Published in final edited form as:

Biol Psychiatry. 2011 April 15; 69(8): 726–733. doi:10.1016/j.biopsych.2010.12.041.

Remission Prognosis for Cognitive Therapy for Recurrent Depression Using the Pupil: Utility and Neural Correlates

Greg J. Siegle, Ph.D.¹, Stuart R. Steinhauer, Ph.D.^{1,2}, Edward S. Friedman, M.D.¹, Wesley S. Thompson³, and Michael E. Thase, M.D.^{1,4}

¹University of Pittsburgh, School of Medicine

²VA Pittsburgh Healthcare System

³University of California, San Diego

⁴University of Pennsylvania School of Medicine and Philadelphia Veterans Affairs Medical Center

Abstract

Background—Although up to 60% of people with major depressive disorder (MDD) respond to Cognitive Therapy (CT) in controlled trials, clinicians do not routinely use standardized assessments to inform which patients should receive this treatment. Inexpensive non-invasive prognostic indicators could aid in matching patients with appropriate treatments. Pupillary response to emotional information is an excellent candidate, reflecting limbic reactivity and executive control. This study examined 1) whether pre-treatment assessment of pupillary responses to negative information were associated with remission in CT, and 2) their associated brain mechanisms.

Methods—We examined whether pre-treatment pupillary responses to emotional stimuli were prognostic for remission in an inception cohort of 32 unipolar depressed adults to 16–20 sessions

Correspondence concerning this article should be addressed to Greg J. Siegle, Ph.D., Western Psychiatric Institute and Clinic, 3811 O'Hara St. Pittsburgh, PA 15213-2593. gsiegle@pitt.edu.

Department in which the work was done: University of Pittsburgh, School of Medicine, and VA Pittsburgh Healthcare System

Participants were recruited from a trial registered in ClinicalTrials.gov, identifier: NCT00183664. This study and the parent trial were approved by the University of Pittsburgh Institutional Review Board.

Pupillary response data from a subset of the participants in the current study, on different tasks, has been published in other articles (19, 23, 35, 37) but not with regard to therapeutic change. Data on a subset of these participants who were assessed with fMRI following their pupil dilation assessment has also been published (6, 19, 28), as have relationships of rumination to therapeutic response in a superset of these participants (5). The data presented in those publications do not overlap with this manuscript. Portions of this manuscript were presented at the meeting of the Association for Behavioral and Cognitive Therapies (2008), Orlando, FL.

Ethics

This study and the parent trial were approved by the University of Pittsburgh Institutional Review Board (IRB). This study was also approved by the VA Pittsburgh IRB. After being fully informed regarding the nature of the study, all participants gave written consent to participate in the study by signing University of Pittsburgh and, if necessary, VA Pittsburgh IRB approved consent forms.

Disclosures

No authors had conflicts relevant to this manuscript. Greg Siegle is a consultant for Trial IQ. Michael Thase reported having served as an advisor/consultant for AstraZeneca, Bristol-Myers Squibb Company, Cephalon, Inc., Cyberonics, Inc., Eli Lilly & Co., GlaxoSmithKline, Janssen Pharmaceutica, MedAvante, Inc., Neuronetics, Inc., Novartis, Organon, Inc., Sepracor, Inc., Shire US Inc., Supernus Pharmaceuticals, and Wyeth Pharmaceuticals. In addition he reported Speakers Bureau affiliations with AstraZeneca, Bristol-Myers Squibb Company, Cyberonics, Inc., Eli Lilly & Co., GlaxoSmithKline, Organon, Inc., Sanofi Aventis, Wyeth Pharmaceuticals. He has given expert testimony on behalf of Jones Day (Wyeth Litigation) and Phillips Lytle (GlaxoSmithKline Litigation), has equity holdings in MedAvante, Inc., has received royalties from American Psychiatric Publishing, Inc., Guilford Publications, Herald House, and has a family relationship with senior staff at Advogent (Formerly Cardinal Health). Dr. Friedman has provided Speaker Bureaus or Advisory Boards: AstraZeneca, Eli Lilly GlaxoSmithKline, Pfizer Wyeth-Ayerst, and has obtained Grant/Research support from Aspect Medical Systems, Indevus, AstraZeneca, Bristol-Myers Squibb, Pfizer, Sanofi-Aventis, Wyeth-Ayerst, Cyberonics, Novartis, NorthStar/St. Jude Medical, Medtronics, Respireonics

of CT. Twenty patients were then assessed on the same task using fMRI. Pupillary responses were assessed in 51 never-depressed controls for reference.

Results—Remission was associated with either low initial severity or the combination of higher initial severity and low sustained pupil dilation responses to negative words (87% correct classification of remitters and non-remitters (93% sensitivity, 80% specificity); 88% correct classification of high-severity participants, $p < .01$, 90% sensitivity, 92% specificity). Increased pupillary responses were associated with increased activity in dorso-lateral prefrontal regions associated with executive control and emotion regulation.

Conclusions—For patients with higher severity, disruptions of executive control mechanisms responsible for initiating emotion regulation, which are indexed by low sustained pupil responses and targeted in therapy, may be key to remitting in this intervention. These mechanisms can be measured using inexpensive noninvasive psychophysiological assessments.

Keywords

Cognitive Therapy; Remission; Psychotherapy; Depression; Pupilometry; Emotion; Affect; Psychophysiology

Introduction

The most effective treatments for major depressive disorder (MDD), including Cognitive Therapy (CT) (1), yield response among 40–60% of depressed patients (2). Patients with low severity are likely to benefit (3–5). Reliable methods to determine which moderately severe patients will most strongly benefit from CT do not exist, leading to delivery of unwarranted treatments, long times to remission, and a high societal toll. Identifying which patients will remit in CT could decrease suffering, healthcare burden, and costs by improving the precision of referrals. Thus, we examined prognosis for remission in depressed patients entering CT using a quick, inexpensive physiological measure.

CT involves, among other skills, learning to interrupt automatic sustained emotional processing (the “downward spiral” of automatic negative thoughts and dysphoric affects). Neuroimaging studies demonstrate associations of symptom change in CT with sustained limbic activity, potentially reflecting automatic emotional engagement, and decreased prefrontal function, often associated with decreased emotion regulation, with recovery (6–8). Potentially, CT helps depressed participants to address neurocognitive features that most strongly characterize their depression. Yet, neuroimaging is lengthy, cumbersome, and expensive, and therefore unlikely to be employed clinically in the near future. Valid, brief, and inexpensive psychophysiological proxies for such measures would thus be useful. Though many behavioral and psychophysiological prognostic indicators of symptom change have been observed, including heart rate variability (9, 10), electroencephalographic asymmetry (11, 12), and dichotic listening performance (13), none have been routinely adopted clinically. Potential reasons include lack of links to psychological processes (i.e., face validity) and having not been examined in patients with at least moderate severity where their effort would be more strongly warranted. If true, it would be useful to show that a quick, inexpensive psychophysiological measure is 1) prognostic among more severe participants, 2) intuitively reflects psychological processes, and 3) reflects expected brain mechanisms.

Accordingly, our goal was thus to associate clinical remission (arguably, the standard patients would use for engaging in a treatment) with an intuitive, validated, easily acquired physiological measure with previously demonstrated differences between depressed and healthy individuals. Thus, we examined whether pupillary responses to negative words are

prognostic for remission in Cognitive Therapy in depressed individuals with moderate or higher levels of symptomatology. We chose pupillary responses as 1) They can be measured in depressed individuals. Pupillary responses to emotional words are increased and sustained in depressed compared to healthy individuals (14, 15) potentially reflecting increased limbic engagement or increased need for regulatory control. 2) They reflect cognitive and emotional processing. Pupils dilate in response to cognitive load (16) and in response to emotional stimuli (17), and remain dilated throughout expenditure of cognitive load (18). To assess both processes, we used a task in which trials alternated between demands previously associated with reliable pupillary responses including attention to emotional features of stimuli (emotion naming (14)) and executive control (digit sorting (19)). 3) They reflect relevant brain mechanisms. Pupillary responses such as dilation are well known to reflect both limbic activity (20) and mechanisms of cognitive control (19). That said, like other unidimensional psychophysiological measures, they are unable to distinguish between different neural inputs, thus not allowing directional inferences regarding association with constructs such as reactivity and regulation.

We hypothesized that if pupil dilation following a negative word primarily reflects limbic hyperactivity it would be positively associated with remission because it would indicate that specific abnormalities addressed in CT were present. Alternatively, if sustained pupil dilation is more reflective of prefrontal regulation, decreased sustained pupil dilation may be associated with better remission, again because CT is presumed to address the underlying mechanism of deficient regulatory control. Thus, as a secondary and more exploratory goal, limbic and prefrontal brain mechanisms were examined in a subset of participants who completed the same tasks again during functional magnetic resonance imaging (fMRI) to aid interpretability. To understand whether patterns associated with remission were consistent with decreased initial disease mechanism load (psychophysiological reactivity more like healthy individuals compared to non-remitters) or increased presence of the types of mechanisms addressed in treatment, secondary analyses compared remitters and non-remitters to healthy never-depressed participants.

Method

Participants

Participants included 36 outpatients with recurrent MDD (via SCID, 21) from a CT trial (Thase, PI; ClinicalTrials.gov: NCT00183664), and 53 controls with no history of Axis I disorder. Participants reported no significant eye problems, interfering health problems, psychoactive drug or alcohol abuse within the past six months¹, antidepressant use within two weeks (six for fluoxetine), history of psychosis, or manic episodes. Participants passed a cognitive screen (NAART-II VIQ > 80, 22). Figure 1 shows patient flow yielding 51 controls and 32 depressed participants who received CT. Of these participants, 20 completed fMRI assessment as part of a later-funded protocol.

Procedures

As in Figure 1, after signing IRB-approved consent forms, a diagnostic interview, and word-list generation, participants completed the Beck Depression Inventory II (BDI-II) to assess depressive severity on the testing day and information processing tasks during pupillary assessment (details in Supplementary Data – Method: Apparatus). A digit sorting task (repeatedly putting digits in numerical order) (19) to decrease novelty effects, was followed by other tasks in a counterbalanced order. Eligible patients were assessed again with fMRI (34 or 30 3.2mm slices acquired parallel to the AC-PC line (3T GE scanner, T2*-weighted

¹one control participant's alcohol use was ambiguous, and they could not be recontacted

images, reverse spiral pulse sequence to increase signal strength in the amygdala and orbitofrontal regions; TR=1500ms, TE=5ms, FOV=24cm, flip=60) within two weeks.

The analyzed task involved 60 27-second compound emotion-identification/digit sorting trials as shown in Figure 1. Emotion-identification consisted of an instruction (“What’s the emotion”; 1s), fixation mask (1s), word (200ms; 10 positive, 10 negative, and 10 neutral personally-relevant and 10 positive, 10 negative, and 10 neutral normed words (methods in 14, 15 and Supplementary Data: Method: Word List Generation)) presented in a different pseudorandom order for each participant, and mask (8.5s) during which participants were instructed to name the emotionality of the word using buttons for “Positive” (labeled “+”), “Negative” (“-”), or “Neutral” (“N”) as quickly and accurately as they could. The digit-sorting portion contained a cue (“Order the numbers”; 1s), mask (1s), five digits (2s), mask (5s), and target digit from the set (7.5s). Participants were told that when the digits appeared, they should read them from left to right, put them in numerical order in memory, and remember the middle digit in the sorted list; when the target appeared, they should push buttons for “yes” (“Y”) or “no” (“N”) to indicate whether the target was the middle digit from the previous set, as quickly and accurately as they could. Participants were told they should not use special strategies such as sorting the digits by moving their eyes on the screen or sorting only the first few digits, because we were examining the process of sorting items in memory. Pupil dilation and prefrontal activity were parametrically associated with the number of digits to be sorted on the digit sorting task in healthy participants from this sample (19, 23).

Depressed participants then received 16 sessions (20 for those with Ham-D reduction <40% at session 9) of CT following Beck’s (1) guidelines. At each session, participants completed the Beck Depression Inventory I (BDI) ² to assess depressive severity.

Data Selection and Analysis

Clinical status—Final severity scores (BDI) were obtained at patients’ blind evaluation, approximately 1–2 weeks following their last treatment session. For patients missing the blind evaluation, severity from their last-attended appointment was used, allowing a last-observation-carried-forward (LOCF) analysis. Remission was defined as having a final depressive severity of BDI<10 and no MDD diagnosis following acute treatment. While there is no agreed-on threshold for remission (24) based on the BDI, criteria have ranged from <9 (25) to <12 (26); our criterion was weighted towards specificity as the probable use of this measure is likely to be in deciding to opt out of rather than into therapy.

Pupillary data—Data were cleaned using our lab’s standard procedures (23) including interpolation through blinks and removal of trials with over 50% blinks, reaction times <150 ms, and incorrect digit sorting responses ($Md=6$, $M(SD)=9.6$ (10.7) trials eliminated per subject), and subtraction of a 100ms prestimulus baseline mean to produce a response index. Data from the second pupilometer was scaled and offset to match the controls (classification only ever improved when this procedure was not implemented).

Because we were interested primarily in sustained pupillary responses, conventional statistics such as peak-response are not appropriate. Rather, contrasts were examined via statistical tests at each sample along pupillary waveforms. To control type 1 error, 35 consecutive samples (.56 seconds) of tests in a row significant at $p<.1$, considered

²The BDI-I (henceforth BDI) was used at clinical assessments as part of the larger associated clinical protocol. The BDI-II was used at physiological testing as part of the physiological protocol and was thus used *only* to separate participants into high and low severity groups at the assessment day. The protocols were linked post-hoc for the current analysis. In our depressed sample, BDI-I could be estimated as $4.7+.83*BDI-II$ with high reliability, $r=.78$.

replications, were considered significant at $p=.05$ (described in (27), (23) and Supplementary Data: Method: Type I Error Control). Results report tests of the mean pupillary response in significant windows.

fMRI—fMRI data were preprocessed via motion and outlier-correction, linear trend removal, temporal smoothing (5-point filter), linear cross-registration to a reference brain, and spatial smoothing (6mm FWHM). Relationships of brain activity to pupillary response were examined in regions of interest (ROIs) previously implicated in sustained processing of emotion in depression (28) and treatment outcome, including the left dorsolateral prefrontal cortex (DLPFC) region that differentiated depressed and control participants on digit sorting (28), a right DLPFC region from the same dataset which reflected parametric activation with task difficulty in both depressed and controls on that task via conjunction analysis (within-group ANOVAs thresholded at $p<0.05$ for each component map yielding $p<0.0025$ for the conjunction, thresholded using an empirically determined contiguity threshold for this threshold via AFNI's (29) AlphaSim), the prognostic BA25 region previously examined in a subset of this sample (6), a hand-traced left amygdala region as in (28), and an anatomically defined ROI from Afni's Talairach atlas for Brodmann's area 24 in the rostral cingulate gyrus, which has been consistently associated with outcome in other studies (reviewed in 30). Raw mean condition-related time-courses represented as %-change from the first-scan of a trial were extracted.

We used partial least squares to capture modes of variation in the fMRI signal associated with maximally discriminant pupillary motility (11.78s) across depressed participants ($N=20$ had fMRI data). Subject was a random factor and each brain region was examined separately.

Results

Sociodemographic and behavioral data

Sociodemographic and behavioral data are presented in Table 1. Participants rated positive words as positive, negative words as negative, and neutral words as neither positive nor negative. Performance was excellent on digit-sorting (means $>90\%$ correct for both groups following all types of words). Depressed participants rated positive words as less positive than control participants and were slower to name the valence of positive words. No other reliable group or condition-related differences in behavioral data (ratings, reaction times to valence identification, percent-correct on digit-sorting) were detected.

Clinical change and initial severity

Depressive symptoms reliably decreased following therapy, $BDI\ M(SD)_{pre-CT}=29.6(12.9)$, $BDI\ M(SD)_{post-CT}=11.7(11.9)$, $t(31)=8.87$, $p<0.005$, standardized difference (Cohen's d) $=1.6$. A total of 17 patients met remission criteria; the remaining 15 patients were non-remitters. Logistic regression of initial severity on remission suggested, as in previous studies, that higher pre-treatment severity was associated with a lower likelihood of remission ($\beta=-.01$, $R^2=.41$, $p<.0005$). A standard high-end pretreatment BDI cut score of 20 (31–34) yielded 59% correct classification. For participants with low initial severity ($BDI\leq 20$), there was 1 non-remitter, and 6 remitters. Thus, most participants with low levels of initial severity remitted in CT. To aid prognosis of clinical remission for more severe participants, subsequent pupil analyses concentrated on discriminating remitters with high pupil-assessment-day severity scores ($BDI-II>20$; henceforth “high-pre-treatment-severity”; $N=25$; 14 non-remitters and 11 remitters) from non-remitters with pupil-assessment-day severity scores below this threshold ($N=7$; 1 non-remitter; 6 remitters – similar to the use of pre-treatment BDI scores).

Associations of pupillary responses to negative stimuli with remission in CT

Figure 2 shows pupillary responses to negative words followed by digit-sorting on the alternating task for controls, and for high-pretreatment-severity remitters and non-remitters. High-pretreatment-severity remitters had lower pupil dilation in response to emotional words than the low-pretreatment-severity participants in multiple significant temporal windows: 0.82–1.93s: $t(23)=2.43$, $p=0.02$, $D=0.04\text{mm}$, $d=0.98$, 11.15–12.83s: $t(23)=2.03$, $p=0.05$, $D=0.10\text{mm}$, $d=0.82$, 13.02–15.55s: $t(23)=2.14$, $p=0.04$, $D=0.10\text{mm}$, $d=0.86$, 19.80–20.83s: $t(23)=2.10$, $p=0.05$, $D=0.10\text{mm}$, $d=0.84$, 26.32–27.00s: $t(23)=2.43$, $p=0.02$, $D=0.07\text{mm}$, $d=0.98$.

Maximal discrimination between high-pretreatment-severity remitters and non-remitters was achieved in multiple intervals from 11.5–12.25 seconds after trial onset (9.5–10.25 seconds after stimulus onset), yielding 88% correct classification, $p<.0005$ via permutation tests. Figure 3 shows a scatter-plot in which final severity is on the x-axis and the representative median pupillary prediction at 11.78 seconds is on the y-axis. As shown in Figure 3, of the high initial-severity participants, 9/11 remitters had pupillary responses less than $-.05$ (82% sensitivity) whereas 13/14 non-remitters had pupillary responses $>-.05$ (93% specificity). Monte Carlo simulations accounting for the number of employed comparisons and temporal smoothness of relevant pupillary waveforms ($r=.998$; 150 iterations) further suggested that prognosis associated with pupillary responses was better than expected by chance, for which a classification rate of 88% correct was never achieved ($p<.007$), and for which a classification rate of .8 was achieved less than 95% of the time. Leave-one-out cross validation using the same method obtained similar intervals and thresholds for all participants (11.7–12.2s, thresholds between $-.09$ and $-.05$) with a 76% correct classification rate suggesting that results are relatively robust. To further demonstrate robustness, when participants were further restricted to those who displayed high severity at both the initial therapy and pupil-assessment days ($\text{BDI-I}>19$, $\text{BDI-II}>20$; $N=22$), leave-one-out cross-validation yielded 86% correct classification.

Sensitivity analyses: Alternate remission criteria, continuous severity, and effects of valence, personal relevance, rumination, and pupilometer

As described in the Supplementary Data, when response was defined using a variety of alternate outcome measures statistics were at least as strong or stronger. When remission was defined as having a final Hamilton Rating Scale Score < 12 , among high-initial-severity participants there were 15 remitters and 8 non-remitters; and the same methods described above, 88% prognostic accuracy was achieved; 22/25 participants were correctly classified. When the initial-severity cutoff threshold and pupillary response were maximized simultaneously, results were similar – having a pupil-assessment-day score below 23 was strongly associated with remission – for those above BDI 23, pupillary responses below $-.07$ were significantly prognostic for remission.

As described in the Supplementary Data, the same qualitative pattern of differences in pupillary response between non-remitters and remitters was observed for using initial BDI-I scores as well as when neutral words were examined, but not for positive words. Remitters with low-BDI scores also appeared to display sustained pupillary responses to neutral words. Prognostic power was stronger for the idiosyncratically generated negative words than the normed words. The threshold which maximally discriminated high BDI remitters from non-remitters was similar regardless of pupilometer.

Also, as shown in the Supplementary Data, though pupillary responses in the region maximally predictive for remission were not strongly associated with continuous change in symptomatology in the high-initial-symptomatology group or the entire cohort, relationships

with pupillary motility in the last part of the trial (~30s following stimulus onset) were moderate. We thus examined whether sustained pupillary responses were associated with rumination as we have previously observed (15). As described in the Supplementary Data, pupillary responses were moderately but not significantly associated with self-reported rumination in the period of maximal discrimination.

Associations of Pupillary Responses with Brain Activity Measured Using fMRI

We employed fMRI data to examine whether pupillary responses strongly reflect activity in limbic regions (e.g., amygdala) or regions more strongly associated with regulatory control (e.g., DLPFC) among depressed participants who received treatment. Figure 4 shows the correlation at each scan within trials from regressions of BOLD response to negative words in the *a priori* regions with pupil responses at 11.78 seconds with statistically significant scans highlighted below the X axis, and scatter plots for the scans with the strongest. As the BOLD response lags neural activity by up to 6 seconds whereas pupillary responses lag neural function by only up to 1 second, BOLD activity at 17 seconds would be expected to be associated with pupillary function at 11.78 seconds proximally. As shown in the figure, positive zero-order correlations between pupillary responses and brain activity before/during the pupillary response were observed only for putative regulatory regions, the left DLPFC and BA24. Partial least squares regressions jointly accounting for the time-course of brain reactivity and its relationship to pupillary reactivity, confirmed that a significant relationship with the predictive aspects of pupil dilation was observed only for the DLPFC ($R^2=.33$, adj $R^2=.25$, $F(2,17)=4.2$, $p=.03$; significant modes of variation were positively associated, peaking at 7.5s, $p=.009$, and peaking at 10.5 and 25.5 seconds, $p=.01$). As shown in Supplementary Figure 10, zero order relationships were similar for controls in the DLPFC but not for Amygdala and BA25. As shown in Supplementary Figure 11, results were qualitatively similar when just the high BDI depressed participants were included rather than all treated patients except for the negative relationships in RDL PFC which was not robust. Finally, as described in the supplementary data, pupillary motility was moderately but not significantly related to functional connectivity across these structures.

Discussion

This experiment examined prognosis for remission in Cognitive Therapy (CT) using pupillary response during cognitive and emotional information processing. Consistent with previous work, individuals with lower pre-treatment depressive severity had better outcomes (3–5), whereas remission was more variable among those with higher pre-treatment severity. In this group, increased sustained pupillary responses were associated with lower remission rates. As in healthy individuals (19), brain imaging on our depressed participants suggested that increased pupillary responses were associated with increased dorso-lateral prefrontal function, which has been linked to executive control and emotion regulation. In support of this idea, an overlapping subset of depressed participants displayed decreased sustained pupillary responses on an executive control (Stroop) task without explicit emotional content (35). Similarities in prefrontal influences for controls and depressed participants could reflect common influences of voluntary regulatory control, and dissimilarities for the Amygdala and BA25 could reflect differential recruitment of these structures in depression. If these data are replicable, there are strong practical and theoretical implications.

Implications

Practically, results suggest an algorithm to help patients to decide whether to expend time and resources in treatment with CT. Consistent with other literature we suggest that patients with low depressive severity are likely remitters and thus may not benefit from further assessment to aid their decision. Among more severely depressed patients, assessment of

pupillary responses could provide a quick, inexpensive, sensitive, and specific pre-treatment prognostic indicator. Given the strong effects using just ten idiosyncratic negative words, the assessment could take less than ten minutes.

Theoretically, some CT techniques may be conceptualized as fostering the depressed person's ability to recruit executive control in response to negative stimuli or thoughts. To the extent that sustained pupil dilation reflects prefrontally mediated executive control, more severely depressed individuals who do not have this capability may benefit from what CT has to offer. Those who already have this capability may benefit from treatments that do not act by teaching executive control, e.g., antidepressant medications (30). Like our more severely depressed treatment-remitters, many depressed children display decreased pupillary responses to negative words on a similar task (36), potentially suggesting that they would be good candidates for CT.

This algorithm applies only to CT and cannot help to decide which of multiple treatments patients are best suited for. Additional studies are needed to determine if this profile is predictive of remission in other interventions, particularly those that do not target executive control. If effects are specific, patients with decreased sustained pupillary responses may also respond well to targeted interventions specifically designed to increase executive control (37).

These data qualify our observations of sustained pupil diameter in depressed individuals using similar tasks (14, 15), suggesting there may be multiple relevant subgroups of depressed individuals. The qualitative similarity of controls to non-remitting depressed (Figure 2) and mechanistic correlates (Supplementary Figure 8) could reflect increased mobilization of prefrontal control in each of these groups. Whether healthy individuals "need" such control as a function of the emotional impact of the stimuli or habitually engage these mechanisms either after an emotional stimulus or before the subsequent cognitive task is not clear from the current design.

Limitations

This study has a number of limitations. Without a no-treatment arm it is unclear whether predicted remitters would remit without CT – thus pupillary responses cannot yet be used in decisions of whether to require Cognitive Therapy. In addition, our sample was small compared to most treatment-moderator studies, requiring replication. The task was long, and part of a three-hour battery. Many facets of the task were analyzed only in an exploratory context (e.g., responses to positive words; see supplementary data) because our goal involved understanding association of physiological responses to negative stimuli with clinical change. This relationship was strong enough that adding other design facets would not have helped quantitatively. That said, the effects of fatigue due to the long task on observed results are difficult to gauge – replication without positive and neutral words and only one task could be useful. Control participants were younger, on average, than depressed participants – for the current study this was not a major confound as primary comparisons did not regard controls. The clinical sample was entirely composed of recurrently depressed participants limiting generalizability. Predictions using pupillary responses to neutral words were similar, potentially suggesting either 1) the operative feature of the task was not negative words, but the cognitive control portion (digit sorting) or anticipation thereof or 2) negative elaboration on neutral stimuli. Brain imaging data was acquired in participants who had already undergone pupillary assessment on the same task, potentially introducing practice effects. The low number of subjects with brain imaging data precludes reliable conclusions regarding mechanistic specificity. Rather, these analyses provide initial exploratory evidence useful in interpreting the most strongly contributing underlying

mechanisms. The current data do not say whether examined phenomena were trait- or state-related.

Summary and future directions

These limitations notwithstanding, results suggest a sensitive, specific, inexpensive, quick, and mechanistically interpretable method to aid in treatment planning for depression. The assessment setup could be implemented for under \$12,000 in virtually any clinic environment. Costs per assessment are trivial – requiring ten to thirty minutes from a minimally-trained technician. The software for implementing this experiment can be obtained, for free, from the PI. Thus, upon replication, this type of psychophysiological assessment could become an integral part of the emerging trend towards personalized treatment for depression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the contributions of Roma Konecky, Lisa Farace, Agnes Haggerty, Mauri Cesare, Mandy Collier, Kelly Magee, and the staff of the Mood Disorders Treatment and Research Program at Western Psychiatric Institute and Clinic.

This research was supported by MH082998, MH064159, MH60473, MH55762, MH58397, MH58356, NARSAD, the Veteran's Administration, and the Veteran's Research Foundation. This material is the result of work supported with resources and the use of facilities at the Pittsburgh VA Healthcare System, Highland Drive Division.

Supported by MH082998, MH064159, MH60473, MH55762, MH58397, MH58356, NARSAD, the Veteran's Administration, and the Veteran's Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. Cognitive therapy of depression. New York: Guilford Press; 1979.
2. Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. *Psychological Science in the Public Interest*. 2002; 3:39–77.
3. Thase ME, Simons AD, Cahalane J, McGeary J, Harden T. Severity of depression and response to cognitive behavior therapy. *Am J Psychiatry*. 1991; 148:784–789. [PubMed: 2035722]
4. Thase ME, Greenhouse JB, Frank E, Reynolds CF 3rd, Pilkonis PA, Hurley K, et al. Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Arch Gen Psychiatry*. 1997; 54:1009–1015. [PubMed: 9366657]
5. Jones NP, Siegler GJ, Thase ME. Effects of rumination and initial severity on response to Cognitive Therapy for depression. *Cognitive Therapy and Research*. (in press).
6. Siegler GJ, Carter CS, Thase ME. Use of fMRI to predict recovery from unipolar depression with cognitive behavior therapy. *Am J Psychiatry*. 2006; 163:735–738. [PubMed: 16585452]
7. McClure EB, Adler A, Monk CS, Cameron J, Smith S, Nelson EE, et al. fMRI predictors of treatment outcome in pediatric anxiety disorders. *Psychopharmacology (Berl)*. 2007; 191:97–105. [PubMed: 16972100]
8. Fu CH, Williams SC, Cleare AJ, Scott J, Mitterschiffthaler MT, Walsh ND, et al. Neural Responses to Sad Facial Expressions in Major Depression Following Cognitive Behavioral Therapy. *Biol Psychiatry*. 2008
9. Rottenberg J, Kasch KL, Gross JJ, Gotlib IH. Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion*. 2002; 2:135–146. [PubMed: 12899187]

10. Rottenberg J, Salomon K, Gross JJ, Gotlib IH. Vagal withdrawal to a sad film predicts subsequent recovery from depression. *Psychophysiology*. 2005; 42:277–281. [PubMed: 15943681]
11. Iosifescu DV, Greenwald S, Devlin P, Mischoulon D, Denninger JW, Alpert JE, et al. Frontal EEG predictors of treatment outcome in major depressive disorder. *Eur Neuropsychopharmacol*. 2009; 19:772–777. [PubMed: 19574030]
12. Bruder GE, Sedoruk JP, Stewart JW, McGrath PJ, Quitkin FM, Tenke CE. Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: pre- and post-treatment findings. *Biol Psychiatry*. 2008; 63:1171–1177. [PubMed: 18061147]
13. Bruder GE, Stewart JW, Mercier MA, Agosti V, Leite P, Donovan S, et al. Outcome of cognitive-behavioral therapy for depression: relation to hemispheric dominance for verbal processing. *J Abnorm Psychol*. 1997; 106:138–144. [PubMed: 9103725]
14. Siegle GJ, Granholm E, Ingram RE, Matt GE. Pupillary response and reaction time measures of sustained processing of negative information in depression. *Biological Psychiatry*. 2001; 49:624–636. [PubMed: 11297720]
15. Siegle GJ, Steinhauer SR, Carter CS, Ramel W, Thase ME. Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self-reported rumination. *Cognitive Therapy and Research*. 2003; 27:365–382.
16. Beatty J. Task-evoked pupillary responses processing load and the structure of processing resources. *Psychological Bulletin*. 1982; 91:276–292. [PubMed: 7071262]
17. Janisse MP. Pupil size and affect: A critical review of the literature since 1960. *The Canadian Psychologist*. 1973; 14:311–329.
18. Beatty J. Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology*. 1982; 19:167–172. [PubMed: 7071295]
19. Siegle GJ, Steinhauer SR, Stenger V, Konecky R, Carter CS. Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *NeuroImage*. 2003; 20:114–124. [PubMed: 14527574]
20. Koikegami H, Yoshida K. Pupillary dilation induced by stimulation of amygdaloid nuclei. *Folia Psychiatrica Neurologica Japonica*. 1953; 7:109–125.
21. First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JB. Structured Clinical Interview for DSM IV Axis I Disorders Patient Edition. New York: Biometrics Research Department New York State Psychiatric Institute; 1996.
22. Nelson, HE.; Willison, J. The National Adult Reading Test. Berkshire, England: Nfer-Nelson Publishing Co; 1991.
23. Siegle GJ, Ichikawa N, Steinhauer S. Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*. 2008; 45:679–687. [PubMed: 18665867]
24. Bandelow B, Baldwin DS, Dolberg OT, Andersen HF, Stein DJ. What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *J Clin Psychiatry*. 2006; 67:1428–1434. [PubMed: 17017830]
25. Keller MB. Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. *JAMA*. 2003; 289:3152–3160. [PubMed: 12813121]
26. Riedel M, Moller HJ, Obermeier M, Schennach-Wolff R, Bauer M, Adli M, et al. Response and remission criteria in major depression - A validation of current practice. *J Psychiatr Res*. (in press).
27. Guthrie D, Buchwald JS. Significance testing of difference potentials. *Psychophysiology*. 1991; 28:240–244. [PubMed: 1946890]
28. Siegle GJ, Thompson W, Carter CS, Steinhauer SR, Thase ME. Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: Related and independent features. *Biol Psychiatry*. 2007; 61:198–209. [PubMed: 17027931]
29. Cox R. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers in Biomedical Research*. 1996; 29:162–173.

30. DeRubeis RJ, Siegle GJ, Hollon SD. Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms. *Nat Rev Neurosci*. 2008; 9:788–796. [PubMed: 18784657]
31. Beck AT, Steer RA, Garbin M. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*. 1988; 8:77–100.
32. Kendall PC, Hollon SD, Beck AT, Hammen CL, Ingram RE. Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy & Research*. 1987; 11:289–299.
33. Oliver JM, Simmons ME. Depression as measured by the DSM-III and the Beck Depression Inventory in an unselected adult population. *J Consult Clin Psychol*. 1984; 52:892–898. [PubMed: 6501675]
34. Sloan DM, Marx BP, Bradley MM, Strauss C, Lang PJ, Cuthbert BN. Examining the high-end specificity of the Beck Depression Inventory using an anxiety sample. *Cognitive Therapy & Research*. 2002; 26:719–727.
35. Siegle GJ, Steinhauer SR, Thase ME. Pupillary assessment and computational modeling of the Stroop task in depression. *International Journal Psychophysiology*. 2004; 52:63–76.
36. Silk JS, Dahl RE, Ryan ND, Forbes EE, Axelson DA, Birmaher B, et al. Pupillary Reactivity to Emotional Information in Child and Adolescent Depression: Links to Clinical and Ecological Measures. *Am J Psychiatry*. 2007; 164:1873–1880. [PubMed: 18056243]
37. Siegle GJ, Ghinassi F, Thase ME. Neurobehavioral therapies in the 21st century: Summary of an emerging field and an extended example of Cognitive Control Training for depression. *Cognitive Therapy & Research*. 2007; 31:235–262.
38. Ratcliff R. Methods for dealing with reaction time outliers. *Psychological Bulletin*. 1993; 114:510–532. [PubMed: 8272468]

Figure 1.

CONSORT diagram of patient flow through the study. The box on the middle right displays the experimental protocol showing the time-course for the digit-sorting and alternating task. Only the alternating task was analyzed in this manuscript.

Figure 2.

Pupillary responses to negative words followed by digit-sorting on the alternating task for controls, and both responders and non-responders with high-initial-severity (BDI-II > 20). Significant differences between responders and non-responders are highlighted below the x-axis; yellow= $p < .1$, red= $p < .05$. Black lines show regions with enough consecutive tests to be considered significant.

Figure 3.

Scatter-plot in which final severity is on the x-axis and pupillary motility at 11.78 seconds is on the y-axis for high-initial-severity participants. Responders were defined as participants with final BDI < 10. The threshold of maximal discrimination between responders and non-responders (-0.07mm) is plotted as a blue line.

Figure 4.

Association in 20 treated depressed patients of BOLD activity in five *a priori* regions with pupil dilation at 11.78 seconds – the point which maximally discriminated depressed responders from non-responders. In the second column, the Y axis represents correlation and the X axis is seconds. Scans with significant correlations are highlighted as yellow: $p < .1$, red: $p < .05$. In the third column, each participant's data on both activity at the maximally predictive scan (X axis) and pupil dilation at 11.78 seconds (Y axis) are plotted.

Table 1

Sociodemographic, clinical, and behavioral data

Measure	Depressed	Control	Significant Difference (ns=not significant (p>.1), blank = not tested)
N	32	51	
# male	15	15	Ns
ethnicities	25 white, 7 African American	39 white, 5 African American, 3 Asian, 4 bi-racial	Ns
Age range	20–56*	19–55	
Age M(SD)	39.84 (10.0)	31.7 (10.8)	$t(81) = -3.4, p = .001$
Education M(SD)	14.94 (3.1)	16.1 (2.7)	$t(81) = 1.82, p = .07$
Severity: BDI at pupil assessment M(SD)	29.5 (12.0)	3.6 (4.3)	$t(36.1) = -11.71, p < .0005$
# depressive episodes	1 → >10 MD (4)	0	
NAART VIQ M(SD)	111.5 (8.5)	109.9 (8.1)	Ns
<u>Post-scan emotion ratings of employed words (1 = very negative and 7 = very positive):</u> Valence x group $F(2,78) = 8.9, p < .005, \eta^2 = .19$			
Positive words M(SD)	5.68 (.60)	6.12 (.41)	$t(79) = 3.89, p < .005, D = .44$
Negative words M(SD)	2.04 (.59)	2.06 (.45)	Ns
Neutral words M(SD)	4.05 (.22)	4.15 (.39)	Ns
<u>Reaction times</u> ** for valence identification in ms: Valence x group $F(2,79) = 8.06, p = .001, \eta^2 = .17$			
Positive words M(SD)	1723.3 (466.9)	1297.8 (546.7)	$t(81) = 3.64, p < .005, D = 425 \text{ ms}$
Negative words M(SD)	1643.1 (415.5)	1492.0 (597.1)	Ns
Neutral words M(SD)	1630.1 (504.7)	1453.8 (558.2)	Ns
<u>Digit-sorting following emotion-identification task Percent Correct.</u> Valence x group ns, Valence main effect ns			
Positive words M(SD)	.91 (.16)	.90 (.22)	
Negative words M(SD)	.92 (.14)	.92 (.20)	
Neutral words M(SD)	.93 (.14)	.92 (.19)	

* one participant was 55 when consented but 56 when tested

** Harmonic means were used for reaction times based on recommendation in (38). To eliminate spurious skew due to outliers while preserving rank-ordering of data, outliers more than 1.5 times the interquartile range from the median harmonic mean on any variable were scaled to the closest obtained value below this cutoff plus the difference between this value and the next closest value as in (14).