity in change detection. *Consciousness and Cognition: An International Journal, 16*, 31–51.

Benton, A. L., Hemsher, K. S., Varney, N., & Spreen, O. (1983). Contributions to neuropsychological assessment: A clinical manual. In (pp. 30–43). Oxford: Oxford University Press.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *The British journal of clinical psychology / the British Psychological Society, 21 (Pt 1)*, 1–16.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*, 547–558.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*, 60–77.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex.*Psychological Science, 4*, 215–222.

Callen, D. J., Black, S. E., Gao, F., Caldwell, C. B., & Szalai, J. P. (2001). Beyond the hippocampus: MRI volumetry confirms widespread limbic atrophy in AD. *Neurology., 57*, 1669–1674.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.

Cosentino, S., & Stern, Y. (2005). Metacognitive theory and assessment in dementia: do we recognize our areas of weakness? *Journal of the International Neuropsychological Society, 11*, 910–919.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). *The California Verbal Learning Test*. New York: Psychological Corp.

Eslinger, P. J., Dennis, K., Moore, P., Antani, S., Hauck, R., & Grossman, M. (2005). Metacognitive deficits in frontotemporal dementia. *Journal of Neurology, Neurosurgery, and Psychiatry, 76*, 1630–1635.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 552–564.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini-mental state': A pratical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 185–198.

Gilovich, T., Griffin, D. W., & Kahneman, D. (2002). *Heuristics of judgment: Extensions and applications*. New York: Cambridge University Press.

Grossman, M. (2002). Frontotemporal dementia: A review. *Journal of the International Neuropsychological Society, 8*, 566–583.

Kaplan, E. G., H., & Weintraub, S. (1982). *Boston Naming Test*. Philadelphia: Lea & Febiger.

Kertesz, A. (1982). *Western Aphasia Battery*. New York: Grune & Stratton.

Kertesz, A., Nadkarni, N., Davidson, W., & Thomas, A. W. (2000). The Frontal Behavioral Inventory in the differential diagnosis of frontotemporal dementia. *Jour-*

nal of the International Neuropsychological Society, 6, 460–468.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*, 180–188.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*, 159–183.

Malle, B. F., & Pearce, G. E. (2001). Attention to behavioral events during interaction: Two actor-observer gaps and three attempts to close them. *Journal of Personality and Social Psychology, 81*, 278–294.

Marcel, A. J., Tegner, R., & Nimmo-Smith, I. (2004). Anosognosia for plegia: specificity, extension, partiality and disunity of bodily unawareness. *Cortex, 40*, 19–40.

McKhann, G., Drachman, D., Folstein, M., 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.

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.

Mendez, M. F., & Shapira, J. S. (2005). Loss of insight and functional neuroimaging in frontotemporal dementia. *The Journal of Neuropsychiatry and Clinical Neurosciences, 17*, 413–416.

Michon, A., Deweer, B., Pillon, B., Agid, Y., & Dubois, B. (1994). Relation of anosognosia to frontal lobe dysfunction in Alzheimer's disease. *Journal of neurology, neurosurgery, and psychiatry, 57*, 805–809.

Moulin, C. J., Perfect, T. J., & Jones, R. W. (2000). Evidence for intact memory monitoring in Alzheimer's disease: metamemory sensitivity at encoding. *Neuropsychologia, 38*, 1242–1250.

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.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol.

26, pp. 125–173). New York: Academic Press.

O'Keeffe, F., Dockree, P., Moloney, P., Carton, S., & Robertson, I. H. (2007). Awareness of deficits in traumatic brain injury: A multidimensional approach to assessing metacognitive knowledge and online-awareness. *Journal of the International Neuropsychological Society, 13*, 38–49.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of neurology, 56*, 303–308.

Rankin, K. P., Baldwin, E., Pace-Savitsky, C., Kramer, J. H., & Miller, B. L. (2005). Self awareness and personality change in dementia. *Journal of Neurology, Neurosurgery, and Psychiatry, 76*, 632–639.

Salmon, E., Perani, D., Herholz, K., Marique, P., Kalbe, E., Holthoff, V., et al. (2006). Neural correlates of anosognosia for cognitive impairment in Alzheimer's disease. *Human Brain Mapping, 27*, 588–597.

Simmond, M., & Fleming, J. (2003). Reliability of the self-awareness of deficits interview for adults with traumatic brain injury. *Brain Injury, 17*, 325–337.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9*, 16–20.

Souchay, C., Isingrini, M., Pillon, B., & Gil, R. (2003). Metamemory accuracy in Alzheimer's disease and frontotemporal lobe dementia. *Neurocase, 9*, 482–492.

Starkstein, S. E., Jorge, R., Mizrahi, R., & Robinson, R. G. (2006). A diagnostic formulation for anosognosia in Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 77*, 719–725.

Starkstein, S. E., Sabe, L., Chemerinski, E., Jason, L., & Leiguarda, R. (1996). Two domains of anosognosia in Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 61*, 485–490.

Vuilleumier, P. (2004). Anosognosia: the neurology of beliefs and uncertainties. *Cortex, 40*, 9–17.

Wallace, J. C., Kass, S. J., & Stanny, C. J. (2002). The cognitive failures questionnaire revisited: dimensions and correlates. *The Journal of General Psychology, 129*, 238–256.

Weschler, D. A. (1987). *Weschler memory scale -Revised.* San Antonio, TX: Psychological Corporation.