Published in final edited form as:

Percept Mot Skills. 2020 April ; 127(2): 401–414. doi:10.1177/0031512519898356.

Questioning the Effort Hypothesis that Depressed Patients Perform Disproportionately Worse on Effortful Cognitive Tasks

Dustin B. Hammers¹, Sara Weisenbach²

¹Center for Alzheimer's Care, Imaging, and Research, Department of Neurology, University of Utah

²Renaissance School of Medicine, Department of Psychiatry, Stony Brook University

Abstract

The debate over Hasher and Zacks' (1979) effort hypothesis – that performance on effortful tasks by patients with depression will be disproportionately worse than their performance on automatic tasks – shows a need for additional research to settle whether or not this notion is “clinical lore.” In this study, we categorized 285 outpatient recipients of neuropsychological evaluations into three groups - No Depression, Mild-to-Moderate Depression, and Severe Depression - based on their Beck Depression Inventory –2 (BDI-2) self-reports. We then compared these groups' performances on both “automatic” and “effortful” versions of the Ruff 2 & 7 Selective Attention Test Total Speed and Total Accuracy Indices, the Digit Span subtest from the WAIS-IV, and Trail Making Test Parts A and B, using a two-way (3×2) mixed MANOVA. Patients with Mild-to-Moderate Depression or Severe Depression performed disproportionately worse than patients with No Depression in our sample on more effortful versions of only one of the four attention or executive functioning measures (Trail Making Test). Thus, these data failed to fully support a hypothesis of disproportionately worse performance on more effortful tasks. While this study does not negate the effort hypothesis in some specific instances, particularly for use in the Trail Making Test, there is cause for clinician caution in routinely applying the effort hypothesis when interpreting test findings in most clinical settings and for most measures.

INTRODUCTION

Patients with depression frequently display worse performance across a number of cognitive domains, including attention, executive functioning, psychomotor speed, and memory (see recent critical reviews by (Ahern & Semkovska, 2017; Roca, Vives, Lopez-Navarro, Garcia-Campayo, & Gili, 2015)). In addition to neurobiological bases for this association (Schmaal et al., 2017; Schmaal et al., 2016), long-standing clinical belief suggests that “limited motivation, persistence, and lack of urgency” may contribute to these relationships (Barisa, 2014). For example, Hasher & Zacks (1979) proposed the “effort hypothesis” that

Address correspondences: Dustin B. Hammers, PhD, Center for Alzheimer's Care, Imaging and Research, Department of Neurology, University of Utah, 650 Komas Drive #106-A, Salt Lake City, UT 84108. Tel: 801-585-3929. dustin.hammers@hsc.utah.edu.

Disclosures: No authors associated with this project have reported conflicts of interest.

Data Availability: The data that support the findings of this study are available on request from the corresponding author (DBH). The data are not publicly available due to their containing information that could compromise the privacy of research participants.

performance on effortful tasks by patients with depression will be disproportionately worse compared to their performance on automatic tasks, owing to motivational factors. This theory was developed following a series of studies by the authors involving children, the elderly, and individuals under stress (Hasher & Zacks, 1979) related to tasks of varying degrees of cognitive load. This hypothesis has been supported by older research, leading some clinicians to use poor performance on effortful tasks as a marker for depression. Cohen and colleagues compared the performance of 11 inpatients with depression with 5 control participants on tasks of grip strength and verbal memory with varying delays, and observed that the greatest depression-related impairment was found on tasks with sustained effort (Cohen, Weingartner, Smallberg, Pickar, & Murphy, 1982). Similarly, Roy-Byrne examined automatic and effortful learning tasks in 10 inpatients with depression and 10 controls, observing that depressed patients' performance was worse on cognitively effortful versions of learning tasks (Roy-Byrne, Weingartner, Bierer, Thompson, & Post, 1986). Also, Tancer and colleagues found that among 17 inpatients with depression and 17 psychiatric controls, depressed patients performed worse on an effort-demanding task of learning and memory, but not on an effortless task (Tancer et al., 1990). Further, Zakzanis conducted a meta-analysis of 22 studies examining the effort hypothesis in individuals with depression (both inpatient and outpatient) and healthy controls across a variety of standard cognitive measures (domains of memory, executive functioning, fluency, speeded processing), and observed that effort-demanding tasks yielded effect sizes "almost capable of discriminating patients with depression from controls", whereas superficial processing tasks could not (Zakzanis, Leach, & Kaplan, 1998). Conversely, not all research findings have supported the effort hypothesis. For example, Den Hartog and colleagues refuted this hypothesis by comparing 30 depressed outpatients, 38 healthy controls, and 25 patients with severe allergic rhinitis on tasks of cognitive speed, concept shifting, and memory retrieval. They observed that the depressed group did not display worse performance on more effortful tasks relative to the control groups, but they did on automatic processing subtasks of speeded processing and memory scanning (Den Hartog, Derix, Van Bommel, Kremer, & Jolles, 2003). Additionally, in 45 outpatients with depression and 32 healthy controls – all of whom passed a measure of effort – Considine and colleagues identified that the depression group performed worse on a task of auditory memory, and they concluded that diminished effort alone cannot uniformly explain reduced cognitive performance in patients with depression (Considine et al., 2011).

Given continued debate about the accuracy of the effort hypothesis, even while it is embraced by many in neuropsychology, additional and more updated research is needed to determine whether this notion represents "clinical lore." Consequently, the aim of the current study was to test the effort hypothesis by examining differences in performance on automatic and effortful versions of cognitive tests by a large sample of patients presenting for outpatient neuropsychological evaluations with varying levels of self-reported symptoms of depression. Cognitive tasks were the Ruff 2 & 7 Selective Attention Task (Ruff; (Ruff, Niemann, Allen, Farrow, & Wylie, 1992), the Digit Span subtest from the Wechsler Adult Intellectual Scale – Fourth Edition (WAIS-IV; (Wechsler, 2008)), and Trail Making Tests Part A and B (Reitan, 1992). These are commonly-administered neuropsychological measures of attention and executive functioning that contain subtests varying in cognitive

load from automatic (Ruff Automatic Detection Speed and Accuracy, Trail Making Test Part A, and Digits Forward) to effortful (Ruff Controlled Search Speed and Accuracy, Trail Making Test Part B, and Digits Backward). Based on the effort hypothesis, we hypothesized that depressed patients would perform generally worse than non-depressed patients on measures of attention and executive functioning, and when administered both “automatic” and “effortful” versions of the measures, more depressed patients would perform disproportionately worse on more cognitively challenging versions of the tests than less depressed patients.

METHOD

Participants

In this retrospective study, we searched a database in the Division of Cognitive Neurology at a university in the Western USA for clinical cases in which patients completed the Beck Depression Inventory –2 (BDI-2; (Beck, Steer, & Brown, 1996)) and the other test measures between August 18, 2014 and October 15, 2019. This search yielded 299 cases, out of a total of 3,354 cases seen in clinic during that time frame. While no formal diagnostic exclusion criteria were utilized, the BDI-2 and the Ruff were not part of this respective clinic’s standard “dementia” test battery, meaning that older patients or those expected to require a less rigorous cognitive test battery were systematically excluded. Additionally, patients failing to pass a measure of effort (see below for procedure for determining failure on the Test of Memory Malingering (Tombaugh, 1997); No Depression = 3, Mild-to-Moderate Depression = 4, and Severe Depression groups = 7 participants) and one participant with an extremely low performance on a scale of intelligence were excluded from the study, consequently the final sample size was 285 participants. The sample tended to be younger-to-middle-aged adults ($M = 39.0$ years of age, $SD = 13.7$, range = 18 to 65), well-educated ($M = 14.3$ years, $SD = 2.5$, range = 8 to 20), predominantly Caucasian (88%), with equivalent proportions of male/female participants (56% male). They endorsed a wide range of depression symptoms on the BDI-2 ($M = 16.1$, $SD = 10.7$, range = 0 to 45), and their intellectual ability was average according to the WAIS-IV Full Scale score ($M = 105.2$, $SD = 14.5$, range = 70 to 144). Although all patients were referred for an evaluation of cognitive concerns, we made no attempt to limit the sample to certain groups associated with psychiatric or medical comorbidity. The participants presented for a wide variety of psychiatric (e.g., bipolar disorder, substance abuse; 8%), emotional (depressed mood, anxiety; 80%), neurological (e.g., multiple sclerosis, stroke, questionable concussion; 31%), or other medical concerns (e.g., cardiovascular disease, cancer; 28%), and came from a diversity of referral sources (e.g., Internal Medicine, Neurology, Psychiatry). No participants were involved in disability determinations or medico-legal examinations. All procedures for this research protocol were approved by the local Institutional Review Board.

Measures—The following test measures were administered to this patient sample during an outpatient neuropsychological evaluation, as part of a larger cognitive evaluation. Readers are referred to Strauss and colleagues (Strauss, Sherman, & Spreen, 2006) and respective test manuals for more detailed test descriptions and psychometric properties of these instruments. The Ruff 2 & 7 Selective Attention Test consists of a series of 20 trials of a

visual search and cancellation task, which evaluated speeded vigilance skills and incorporated Indices of Total Speed and Accuracy (Ruff et al., 1992). Both Indices were comprised of sub-indices for Automatic Detection (identifying 2s and 7s from a collection of letters on a page) and Controlled Search (identifying 2s and 7s from a collection of other numbers on a page). Ruff 2 & 7 test variables have adequate to high test-retest reliability ($r_{12} = .70$ to $.90$), with test-retest reliability higher for Speed than Accuracy scores (Allen & Ruff, 1990), and this test has been studied in patients with a wide range of psychiatric and neurological presentations (Strauss et al., 2006). The Digit Span subtest from the WAIS-IV (Wechsler, 2008) was also administered, which included subtests of Digits Forward and Digits Backward. Digits Forward requires examinees to repeat a number sequence in the same order as presented, and Digits Backward requires examinees to repeat a number sequence in reverse order. Test re-test reliability is very good ($r_{12} = .82$), and both criterion and construct validity for these variables are sound (Wechsler, 2008). The Digit Span subtest has been studied in a variety of neurological and psychiatric presentations (Wechsler, 2008). The Trail Making Test, Parts A (visual scanning) and B (set shifting) were additionally collected (Reitan, 1992). In this task, examinees are asked to draw a continuous line to connect bubbles on a page containing numbers (Part A) or alternating numbers and letters (Part B) in ascending order as quickly as possible, with time to completion the resultant test score. Test-retest reliability was adequate ($r_{12} = .79$) for Part A and high ($r_{12} = .89$) for Part B for normal and neurologically stable older adults tested 11 months apart, and the Trail Making Test has been studied in a variety of neurological and psychiatric presentations (Strauss et al., 2006).

We used raw scores on these measures, with higher scores indicating better cognitive functioning for the Ruff and Digit Span, and higher scores indicating worse performance for Trail Making Test, Parts A and B. As indicated above, Ruff Total Speed and Accuracy Automatic Detection, Digits Forward, and Trail Making Test Part A were conceptualized as being less cognitively challenging (automatic) tasks than Ruff Total Speed and Accuracy Controlled Search, Digits Backwards, and Trail Making Test Part B (conceptualized as effortful). We additionally assessed self-reported depression levels using the 21-item BDI-2 (Beck et al., 1996). Scores ranged from 0 to 63, with scores of 0 – 13 indicating minimal/no depression, 14 – 19 indicating mild depression, 20 – 28 indicating moderate depression, and 29 – 63 indicating severe depression. From these BDI-2 categorizations, we formed three groups of patients: (1) No Depression ($n = 142$), with BDI-2 scores ranging from 0 – 13; (2) Mild-to-Moderate Depression ($n = 99$), with BDI-2 scores ranging from 14 – 28; and (3) Severe Depression ($n = 44$), with scores ranging from 29 – 63. Further, the TOMM (Tombaugh, 1997) was administered as a measure of objective effort to ensure that the level of effort was not so poor as to have invalidated test performance. Performance below cutoff points in the TOMM test manual on all three trials of the test determined TOMM failure and subsequent exclusion from the study.

Data Analyses

We grouped cases by severity, based on BDI-2 scores, according to test manual guidelines (Beck et al., 1996). Following this classification, we compared depression groups on continuous demographic variables using analyses of variance (ANOVA) for interval data,

and chi-square analyses for categorical data. For the primary analyses, we calculated a two-way (3×2) mixed multivariate analysis of variance (MANOVA) with one within-subjects factor and one between-subjects factor for the Ruff Total Speed and Accuracy Indices, Digit Span, and Trail Making Test. Test Difficulty was the within-subjects factor (“automatic” and “effortful” versions of respective tests), and Depression group was the between subjects factor (No Depression, Mild-to-Moderate Depression, and Severe Depression). We used the mixed MANOVA to compare the main effects of Depression group and Test Difficulty – and their interaction – on cognitive test performance, with subsequent individual mixed ANCOVAs for each dependent variable. We conducted post-hoc analysis for any significant omnibus test, using *Tukey’s honest significance test*. We expressed measures of effect size throughout as η^2 values for continuous data and *Phi* coefficients for categorical data. We set a two-tailed alpha significance level at .05 for all statistical analyses. Based on previous findings in the literature (*effect size* = 0.70; (Cohen et al., 1982; Tancer et al., 1990), we determined that a sample size of approximately 156 patients (52 patients per group) would be necessary to obtain power ($1 - \beta$) at .95 based on this alpha level.

Regarding assumptions of normality in the data distribution for the BDI-2 variable, a visual inspection of Histograms/Normal Q-Q plots/Box plots indicated that the BDI-2 variable was approximately normally distributed, with a skewness of 0.65 (*Standard Error* [*SE*] = .14) and a kurtosis of −0.28 (*SE* = .29). While slightly positively skewed, in reasonably large sample sizes skewness should “not make a substantive difference in the analysis” (Tabachnick & Fidell, 2007). Although the *Kolmogorov-Smirnov* (K-S) Test of Normality was significant, K-S (285) = 0.111, $p = .001$, Ghasemi and Zahediasl (Ghasemi & Zahediasl, 2012) observed that the K-S test is overly sensitive to deviations from normality with increasing sample sizes beyond $n = 50$. Thus, in light of our very large sample size and our visual inspection, we considered the BDI-2 data as being approximately normally distributed.

RESULTS

Among patient groups categorized for levels of depression symptoms, we observed no differences between the three groups for demographic variables of sex, $\chi^2(2) = 5.01$, $p = .08$, *Phi* = .13, age, $F(2, 282) = 2.78$, $p = .06$, $\eta^2 = .019$, or education, $F(2, 282) = 1.74$, $p = .19$, $\eta^2 = .012$. See Table 1.

As noted, we conducted a two-way (3×2) mixed MANOVA comparing the main effects of Depression group and Test Difficulty – and their interaction – on cognitive performance on the Ruff 2 & 7 Selective Attention Task (Total Speed and Accuracy Indices), Digit Span, and Trail Making Test. Omnibus testing indicated that there were statistically significant differences between Depression groups on the dependent variables, $F(8, 552) = 2.62$, $p = .008$; *Wilks’ Lambda* = .93, $\eta^2 = .037$, between levels of Test Difficulty, $F(4, 276) = 218.26$, $p < .001$; *Wilks’ Lambda* = .24, $\eta^2 = .760$, and for the interaction between Depression group and Test Difficulty, $F(8, 552) = 2.03$, $p = .04$; *Wilks’ Lambda* = .94, $\eta^2 = .029$.

When considering the results for the dependent variables separately, the findings were as follows (see Table 2 and Figure 1). For the Ruff Total Speed Index, the main effect of

Depression group was significant, $F(2, 279) = 6.52, p = .002, \eta^2 = .045$, with the Ruff Total Speed Index performed significantly more poorly by patients in the Severe Depression group ($M = 108.73, SE = 4.8$) than the No Depression group ($M = 128.54, SE = 2.7$) or the Mild-to-Moderate Depression group ($M = 125.53, SE = 3.2$), $p = .001$ and $p = .011$, respectively; no differences were observed between the No Depression group and the Mild-to-Moderate Depression group, $p = .77$. The main effect of Test Difficulty was also significant, $F(1, 279) = 192.72, p < .001, \eta^2 = .409$, indicating that performance across all groups was significantly worse for the Controlled Search “effortful” version of the Ruff Total Speed Index ($M = 113.50, SE = 1.9$) than for the Automatic Detection “automatic” version of this test ($M = 128.36, SE = 2.5$). The interaction effect between Depression group and Test Difficulty was not significant, $F(2, 279) = 2.68, p = .07, \eta^2 = .019$. For the Ruff Accuracy Index, the main effect of Depression group was not significant (though a trend was observed), $F(2, 279) = 1.84, p = .16, \eta^2 = .013$, though the main effect of Test Difficulty was significant, $F(1, 279) = 150.35, p < .001, \eta^2 = .350$, indicating that performance across all groups was significantly worse for the Controlled Search “effortful” version ($M = 92.77, SE = 0.39$) than for the Automatic Detection “automatic” version of the Ruff Accuracy Index ($M = 95.77, SE = 0.34$). The interaction effect was non-significant as well, $F(2, 279) = 0.14, p = .87, \eta^2 = .001$.

For the Digit Span test of the WAIS-IV, the main effect of Depression group was significant, $F(2, 279) = 3.98, p = .02, \eta^2 = .028$, with performance on Digit Span significantly worse for the Severe Depression group ($M = 8.48, SE = 0.3$) than for the No Depression group ($M = 9.50, SE = 0.2$), $p = .014$, but not for the Mild-to-Moderate Depression group ($M = 9.27, SE = 0.2$), $p = .095$ (though a non-significant trend was present); no differences were observed between the No Depression group and the Mild-to-Moderate Depression group, $p = .68$. The main effect of Test Difficulty was also significant, $F(1, 279) = 130.54, p < .001, \eta^2 = .319$, with performance across all groups being significantly worse for the Digits Backward “effortful” version ($M = 8.26, SE = 0.16$) than for the Digits Forward “automatic” version ($M = 9.90, SE = 0.15$). The interaction effect was not significant, $F(2, 279) = 1.20, p = .60, \eta^2 = .009$. Finally, for the Trail Making Test, the main effect of Depression group was significant, $F(2, 279) = 4.97, p = .008, \eta^2 = .034$, with Trail Making Test performed significantly worse by the Severe Depression group ($M = 69.97, SE = 4.9$) than by the No Depression group ($M = 55.68, SE = 2.5$) or the Mild-to-Moderate Depression group ($M = 53.60, SE = 3.0$), $p = .016$ and $p = .007$, respectively; no differences were observed between the No Depression group and the Mild-to-Moderate Depression group, $p = .85$. The main effect of Test Difficulty was also significant, $F(1, 279) = 426.99, p < .001, \eta^2 = .605$, with performance across all groups being significantly worse for the Trail Making Test Part B “effortful” version ($M = 85.42, SE = 3.1$) than for the Trail Making Test Part A “automatic” version ($M = 34.08, SE = 1.07$). In contrast to the other analyses, we observed a significant Depression group * Test Difficulty interaction for this task, $F(2, 279) = 5.37, p = .005, \eta^2 = .037$. An inspection of Figure 1 indicates that the Severe Depression group performed disproportionately worse on Trail Making Test Part B (relative to Part A) than either the No Depression or the Mild-to-Moderate Depression group.

DISCUSSION

The current study sought to examine cognitive performance differences in patients with varying levels of self-reported symptoms of depression, with particular attention to performance on automatic and effortful versions of cognitive tests. Specifically, we categorized patients into No Depression, Mild-to-Moderate Depression, and Severe Depression groups based on their self-reported endorsements of depression on the BDI-2 and compared their performance across both automatic and effortful versions of the Ruff 2 & 7 Selective Attention Test Total Speed and Total Accuracy Indices, the Digit Span Test from the WAIS-IV, and the Trail Making Test.

Within our sample of 285 recent participants seeking outpatient neuropsychological evaluations, we found that patients with severe depression performed worse than patients with mild-to-moderate depression or no depression when collapsed across global scores for most measures administered (significantly worse for Ruff Total Speed Index, Digit Span, Trail Making Test, and a non-significant trend toward lower performance on the Ruff Accuracy Index). Our results are consistent with the expectation, based on early writings and research, that patients with serious depression are susceptible to decrements in attention and executive functioning (Ahern & Semkowska, 2017; Roca et al., 2015). Also of relevance to an interpretation of this finding, it is possible that some patients with depression in this study experienced decreased mood due to medical comorbidity or anxiety, or because of self-recognized cognitive deficits or other symptoms associated with neurological conditions that may have precipitated the evaluation. Across the entire sample, performance on more conceptually challenging versions of the test measures was significantly worse than performance on more effortless versions (Ruff Total Speed Index, Ruff Accuracy Index, Digit Span, Trail Making Test), which is not necessarily surprising and potentially speaks to the clinical nature of our sample.

Of direct relevance to the purpose of this study, as seen in Table 2 and Figure 1, we observed mixed results regarding the interaction between Depression group and Test Difficulty for the cognitive measures. For the Trail Making Test, the interaction effect was significant, with the Severe Depression group performing disproportionately worse on Trail Making Test Part B (relative to Part A) than either the No Depression or the Mild-to-Moderate Depression group. This result is consistent with past research in both inpatient and outpatient depressed individuals (Cohen et al., 1982; Tancer et al., 1990; Zakzanis et al., 1998), lending partial support to Hasher and Zacks' effort hypothesis (1979), which theorizes that depressed patients perform *disproportionately* worse on more effortful cognitive tests as a result of reduced ability to allocate effortful resources towards the measure at hand. However, our results fail to support Hasher and Zacks' effort hypothesis entirely, as well as our expectations. Specifically, there was no significant interaction between Depression Group and Test Difficulty for Digit Span, Ruff Total Speed Index, or Ruff Accuracy Index, meaning that more severe depression groups did not perform disproportionately worse on more effortful cognitive tasks for three of the four tasks administered. Such findings are consistent with Den Hartog and colleagues' (2003) results that failed to observe lower cognitive test performance on more effortful cognitive tasks for depressed versus non-depressed patients in an outpatient setting. When considering that many of the early studies to support the effort

hypothesis were of small sample size and assessed psychiatric inpatients with depression, our null findings may speak to limitations in the effort hypothesis for outpatient samples with elevated depression or emotional distress.

Den Hartog's work also potentially illuminates a reason that more severe Depression groups performed disproportionately worse on the Trail Making Task. Their study observed that cognitive-speed-based tasks were particularly susceptible to the effects of depression, which is consistent with the nature of the timed Trail Making Test. Additionally, select impairment in attentional shifting (as seen in the Trail Making Test) may be present in more severe cases of depression – particularly given that cognitive control is known to be particularly vulnerable to the impact of depression (Peters et al., 2017) – which may drive the interaction effect observed for this task. Even with this Trail Making Test finding, our varied results suggest that motivational factors or effort alone may not fully explain cognitive dysfunction in outpatients with depression. Underlying neurological factors (e.g., neuroinflammation, (Kohler et al., 2017); monoamine neurotransmitter abnormalities, (Morgese & Trabace, 2019); hippocampal volume loss and cortical disconnectivity; (Sampath, Sathyanesan, & Newton, 2017) may also be playing a role that would explain these performance differences.

Our equivocal findings do not negate the effort hypothesis in some specific instances, particularly for use in the Trail Making Test, though it is not supported for the majority of test measures administered to our sample. It remains possible that a different sample of patients tested – with different neuropsychological measures – might show different test performance results. For example, our selection of test measures tended to exclude older adults (only 7% of the sample was over the age of 60) who would have been given shorter test batteries and adults for which primarily referral concerns were dementia due to neurodegenerative disease. Thus, our results cannot be generalized to those populations. Additionally, our sample was only comprised of outpatients with varying levels of self-reported depression; while our BDI-2 scores reflected a wide range of endorsed depression (observed BDI-2 scores of 0 – 45), unlike in much of the early effort hypothesis research (Cohen et al., 1982; Roy-Byrne et al., 1986; Tancer et al., 1990), we did not include patients with sufficiently severe depression to warrant current psychiatric inpatient hospitalization. Consequently, caution should be taken before extending these results to psychiatric inpatients with depression-related affective disorders. Third, our findings are specific to the cognitive measures administered in this test battery, limiting their generalization to other measures of attention and executive functioning, or other domains of cognition (e.g., memory). Fourth, these results may not generalize to more heterogeneous participants with regard to premorbid functioning, education, race, and sex. Fifth, we selected patients into this study irrespective of their treatment with anti-depressant medication or psychotherapy, or age-of-onset or recurrence of decreased mood, and used no formal diagnostic criteria to form our depression groups. Finally, this study examined only two levels of difficulty across each task, and cannot speak to the possibility of a non-linear effect of effort. Future studies examining tasks with more gradations of complexity are encouraged to ascertain whether engagement might remain stable to a certain level of task difficulty before participants with more severe depression resign. Despite these limitations, however, these results lend further caution to the depression-based effort hypothesis. While the presence of depression appears to negatively impact attention and executive functioning performance generally, worsening

depression only resulted in disproportionately worse cognitive performance on more effortful, relative to less effortful, cognitive tasks for one of four test measures of executive functioning and attention in this sample. Clinicians should exercise some reluctance in incorporating the effort hypothesis when interpreting clinical test data in most clinical settings and for most measures.

Acknowledgements:

The project described was supported by an anonymous grant to the Center for Alzheimer's Care, Imaging and Research.

REFERENCES

- Ahern E, & Semkovska M (2017). Cognitive functioning in the first-episode of major depressive disorder: A systematic review and meta-analysis. *Neuropsychology*, 31(1), 52–72. doi:10.1037/neu0000319 [PubMed: 27732039]
- Allen CC, & Ruff RM (1990). Self-rating versus neuropsychological performance of moderate versus severe head-injured patients. *Brain Inj*, 4(1), 7–17. [PubMed: 2297602]
- Barisa M (2014). Mood Disorders: Depression, Mania, and Anxiety In Stucky K, Kirkwood M, & Donders J (Eds.), *Neuropsychology Study Guide & Board Review*. New York: Oxford University Press.
- Beck A, Steer R, & Brown G (1996). *Manual for the Beck Depression Inventory - II*. San Antonio, TX: Psychological Corporation.
- Cohen RM, Weingartner H, Smallberg SA, Pickar D, & Murphy DL (1982). Effort and cognition in depression. *Arch Gen Psychiatry*, 39(5), 593–597. [PubMed: 7092490]
- Considine CM, Weisenbach SL, Walker SJ, McFadden EM, Franti LM, Bieliauskas LA, ... Langenecker SA (2011). Auditory memory decrements, without dissimulation, among patients with major depressive disorder. *Arch Clin Neuropsychol*, 26(5), 445–453. doi:10.1093/arclin/acr041 [PubMed: 21593060]
- Den Hartog H, Derix M, Van Bommel A, Kremer B, & Jolles J (2003). Cognitive functioning in young and middle-aged unmedicated out-patients with major depression: testing the effort and cognitive speed hypotheses. *Psychological Medicine*, 33(8), 1443–1451. [PubMed: 14672253]
- Ghasemi A, & Zahediasl S (2012). Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab*, 10(2), 486–489. doi:10.5812/ijem.3505 [PubMed: 23843808]
- Hasher L, & Zacks R (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108(3), 356–388.
- Kohler CA, Freitas TH, Maes M, de Andrade NQ, Liu CS, Fernandes BS, ... Carvalho AF (2017). Peripheral cytokine and chemokine alterations in depression: a meta-analysis of 82 studies. *Acta Psychiatr Scand*, 135(5), 373–387. doi:10.1111/acps.12698 [PubMed: 28122130]
- Morgese MG, & Trabace L (2019). Monoaminergic system modulation in depression and Alzheimer's disease: A new standpoint? *Frontiers in Pharmacology*, 10, 483. [PubMed: 31156428]
- Peters AT, Jacobs RH, Crane NA, Ryan KA, Weisenbach SL, Ajilore O, ... Langenecker SA (2017). Domain-specific impairment in cognitive control among remitted youth with a history of major depression. *Early Interv Psychiatry*, 11(5), 383–392. doi:10.1111/eip.12253 [PubMed: 26177674]
- Reitan R (1992). *Trail Making Test: Manual for administration and scoring*. Tucson, AZ: Reitan Neuropsychology Laboratory.
- Roca M, Vives M, Lopez-Navarro E, Garcia-Campayo J, & Gili M (2015). Cognitive impairments and depression: a critical review. *Actas Esp Psiquiatr*, 43(5), 187–193. [PubMed: 26320897]
- Roy-Byrne PP, Weingartner H, Bierer LM, Thompson K, & Post RM (1986). Effortful and automatic cognitive processes in depression. *Arch Gen Psychiatry*, 43(3), 265–267. [PubMed: 3954546]
- Ruff RM, Niemann H, Allen CC, Farrow CE, & Wylie T (1992). The Ruff 2 and 7 Selective Attention Test: a neuropsychological application. *Percept Mot Skills*, 75(3 Pt 2), 1311–1319. doi:10.2466/pms.1992.75.3f.1311 [PubMed: 1484803]

- Sampath D, Sathyanesan M, & Newton SS (2017). Cognitive dysfunction in major depression and Alzheimer's disease is associated with hippocampal-prefrontal cortex dysconnectivity. *Neuropsychiatr Dis Treat*, 13, 1509–1519. doi:10.2147/NDT.S136122 [PubMed: 28652752]
- Schmaal L, Hibar DP, Samann PG, Hall GB, Baune BT, Jahanshad N, ... Veltman DJ (2017). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol Psychiatry*, 22(6), 900–909. doi:10.1038/mp.2016.60 [PubMed: 27137745]
- Schmaal L, Veltman DJ, van Erp TG, Samann PG, Frodl T, Jahanshad N, ... Hibar DP (2016). Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Mol Psychiatry*, 21(6), 806–812. doi:10.1038/mp.2015.69 [PubMed: 26122586]
- Strauss E, Sherman E, & Spreen O (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (Third ed.). New York, NY: Oxford University Press.
- Tabachnick B, & Fidell L (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.
- Tancer ME, Brown TM, Evans DL, Ekstrom D, Haggerty JJ Jr., Pedersen C, & Golden RN (1990). Impaired effortful cognition in depression. *Psychiatry Res*, 31(2), 161–168. [PubMed: 2326395]
- Tombaugh TN (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, 9(3), 260–268.
- Wechsler D (2008). *Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV)*. San Antonio, TX: The Psychological Corporation.
- Zakzanis KK, Leach L, & Kaplan E (1998). On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsychiatry Neuropsychol Behav Neurol*, 11(3), 111–119. [PubMed: 9742509]

Figure 1. Neuropsychological test performance for the Depression groups for “automatic” and “effortful” versions of the A) Ruff 2 & 7 Selective Attention Test Total Speed Index, B) Ruff 2 & 7 Selective Attention Test Accuracy Index, C) Digit Span subtest from the Wechsler Adult Intelligence Scale – IV, and D) Trail Making Test. Scales on the vertical axes reflect raw score values for the respective tests.

Table 1.

Demographic and neuropsychological test performances for each of the Depression groups

	No Depression	Mild-to-Moderate Depression	Severe Depression
<i>n</i>	142	99	44
Age	37.16 (13.8)	40.18 (13.8)	42.09 (13.7)
Education	15.08 (2.4)	14.80 (2.5)	14.30 (2.5)
Sex (<i>n</i>)			
Male	89	50	21
Female	53	49	23
Race (<i>n</i>)			
African American	2	3	0
Hispanic/Latino	10	3	3
Native American	3	1	0
Asian/Pacific Islander	5	1	3
Caucasian	122	91	38
BDI-2	7.49 (4.1)	20.04 (4.1)	34.95 (4.7)

Note: BDI-2 = Beck Depression Inventory –2. Unless labeled otherwise, all values are *Mean (Standard Deviation)*.

Table 2.

Significance levels for main and interaction effects from individual ANOVAs for the cognitive measures administered

	Depression Group Main Effect	Task Difficulty Main Effect	Depression Group * Task Difficulty Interaction Effect
Ruff Total Speed Index	$p = .002$	$p < .001$	$p = .07$
Ruff Accuracy Index	$p = .16$	$p < .001$	$p = .87$
Digit Span	$p = .02$	$p < .001$	$p = .60$
Trail Making Test	$p = .008$	$p < .001$	$p = .005$

Note: ANOVAs = Analyses of Variance.