METACOGNITION AND DEPRESSIVE SYMPTOMOLOGY: USING JUDGMENTS OF LEARNING (JOLs) TO EVALUATE THE DEPRESSIVE REALISM VERSUS NEGATIVITY HYPOTHESES

by

Zehra Haque

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Arts
with a Major in Psychological Research
May 2023

Committee Members:

William Kelemen, Chair

Rebecca Deason

Joseph Etherton

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

I would like to thank my friends, family, and my cat for all of their unwavering support and love. You all have played a significant role in my accomplishments, and I would not have been able to do this without you.

I owe a great amount of my success to my thesis adviser, Dr. Kelemen, and to my committee members, Dr. Deason and Dr. Etherton. Thank you for working with me for the past two years to help me accomplish my goals.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS…….. iv

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF ABBREVIATIONS viii

ABSTRACT ix

CHAPTER

I. INTRODUCTION ……..... 1

II. METHOD ……..... 10

III. RESULTS 14

REFERENCES 32

LIST OF TABLES

Table **Page**

1. Descriptive Statistics for Bias Across Trials……..14

2. Descriptive Statistics for JOL, Recall, and PTPE Measurements……..…...17

3. Descriptive Statistics for Gamma Correlations Across Trials……..…..23

LIST OF FIGURES

Figure **Page**

1. Sample Item from JOL Task……12

2. Mean Item-by-item JOL Bias by Trial……..……19

3. Mean Aggregate JOL Bias by Trial……...……22

LIST OF ABBREVIATIONS

Abbreviation	Description
BDI	Beck Depression Inventory
CES-D	Center for Epidemiologic Studies Depression Scale
GAD-7	Gender Anxiety Disorder-7 item
JOL	Judgment(s) of learning
PTPE	Post-test performance estimate
UWP	Underconfidence-with-practice

ABSTRACT

This study sought to evaluate the validity of the depressive realism and negativity hypotheses, and to determine if an underconfidence with practice effect would be obtained. Several metacognitive judgments were studied, including prospective judgments (item-by-item JOLs and aggregate JOLs) and retrospective judgments (post-test performance estimates). A total of 58 participants studied Swahili-English word translations and were asked to make JOLs before being tested on the material. This study-test period was repeated three times. Depressive symptomology was measured using the CES-D, and individuals were split into low- or high-depressive groups based on their scores. A strong underconfidence with practice effect was obtained for item-by-item JOL bias scores and aggregate JOL bias scores, but not for PTPE bias scores. Generally, participants became underconfident by the second and third study-test period, as is seen with their item-by-item JOL bias and aggregate JOL bias scores. There was no statistically significant evidence found for either hypothesis, but there was evidence of support for the depressive realism hypothesis since high-depressive participants were less underconfident when making item-by-item JOLs. There was no evidence of support for the negativity hypothesis. Several limitations and future directions are discussed to further enhance this study.

I. INTRODUCTION

Metacognition involves the awareness and control of one's own thoughts and cognitions (Dunlosky & Metcalfe, 2008). An individual's capability to think, learn, and process emotions is influenced by both cognitive and metacognitive elements. Metacognitive elements allow people to monitor and control their own cognitive processes. Monitoring is typically studied by eliciting judgments, which can be classified as either retrospective (e.g., confidence judgments) or prospective (e.g., judgments of learning, feelings of knowing). Retrospective judgments concern one's own past performance, whereas prospective judgments refer to one's own future performance (Vaccaro & Fleming, 2018). In general, metacognitive approaches to memory examine how individuals may monitor their memory performance to better improve or optimize their memory performance. Metacognitive control processes allow individuals to properly direct their study time toward heavy or difficult material based on their monitoring (Mazzoni & Cornoldi, 1993). Because metacognitive knowledge and activities involve memory, comprehension monitoring, and strategy selection, a great amount of research revolves around understanding how these strategies contribute to learning and academic performance (Ohtani & Hisasaka, 2018).

Changes in metacognition are often observed in neurological and psychiatric disorders, such as depression, which is described as one of the most prevalent psychiatric disorders (Seow et al., 2021; Kessler et al., 2003). There is an indication that negative metacognitive beliefs (i.e., "I can't control my thinking") are positively associated with symptoms of depression and anxiety (Capobianco, 2020). Individuals with depression tend to believe they perform less well on tasks compared with their non-depressed

counterparts (Dunlosky & Metcalfe, 2008). Despite the widespread occurrence of depression, relatively little is known about the effects of depressive symptoms on metacognitive elements of cognition.

Instances where individuals judge how likely they will be to correctly recall a response on an upcoming test are referred to as judgments of learning (JOLs). JOLs serve a vital role in memory and confidence, and they have been studied extensively over the past 40 years. For example, it has been shown that making JOLs are not merely diagnostic, but that they can influence the items they are intended to assess. Making JOLs can enhance memory for word pairs when a recall test is later administered (Myers et al., 2020). Increasing the number of study trials can change JOL accuracy (Dunlosky & Metcalfe, 2008). Although there is literature pertaining to the effects between depression and metacognition, little is known about the relationship between prospective confidence judgments and depression, particularly after multiple trials. Further research is necessary to fill the gap in this literature and understand how depression may affect a person's ability to make well-calibrated prospective confidence judgments (e.g., JOLs) over multiple trials.

Depressive effects on metacognition

Major depressive disorder (MDD) is a prevalent mental disorder characterized by feelings of sadness and guilt, anhedonia, apathy, and disruptions in sleeping patterns and appetite (Iranpour, 2022). Globally, around 5% of adults suffer from depression (Iranpour, 2022). Rates of depression in the U.S. have increased with the rise of COVID-19; about 14.2% of U.S. adults reported depression in April 2020, a sharp increase from the 8.9% of U.S. adults that reported having depression in 2017-2018 (Daly et al., 2021).

Although cognitive-behavioral therapy (CBT) is the most recommended treatment for depression, it is not completely effective. In one study, only 40-58% of patients receiving CBT recovered at post-treatment, and relapse rates are between 40-60% within a period of two years (Dimidjian et al., 2006; Hollon et al., 2006; Vittengl et al., 2007). Because of this prevalence, there is a need to understand how depressive symptoms can affect various aspects of cognition, including metacognitive judgments.

Depressive Realism Hypothesis. According to the depressive-realism hypothesis, individuals who are depressed are more realistic in their judgments compared with those who are not, whereas nondepressed individuals tend to be overconfident and unrealistically optimistic. This hypothesis states that depressed individuals do not systematically over- or under-estimate the probability of events; they are more accurate and "realistic" (Fu et al., 2005). Previous studies indicate that nondepressed students had rated themselves more positively in terms of social competence compared with ratings from depressed controls (Lewinsohn et al., 1980). However, the fact that there was no objective definition of what is "realistic" was a limitation with this study. Depressed individuals judged themselves more harshly, but it was impossible to determine whether those judgments were more, equal, or less realistic than those made by nondepressed individuals.

Hancock et al. (1996) conducted a study to address this limitation and further examine the depressive-realism hypothesis by examining three separate groups. These groups consisted of individuals who were currently suffering from major depression, individuals who had recovered from depression, and a group of healthy controls. To assess confidence, subjects were asked to answer general knowledge questions and asked

to rate their confidence in their responses. Individuals with higher BDI scores were less confident in their responses, and individuals with lower BDI scores were more confident. Healthy controls were generally more confident in their responses than the depressed group (Hancock et al., 1996). Findings from this study indicate that depressed individuals were less overconfident in their judgments when compared to their nondepressed counterparts. This study is significant because it provided strong support for the depressive-realism hypothesis, which has received additional support (Soderstorm et al., 2011). However, there may be alternative explanations that exist to explain the impact of depression on individuals' confidence levels.

Negativity Hypothesis. One alternative view, known as the negativity hypothesis, states that depressed individuals are overly pessimistic in their evaluations overall. Following this hypothesis, depressed individuals tend to be molded by negative tendencies (Stone et al., 2001). Because of this, their judgments tend to be more negative. In terms of prospective judgments, they tend to display more underconfidence than their non-depressed counterparts, meaning they predict their performance will be poorer than it actually is. With retrospective judgments, depressed individuals assess that they previously performed poorer than they actually did.

Fu et al. (2012) sought to compare the negativity hypothesis and the depressive-realism hypothesis by studying three separate groups: a clinically depressed group, a dysphoric group, and a group of healthy controls. The dysphoric group was comprised of individuals with chronic fatigue syndrome (CFS). The decision to include a group of dysphoric CFS individuals was based on the need to test the validity of depressive realism in individuals with mild depressive symptoms without having been diagnosed

with depression. This study utilized an adjective recognition task. These items were centered around the self, the world, and the future. Following this, participants were asked if they recognized a word previously shown during an encoding period (e.g., 'Weak'). The second test contained the 36 previously presented items as well as 36 new items that were not previously shown to the participant. Following each item, they were asked to make a confidence judgment ranging from 1-6 (1 indicating a total guess, 6 indicating 100% confidence) to define how sure they were in their answers. Upon completion of these tasks, participants were asked to verbally estimate the percentage of questions they believed they answered correctly, which was measured as post-test performance estimate (PTPE).

Rather than using a standard bias calculation, which is calculated by subtracting mean performance from mean judgment, Fu et al. (2012) used a similar calculation by taking the difference between confidence and performance, denoted as absolute accuracy. Depressed participants recognized significantly fewer adjectives when compared to the healthy controls and dysphoric individuals. The groups differed significantly on their over/underconfidence in PTPEs. Depressed individuals showed significantly greater underconfidence than the healthy controls (Fu et al., 2012). Altogether, the depressed participants demonstrated significant item-by-item (i.e., judgments made after each item) overconfidence and, yet, underconfidence in their PTPEs (i.e., cumulative judgments collected after each test). After each item, they believed they performed better than they actually performed; after each test period, they believed they performed worse when considering the cumulative number of items they got correct. Although these results focused on retrospective judgments, the authors concluded that these results were better

aligned with the negativity hypothesis, considering that the greater underconfidence in the depressed participants indicated greater negativity.

To examine self-regulatory judgments (i.e., JOLs made without feedback from the environment) in dysphoric and clinically depressed individuals, Dunn et al. (2007) asked participants to judge whether they were correct or not after completing trials of spatial span working memory tasks. After each trial of the Corsi block tapping spatial span paradigm (Milner, 1971), participants were asked to evaluate their performance by assessing whether they judged each span correctly or incorrectly. Although Study 1 focused on dysphoric participants and Study 2 focused on clinically depressed participants, the researchers' hypotheses remained the same: they proposed that dysphoric and depressed individuals would show a bias towards negativity in their evaluations on the memory span task and that they would believe they performed more trials incorrectly when compared to control participants. In Study 1, results indicated that both the dysphoric and control groups performed equivalently, but judgment accuracy in dysphoric participants was consistent with the negativity hypothesis. Dysphoric participants believed they completed the trials with more mistakes relative to the control group. As found in Study 1, depressed participants in Study 2 were more negative in their assessments of their performance despite having significantly greater scores. Dunn et al. (2007) suggest that demonstrated a negatively biased judgment accuracy regarding their task performance.

Underconfidence with Practice

Lay beliefs about learning often reflect the idea that individuals perform better with more practice or studying. However, overlearning often results in underconfidence in one's performance (Koriat, 1997). Koriat conducted a series of experiments in order to study the influence of multiple study-test trials on accuracy. He found that test performance and JOL accuracy increased across trials. At the beginning of the trials, individuals tend to show overconfidence in their judgments – that is, they tend to make higher JOLs and score lower on the subsequent memory task. If the individuals undergo multiple trials, however, they may display underconfidence (i.e., making lower JOLs but scoring higher on subsequent memory tasks; Koriat, 1997). Over a series of subsequent study-test trials, they displayed underconfidence, known as the underconfidence-with-practice (UWP) effect.

This current research was based on the UWP methodology used by Scheck and Nelson (2005). They conducted a study to test whether a UWP effect would emerge for easy versus difficult items. By testing JOL accuracy across multiple study-test trials and different difficulties of items (easy or difficult) and immediate versus delayed JOLs, Scheck and Nelson (2005) found that the standard UWP effect occurred only for easy items, whether participants made immediate or delayed JOLs (JOLs made after each item were considered immediate, and JOLs made 30 seconds after the item presentation were considered delayed). For easy items, individuals displayed underconfidence across trials, but with difficult items, there was no significant bias toward underconfidence or overconfidence for immediate JOLs. For this reason, the present research used easy items and immediate JOLs and replicated their presentation rate (i.e., 1.5 seconds of study time

per item), in order to maximize the likelihood of obtaining a UWP effect across participants.

Purpose of Research

The purpose of this study is to understand how the severity of symptoms related to the most common psychological disorder, depression, may affect one's ability to feel confident in their responses in a cognitive task. Understanding what makes individuals underconfident and overconfident is vital to recognizing how metacognitive elements vary across individuals in various mental states. Due to the lack of conclusive research surrounding the depressive realism hypothesis and negativity hypothesis, further research is necessary to conduct a comprehensive analysis between them.

This study sought to disentangle predictions made by the depressive realism hypothesis and negativity hypothesis regarding prospective metacognitive JOLs. Given that participants tend to be overconfident in single study-test trial JOL tasks, both hypotheses would make the same prediction (i.e., reduced overconfidence) with only one study trial. Therefore, we selected a multiple study-test trial task in which overconfidence changes to underconfidence within the same participants: the UWP effect. In the case of UWP, the two hypotheses make competing predictions for participants with depressive symptoms. The depressive realism hypothesis suggests that depressive symptoms should be associated with more accurate judgements across all trials. The negativity hypothesis suggests that depressive symptoms should be associated with more accurate JOLs when participants are overconfident (Trial 1) but JOLs should be less accurate when participants are expected to be underconfident (Trials 2 and 3).

This study used the standard item-by-item JOLs employed by Koriat (1997) and Scheck and Nelson (2005). It also used an overall prospective JOL, known as an aggregate JOL, after each study-test trial to see if a similar pattern would emerge. Finally, a type of retrospective judgment, PTPEs, was included to explore the impact of depression on confidence in answers that had been provide (cf., Fu et al., 2012). The latter two types of judgments, aggregate JOLs and PTPEs, have not been tested in the context of a UWP paradigm, and are largely exploratory.

II. METHOD

Participants

Participants were recruited through Texas State University's human subjects pool and received course credit for compensation. A total of 71 participants responded to the survey; 9 participants were excluded for providing blank responses throughout all three trials (N = 62). An additional 4 participants were excluded for providing blank responses during the JOL tasks in trials 2 and 3 (N = 58). Thus, the final sample size for the student was 58 students who completed all portions of the procedures.

The average age of the participants was approximately 19 years old (M = 19.64, SD = 2.41). In terms of gender identity, 74.1% of respondents identified as a woman (n = 43), and 20.7% identified as a man (n = 12). Two participants identified as non-binary or genderfluid, and 1 participant chose not to respond. In regard to ethnicity, 46.6% of participants identified as Hispanic/Latino (n = 27) and the rest reported they were not Hispanic/Latino. In regard to race, 69% of participants identified as White (n = 40), and 6.9% of participants were Black or African American (n = 4), 5.2% of participants were American Indian/Alaskan Native (n = 3), 3.4% of participants were Native Hawaiian or other Pacific Islander (n = 2), and 1.7% of participants were Asian (n = 1).

Materials and Design

Questionnaires. The Center for Epidemiologic Studies Depression Scale (CES-D) was used as a tool to measure participants' recent depressive symptomology. The CES-D is composed of 20 items scored on a 4-point scale. Participants indicated how often they felt depressive symptoms in the past week, with items such as "I felt lonely" and possible responses ranging from "rarely or none of the time (less than 1 day)", "some

or a little of the time (1-2 days)", "occasionally or a moderate amount of time (3-4 days)", or "most or all of the time (5-7 days)" (Radloff, 1977). Consistent with past research, individuals with a total CES-D score under 16 were classified as low-depressive participants, and individuals with a total score over 16 were defined as high-depressive participants. Higher scores indicate greater symptoms, and a cutoff of 16 is considered to be indicative of significant depressive symptoms (Henry et al., 2018).

The General Anxiety Disorder-7 (GAD-7) was used to measure participants' recent anxiety symptoms, as symptoms of anxiety are often comorbid with depressive symptomology (Hirschfeld, 2001). GAD-7 scores were examined to possibly control for this variable. This scale contains 7 items; participants were asked to rate how often they were bothered by certain symptoms over the past 2 weeks, with items such as "Trouble relaxing" and responses ranging from "not at all", "several days", "more than half the days", and "nearly every day" (Spitzer et al., 2006).

JOL Task. Swahili-English word pairs and translations were used for this JOL task. The easiest 20 translations were taken from a norm-referenced list to ensure that memory recall will increase from initially moderate levels over trials to create underconfidence with practice (Nelson & Dunlosky, 1994). Participants were shown a Swahili-English translation on a screen for 1.5 seconds before they were prompted to the next screen. See Figure 1 for an example. Participants were then asked to rate how confident they were that they would correctly recall the translation in 5 minutes, with responses ranging from 0-100% in 20% increments. This rating was used as the item-by-item JOL. After the participant completed the study phase, they were asked to estimate how many items of the 20 they could confidently recall during a test later. This rating

was used as the aggregate JOL. Each participant went through three randomized study-test sessions.

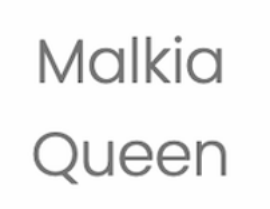

Figure 1
Sample Item from JOL Task
Note. The word presented on the bottom is the English translation of the Swahili word presented above.

Distractor Task. A distractor task was used following the JOL task. Participants were asked to list the names of geographical locations, each starting with the last letter of the previous location (e.g., Houston-Nigeria-Alabama). If they were not able to think of another location, they were instructed to start a new list of locations following a space. The distractor was 3 minutes long. Mean distractor performance was 9.33 (*SD* = 8.22) items on Trial 1, 10.98 (*SD* = 7.41) items on Trial 2, and 10.97 (*SD* = 7.58) items on Trial 3. These levels of performance indicate that participants were engaged during the 3-minute interval. The means did not vary significantly across trials, and so this task is not discussed further, $F(1.77, 101.13) = 3.08$, $p = .056$.

Memory Task. Following the distractor task, participants were asked to correctly translate the word pairs: they saw the Swahili word and were asked to provide the English translation by entering their response in a text box. Obvious spelling errors were counted as correct (generally, if the first 3 letters matched; e.g., umbrela = umbrella). If a participant's response greatly differed in the proper spelling or their response differed in meaning from the correct response (e.g., surprised ≠ suppressed), their response was marked as incorrect. Finally, PTPEs were collected after each study and test trial by

asking participants to provide an estimate of how many items they answered correctly during the test period.

Procedure

This study was conducted remotely using Qualtrics software. Participants were able to complete the procedures on a phone, tablet, or computer. Participants signed up for the study through an electronic system. After registering for the study, participants were directed to the Qualtrics survey, which took approximately 50 minutes to complete. They received course credit for participation.

Data Analysis

Preliminary analyses were conducted to ensure that all assumptions for the appropriate tests were met. Anxiety scores were examined as a potential covariate. The analysis of primary interest involved item-by-item JOL bias scores (mean JOL - mean performance), which indicated participants' absolute over- or under- confidence on each trial. Participants were assigned into either a high- and low- depressive symptom group based on the CES-D cutoff score of 16 reported earlier. For each group, a within-subjects ANOVA was conducted to determine if the magnitude of bias scores changed over the series of trials, and a-priori *t*-tests were conducted to compare each trial to a bias score of 0 to test for over or underconfidence. In addition to Bias scores, Gamma correlations were conducted to test participants' relative accuracy in predicting subsequent performance. Although not directly related to the hypotheses, *G* correlations were calculated to tie the study to the broader literature which often reports this measure.

III. RESULTS

Underconfidence with Practice

Item-by-item JOL Bias. As expected, a substantial UWP effect was obtained for mean item-by-item JOL bias (see top row of Table 1). A one-way, repeated-measures ANOVA was used to examine differences in item-by-item JOL bias across the three study-test sessions, denoted as trials, for all participants. According to Mauchly's test, the assumption of sphericity was violated, $\chi^2(2) = 9.68$, $p = .008$. For this analysis, *df* and *p*-values were adjusted using Huynh-Feldt estimates ($\varepsilon = .887$). Item-by-item JOL bias scores were significantly different across trials, $F(1.78, 101.17) = 27.61$, $p < .001$, $\eta^2 = .33$. Mean JOL bias for Trial 1 ($M_1 = 0.05$) was significantly higher compared to JOL bias for Trial 2 ($M_2 = -0.23$), $F(1, 57) = 47.38$, $p < .001$, $\eta^2 = .45$. JOL bias for trial 2 was significantly lower ($M_2 = -0.23$) than bias for trial 3 ($M_3 = -0.15$), $F(1, 57) = 6.63$, $p = .013$, $\eta^2 = .10$. Finally, JOL bias for Trial 1 ($M_1 = 0.05$) was significantly higher than Trial 3 ($M_3 = -0.15$), $F(1, 57) = 20.85$, $p < .001$, $\eta^2 = .27$. Thus, mean item-by-item JOL bias scores changed significantly across all trials.

Table 1
Descriptive Statistics for Bias Across Trials

Dependent Measure	Trial 1	Trial 2	Trial 3
Bias (Item-by-Item JOLs)	0.05 (0.29)	-0.23 (0.25)	-0.15 (0.23)
Bias (Aggregate JOLs)	-0.02 (0.27)	-0.29 (0.24)	-0.20 (0.27)
Bias (PTPEs)	-0.06 (0.15)	-0.07 (0.15)	-0.04 (0.12)

Note. Main entries are means; entries in parentheses are standard deviations.

A series of one-sample *t*-test were conducted to compare mean scores across trials to 0. Mean item-by-item JOL bias scores for Trial 1 were positive but not significantly different from 0, $t(57) = 1.25$, $p = .215$. However, item-by-item JOL bias scores were negative and significantly less than 0 for Trial 2, $t(57) = -6.97$, $p < .001$, which was a large effect (Cohen's $d = -.915$). JOL bias scores were also different from 0 for Trial 3, $t(57) = -4.89$, $p < .001$, which was a moderate effect (Cohen's $d = -.642$). Because JOL bias scores were significantly less than 0 in Trials 2 and 3, this indicates that participants did show underconfidence with practice. Although mean JOL bias scores for Trial 1 were positive ($M_1 = 0.05$), they were not different from 0, but participants became underconfident by Trial 2 ($M_2 = -0.23$) and Trial 3 ($M_3 = -0.15$) as predicted.

Aggregate JOL Bias. Aggregate JOLs were collected at the end of each study period by asking participants to provide an estimate of the 20 items they believed they would correctly translate during a test later. To determine whether or not the UWP effect extended to another type of prospective judgment, the same analyses were conducted using aggregate JOL bias scores. Mean scores are shown in the second row of Table 1. A one-way, repeated-measures ANOVA was conducted to examine differences in mean aggregate JOL bias across trials. Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(2) = 9.14$, $p = .010$. For this analysis, *df* and *p*-values will be adjusted using Huynh-Feldt estimates ($\varepsilon = .894$). Results indicated that aggregate JOL bias scores were different across trials, $F(1.79, 101.92) = 23.06$, $p < .001$, $\eta^2 = .29$. Aggregate JOL bias scores in Trial 1 ($M_1 = -0.02$) were higher than Trial 2 ($M_2 = -0.28$), $F(1, 57) = 46.60$, $p < .001$, $\eta^2 = .45$. Aggregate JOL bias scores for Trial 2 ($M_2 = -0.28$) were lower compared to Trial 3 ($M_3 = -0.20$), $F(1, 57) = 7.44$, $p = .008$, $\eta^2 = .12$. Finally,

aggregate JOL bias scores for Trial 1 (M_1 = -0.02) were higher than Trial 3 (M_3 = -0.20), $F(1, 57) = 14.38$, $p < .001$, $\eta^2 = .20$.

Further analysis with a one-sample *t*-test indicated that aggregate JOL bias scores for Trial 1 were not different from 0, $t(57) = -0.64$, $p = .527$. However, they were different from 0 for Trial 2, $t(57) = -9.32$, $p < .001$, which was a large effect (Cohen's d = -1.22). They were also different from 0 for Trial 3, $t(57) = -5.67$, $p < .001$, which was a moderate effect (Cohen's d = -.745). These analyses followed a comparable pattern to the item-by-item JOL bias scores, indicating that the UWP effect can be demonstrated with a different type of prospective judgment.

PTPE Bias. For exploratory purposes, similar analyses were conducted with PTPE bias scores (see bottom row of Table 1) to determine if a UWP effect would be evident using retrospective judgments. A one-way repeated-measures ANOVA was conducted to determine if mean differences in PTPE bias scores varied across trials. The assumption of sphericity was violated according to Mauchly's test, $\chi^2(2) = 7.23$, $p = .027$, so the following analysis is reported using adjusted *df* and *p*-values based on Huynh-Feldt estimates ($\varepsilon = .919$). PTPE bias scores were not different across trials, $F(1.84, 104.75) = 1.56$, $p = .216$, $\eta^2 = .03$.

Underconfidence with Practice and Depressive Symptomology

For depression, the mean CES-D score was 19.69 (*SD* = 10.00). Using a cut-off score of 16, 43.1% of participants belonged to the low-depressive symptom group (*n* = 25), and 56.9% of participants belonged to the high-depressive symptom group (*n* = 33). For anxiety, the mean GAD-7 score was 8.10, (*SD* = 4.96). Generally, a score of 5-10

represent moderate anxiety when using the GAD-7 (Spitzer et al., 2006). Participants showed moderate levels of depression and anxiety.

Depression symptoms were not significantly correlated with bias scores in Trial 1, $r(56) = -.013$, $p = .923$, Trial 2, $r(56) = .251$, $p = .057$, or Trial 3, $r(56) = .232$, $p = .159$. Given the co-morbidity often reported between depression and anxiety, Pearson correlations were computed between anxiety scores and JOL bias scores to determine if the former should be used as potential covariate in subsequent analyses. Anxiety scores were not significantly correlated with JOL bias scores in Trial 1, $r(56) = .005$, $p = .970$, or Trial 3, $r(56) = .252$, $p = .056$. JOL bias scores were, however, modestly correlated with anxiety in Trial 2, $r(56) = .289$, $p = .028$. Based on these outcomes, anxiety scores were not included as a covariate in the analyses reported below. For completeness, however, all analyses were conducted both with and without anxiety as a covariate, and no differences were noted in tests of statistical significance.

Table 2 reports the mean scores for the major dependent variables for all participants across trials. Mean item-by-item JOLs and aggregate JOLs were computed across trials, which reflected mean confidence during the learning task (i.e., item-by-item JOLs) and overall confidence at the end of the learning task (aggregate JOLs). In addition, mean recall scores were calculated and mean PTPE scores.

Table 2

Descriptive Statistics for JOL, Recall, and PTPE Measurements Across Trials

Dependent Measure	Trial 1	Trial 2	Trial 3
JOL (Item-by-Item)	0.45 (0.23)	0.46 (0.24)	0.69 (0.28)
JOL (Aggregate)	0.38 (0.21)	0.40 (0.23)	0.65 (0.28)
Recall	0.40 (0.24)	0.69 (0.22)	0.84 (0.21)

PTPE	0.34 (0.20)	0.62 (0.22)	0.80 (0.21)

Note. Main entries are means; entries in parentheses are standard deviations.

Item-by-item JOL Bias. A one-way, repeated-measures ANOVA was conducted to test for differences in mean item-by-item JOL bias scores across trials for both symptom groups (see Figure 2). For the low-depressive symptom group, Mauchly's test indicated that the assumption of sphericity was met, $\chi^2(2) = 5.75$, $p = .057$. Results of the ANOVA indicated that item-by-item JOL bias was significantly different across all three trials[1], $F(2, 48) = 18.85$, $p < .001$, $\eta^2 = .44$. Item-by-item bias for Trial 1 ($M_1 = 0.04$) was significantly higher compared to item-by-item bias for Trial 2 ($M_2 = -0.29$), $F(1, 24) = 33.65$, $p < .001$, $\eta^2 = .58$. Item-by-item bias for Trial 2 ($M_2 = -0.29$) was significantly lower when compared to Trial 3 ($M_3 = -.18$), $F(1, 24) = 6.77$, $p = .016$, $\eta^2 = .22$. Finally, item-by-item bias for Trial 1 ($M_1 = 0.04$) was significantly higher than item-by item bias for Trial 3 ($M_3 = -.18$), $F(1, 24) = 12.15$, $p = .002$, $\eta^2 = .34$.

For the high-depressive symptom group, the assumption of sphericity was not violated according to Mauchly's test, $\chi^2(2) = 4.28$, $p = .118$. Item-by-item bias was significantly different across some trials[2], $F(2, 64) = 10.60$, $p < .001$, $\eta^2 = .25$. Item-by-item bias for Trial 1 ($M_1 = 0.05$) was higher compared to item-by-item bias for Trial 2 ($M_2 = -0.18$), $F(1, 32) = 17.87$, $p < .001$, $\eta^2 = .36$. However, item-by-item bias for Trial 2 was not significantly different from item-by-item bias for Trial 3, $F(1, 32) = 1.62$, $p = .213$, $\eta^2 = .05$. Item-by-item bias for Trial 1 ($M_1 = 0.05$) was significantly higher

[1] With anxiety as a covariate, results were significant with a smaller *F*, larger *p*-value, and a smaller effect size, $F(2, 46) = 15.42$, $p = .009$, $\eta^2 = .18$.
[2] With anxiety as a covariate, results were significant with a smaller *F*, larger *p*-value, and a smaller effect size, $F(2, 62) = 6.91$, $p = .002$, $\eta^2 = .18$.

compared to item-by-item bias for Trial 3 (M_3 = -0.13), $F(1, 32) = 9.02$, $p = .005$, $\eta^2 = .22$.

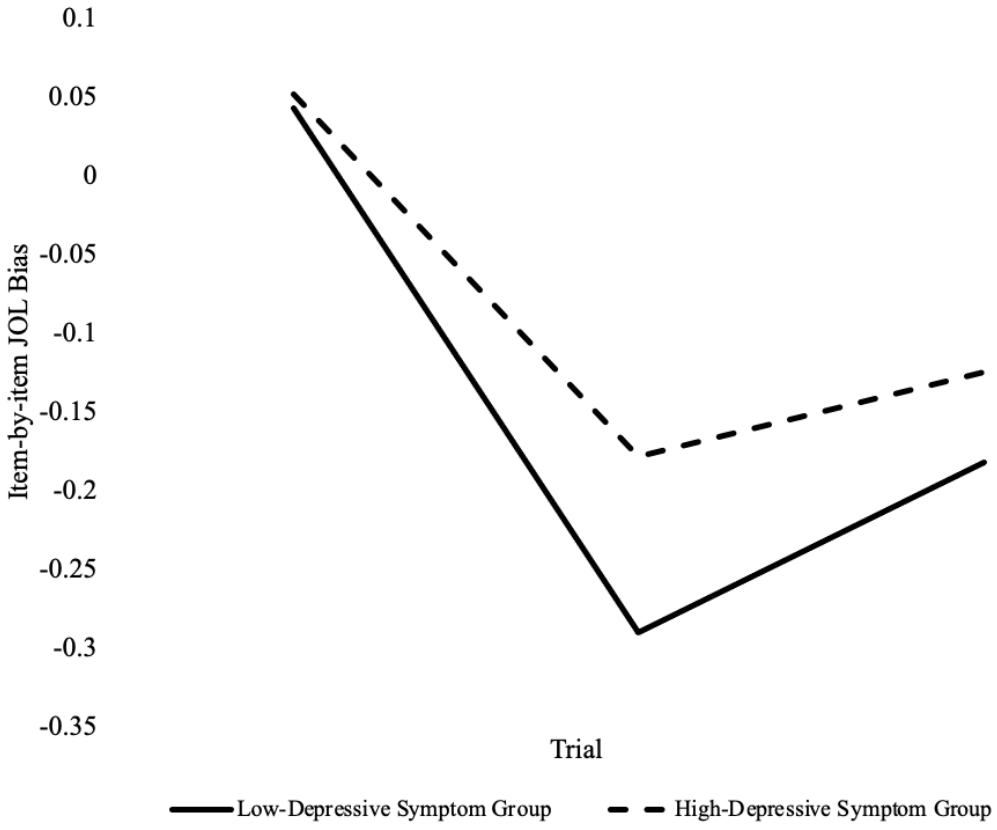

Figure 2
Mean Item-by-item JOL Bias by Trial

One-sample *t*-tests were used to assess whether each group was different from 0 across trials. For the low-depressive symptom group, item-by-item JOL bias was not different from 0 in Trial 1, $t(24) = 0.764$, $p = .452$. Mean JOL bias was significantly different from 0 in Trial 2, $t(24) = -5.94$, $p < .001$, which was a large effect (Cohen's $d = -1.19$). Similarly, JOL bias was significantly different from 0 in Trial 3, $t(24) = -3.95$, $p = .001$, which was a moderate effect (Cohen's $d = -.791$). For the high-depressive symptom group, mean JOL bias was not different from 0 in Trial 1, $t(32) = 0.982$, $p = .334$. JOL bias scores were different from 0 in Trial 2, $t(32) = -4.22$, $p < .001$, which was a moderate

effect (Cohen's d = -.735). Finally, JOL bias scores were significantly different from 0 in Trial 3, $t(32) = -3.05$, $p = .005$, which was a moderate effect (Cohen's d = -.530). Thus, both the low-depressive and high-depressive symptom groups showed a UWP effect. The effect size on Trial 2 was larger for the low-depressive symptom group.

To further explore this UWP effect between groups, and to determine whether the high-depressive symptom participants were significantly less underconfident or overconfident in Trials 2 and 3 compared to low-depressive symptom participants, an independent-samples *t*-test was conducted to compare these trials between the groups. Results indicated that JOL bias scores were not different between the two groups for Trial 2, $t(56) = -1.72$, $p = .090$, or Trial 3, $t(56) = -.926$, $p = .358$. Overall, the pattern of mean item-by-item JOL bias scores between the groups is suggestive of evidence for the depressive-realism hypothesis because high-depressive symptom participants indicated less underconfidence and the low-depressive group showed a larger effect size for their underconfidence on Trial 2, which suggests the high-depressive group may be more realistic in their JOLs. However, this conclusion is not definitive because the two specific comparisons between groups did not reach statistical significance.

Aggregate JOL Bias. Similar analyses were conducted for aggregate JOL bias scores, which showed a similar pattern to aggregate JOL scores (see Figure 3). A one-way repeated-measures ANOVA was conducted to test for differences between mean aggregate-JOL bias scores across trials. For the low depressive symptom group, the assumption of sphericity was met according to Mauchly's test, $\chi^2(2) = 2.43$, $p = .297$. Results of the ANOVA indicated that aggregate bias scores were different across several trials, $F(2, 48) = 13.60$, $p < .001$, $\eta^2 = .36$. Aggregate bias JOL scores from Trial 1 (M_1 =

-0.01) were higher than Trial 2 (M_2 = -0.30), $F(1, 24) = 28.51$, $p < .001$, $\eta^2 = .54$. Aggregate bias scores from Trial 2 were not different from Trial 3, $F(1, 24) = 1.17$, $p = .291$, $\eta^2 = .05$. Finally, aggregate JOL bias scores from Trial 1 (M_1 = -0.01) were higher compared to Trial 3 (M_3 = -0.24), $F(1, 24) = 11.81$, $p = .002$, $\eta^2 = .33$.

For the high-depressive symptom group, the assumption of sphericity was violated, $\chi^2(2) = 7.00$, $p = .030$. For this group, *df* and *p*-values will be adjusted using Huynh-Feldt estimates ($\varepsilon = .872$). Aggregate JOL bias scores were different across the three trials, $F(1.74, 55.80) = 10.50$, $p < .001$, $\eta^2 = .25$. Aggregate JOL bias scores from Trial 1 (M_1 = -0.03) were higher than aggregate bias scores from Trial 2 (M_2 = -0.28), $F(1, 32) = 20.15$, $p < .001$, $\eta^2 = .39$. Aggregate JOL bias scores from Trial 2 (M_2 = -0.28) were lower compared to Trial 3 (M_3 = -0.17), $F(1, 32) = 7.50$, $p = .010$, $\eta^2 = .20$. Finally, aggregate JOL bias scores for Trial 1 (M_1 = -0.03) were higher compared to Trial 3 (M_3 = -0.17), $F(1, 32) = 4.45$, $p = .043$, $\eta^2 = .12$.

To determine whether aggregate JOL bias was different from 0 across trials for both groups, a series of one-sample *t*-tests were used. For individuals in the low-depressive symptom group, aggregate JOL bias scores were not different from 0 in Trial 1, $t(24) = -0.152$, $p = .881$. These scores were different from 0 in Trial 2, $t(24) = -5.77$, $p < .001$, which was a large effect (Cohen's $d = -1.15$). Aggregate JOL bias scores were also significantly different from 0 in Trial 3, $t(24) = -3.86$, $p = .001$, which was a moderate effect (Cohen's $d = -.771$). For the high-depressive symptom group, aggregate JOL bias scores were not different from 0 in Trial 1, $t(32) = -0.637$, $p = .528$. In trial 2, aggregate JOL bias was different from 0, $t(32) = -7.29$, $p < .001$, which was a large effect

(Cohen's d = -1.27). Finally, aggregate JOL bias scores were different from 0 in Trial 3, $t(32) = -4.21$, $p < .001$, which was a moderate effect (Cohen's d = -.732).

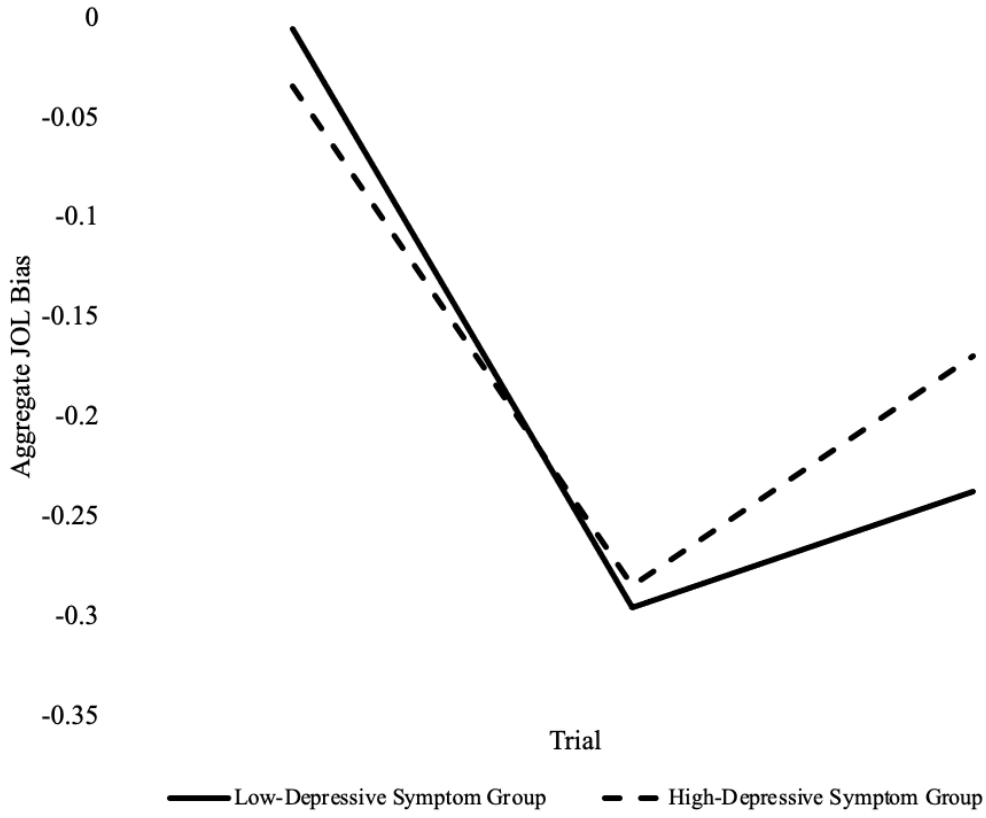

Figure 3
Mean Aggregate JOL Bias by Trial

PTPE Bias. A one-way, repeated-measures ANOVA was conducted to determine differences in PTPE biases in both groups during all three trials. Low depressive-participants had a PTPE bias mean of -0.08 for Trial 1 (*SD* = 0.17), -0.09 for Trial 2 (*SD* = 0.16), and -0.06 for Trial 3 (*SD* = 0.12). For the low-depressive group, Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(2) = 6.41$, $p = .040$. For this test, df and p-values were adjusted using Huynh-Feldt estimates ($\varepsilon = .853$). PTPE bias scores were not significantly different across trials for the low-depressive group, $F(1.71, 40.96) = .263$, $p = .735$, $\eta^2 = .01$. High depressive-participants had a PTPE bias mean of -0.05 for Trial 1 (*SD* = 0.14), -0.06 for Trial 2 (*SD* = 0.14), and -0.02 for Trial 3 (*SD* =

0.12). For the high-depressive group, the assumption of sphericity was met according to Mauchly's test, $\chi^2(2) = 1.69$, $p = .430$. PTPE bias scores were not different across trials for the high-depressive group, $F(2, 64) = 1.64$, $p = .202$, $\eta^2 = .05$.

Item-by-item JOL Gamma. Most of the analyses reported so far have focused on absolute metacognitive accuracy, which is the basis for the UWP effect. However, JOL studies typically report relative measures of metacognitive accuracy as well (*G* correlations; see Table 3). For completeness, a series of one-sample *t*-tests was conducted to determine whether *G* correlations across trials were significantly different from 0. JOL *G*s were found to be different from 0 in Trial 1 [$t(52) = 7.57$, $p < .001$], Trial 2 [$t(51) = 8.05$, $p < .001$], and Trial 3 [$t(31) = 3.51$, $p = .001$]. Similarly, PTPE *G*s were found to be different from 0 in Trial 1 [$t(54) = 20.57$, $p < .001$], Trial 2 [$t(54) = 19.22$, $p < .001$], and Trial 3 [$t(33) = 13.03$, $p < .001$]. In terms of JOL *G*s, these findings are consistent with past literature and indicate that participants were able to predict their relative performance across items at a level greater than chance for all trials; similarly, participants were generally able to successfully evaluate their past performance. Thus, JOL and PTPE *G*s were significantly non-zero in all cases, indicating statistically reliable levels of metacognitive monitoring of relative performance for item-by-item JOLs and PTPEs.

Table 3

Descriptive Statistics for Gamma Correlations Across Trials

Dependent Measure	Trial 1	Trial 2	Trial 3
JOL Gammas	0.48 (0.46)	0.52 (0.46)	0.41 (0.67)
PTPE Gammas	0.85 (0.30)	0.84 (0.32)	0.82 (0.37)

Note. Main entries are means; entries in parentheses are standard deviations.

Discussion

The present research was designed to explore the effects of depressive symptomology on JOLs and test predictions from two theories, the negativity hypothesis and depressive realism hypothesis, in a JOL task across multiple trials. The main hypothesis was that a UWP effect would emerge across trials (Koriat 1997), and that the magnitude of UWP on Trials 2 and 3 would differ by depressive symptomology. If the UWP effect on Trials 2 and 3 were reduced in the high-depressive group, this would support the depressive realism hypothesis (Fu et al., 2012). We found suggestive evidence supporting this, where high-depressive symptom participants collectively reported mean item-by-item JOL bias scores closer to 0, indicating less underconfidence.

If the magnitude of the UWP effect was greater in the high-depressive group, that would indicate greater negativity and be consistent with the negativity hypothesis (Fu et al., 2012). However, there were no instances where the high-depressive group was even more underconfident or more negative with their judgments when compared to the low-depressive group. For both item-by-item JOLs and aggregate JOLs, both groups became underconfident by Trials 2 and 3, but low-depressive participants were more negative with their judgments than high-depressive participants. Thus, there was a complete lack of support for the negativity hypothesis in this study.

Overall, a strong UWP effect emerged, reinforcing the idea that individuals tend to underestimate their performance or ability over a period of several trials. Individuals tend to be overconfident in early trials, but with repeated study and test periods, they underestimate their improvement in memory and become underconfident overall (Koriat 1997). This effect may occur because people do not fully understand the bandwidths of

their improved ability, and because they may feel there is still much to learn to refine their ability. After the test period during Trial 1 and study period during Trial 2, they may realize the limitations of their own knowledge, which leads discount the substantive learning that occurs on Trials 2 and 3 and produce underconfidence.

For low-depressive participants, mean item-by-item bias scores were different across all three trials. During Trial 1, participants were overconfident and became significantly underconfident by Trial 2. By Trial 3, participants became slightly more confident compared to Trial 2 but still remained underconfident overall. When examining differences between Trials 1 and 3, participants became underconfident as trials increased. These findings support a UWP effect overall (Koriat, 1997). High-depressive participants showed a similar pattern: they were overconfident in Trial 1 and became underconfident by Trial 2. Although, there was no difference in confidence between Trials 2 and 3, participants became underconfident between Trials 1 and 3. It is worth noting the effect sizes changed across trials: when examining item-by-item JOL bias scores and aggregate JOL bias scores, the effect size for Trial 2 was generally large and had been reduced to a moderate effect by Trial 3. This pattern indicates that, participants had greater JOL bias scores in Trial 2 than in Trial 3. However, when comparing item-by-item JOL bias scores to 0 between the groups, low-depressive participants had greater effect sizes for Trials 2 and 3; that is, they were significantly lower than 0 compared to high-depressive participants.

These findings provide tentative support for the depressive realism hypothesis. The effect size of the UWP effect on Trial 2 was lower for the high-depressive group and the overall pattern on both Trials 2 and 3 suggest depressive realism may play a role in

the UWP effect for depressed individuals. No evidence was found to support the negativity hypothesis in this cognitive task. If high-depressive participants had shown more negativity than their low-depressive counterparts, this would have been a display of greater negativity, thus supporting the negativity hypothesis (Stone et al., 2001). If high-depressive participants displayed mean JOL bias scores that were more negative and further away from 0, it would have been an indication that high-depressive individuals displayed greater negativity and less realism.

Similar findings emerged using other types of judgments. Aggregate JOLs were collected by asking participants how many items they would remember of the total 20 after each study period. For the low-depressive participants, individuals were slightly underconfident in Trial 1, but became more underconfident in Trial 2, indicating that participants generally remembered more items than they predicted they would in Trials 1 and 2. There was no significant difference between Trials 2 and 3, but participants remained underconfident between Trials 1 and 3. High-depressive participants followed a similar pattern. They were underconfident in Trial 1 and became even more underconfident in Trial 2. By Trial 3, they gained slight confidence but still remained underconfident. These patterns are similar to the findings for item-by-item JOL bias scores, indicating that an UWP effect can be induced while studying aggregate JOLs as well. For both low-depressive and high-depressive participants, there were no differences in PTPE bias scores across trials, indicating that depressive symptoms had no effect on retrospective judgments in this study.

G correlations revealed good relative metacognitive monitoring, which is consistent with previous findings in this task. Because *G* correlations for JOLs were

significantly different from 0 across trials, this indicated that participants were able to distinguish when they would and would not be able to remember the correct translation at a later point. This is consistent with moderate G correlations reported in other JOL studies using judgments immediately after study (e.g., Nelson & Dunlosky, 1991).

Limitations

This study had some limitations that should be considered when interpreting the results. Despite evidence from previous literature for a clinical cutoff score of 16 on the CES-D (Lewinsohn et al., 1980), some researchers have argued that this score is too rigid to determine clinical levels for depression. Others have suggested that a cutoff score of 25 would yield greater sensitivity and specificity for identifying depressive symptoms (Haringsma et al., 2004). This higher cutoff score was not used for this study. Additionally, past literature has indicated that race and gender differences may be worth considering when considering different cutoffs for the CES-D (Callahan & Wolinsky, 1994). For example, women are more likely to report depressive symptoms and tend to have an earlier onset compared to men (Albert, 2015). Separate analyses can be used to consider differences in the presentation of depressive symptomology based on gender. A higher cutoff or a specialized cutoff for differences in race and gender may be something to consider when attempting to replicate this study.

We used a sample that showed moderate levels of depression overall. Results may have varied if participants were recruited based off of previous diagnoses for clinical depression (e.g., Hancock et al., 1996). The difference in using participants who have experienced recent depressive symptoms versus participants who have been screened for

major depressive disorder may impact the results in a significant manner and change our pattern of results.

In addition, this study was administered completely online, outside of a lab setting with no proctor to monitor participants. Because of this, there is no way to know if participants used aids or online translators to assist in their test periods. Additionally, it is unknown if participants were distracted by external events during the study, perhaps adding error variance into our measures.

Clinical Implications

This study examined how depressive symptoms can influence monitoring aspects metacognition. However, the metacognition can also be used to control certain aspects of cognition, and some research suggestions that metacognitive control can be effective in treating depressive symptoms. For example, metacognitive therapy (MCT) has been studied as an alternative to CBT. This treatment approach is based on a metacognitive model. Under this model, psychological disorders stem from an inflexible response pattern to cognitive events labeled the Cognitive Attentional Syndrome (CAS; Wells, 2000; Wells & Matthews, 1994). Within the CAS lies constant worry, threat monitoring, and unsuccessful coping strategies. Negative and positive metacognitive beliefs are both integral parts of the CAS. Positive metacognitions are composed of the benefits of worry and rumination, which are portrayed through statements such as, “Understanding the reason for my sadness will help me feel better.” These positive metacognitions lead to repeated and/or lengthened engagement in ruminative thinking. Negative metacognitions are triggered as the rumination process becomes plagued with distress and/or as a result of what an individual may come to understand about their own depression. These are

seen in patterns that follow the same idea of thoughts such as, “This way of thinking can lead to a mental breakdown” or “My broken brain is the reason behind my negative thoughts.” Negative metacognitions such as these lead to an increase in distress and to unhealthy behaviors which reduce effective coping (Hagen et al., 2017).

To understand the effects of a metacognitive approach as compared to the standard pair of CBT and antidepressant medications, which often lead to relapse or recurrence, Hagen et al. (2017) designed a randomized controlled trial with a group of patients treated with MCT. Participants either received 10 sessions of MCT immediately or were assigned to a 10-week wait list (WL) period. The latter group then received 10 sessions of MCT following the waiting period. The MCT treatment consisted of identifying and expanding meta-awareness by identifying thoughts that may trigger rumination, discovering metacognitive control by learning about attention training, challenging previously held beliefs on the uncontrollability of worry and rumination, challenging beliefs centered around threat monitoring and the dangers of worry and rumination, modification of positive feelings toward worry and rumination, and relapse prevention. Although MCT bears similarities with CBT, the intention behind MCT is to change how patients respond to thoughts by altering metacognitions driven by the CAS (Johnson et al., 2017).

Hagen et al. (2017) found a statistically significant improvement in depressive symptoms from pre to post intervention in both groups when compared to individuals who underwent CBT. Additionally, individuals who underwent MCT did not experience a worsening of symptoms like some of the participants who underwent CBT did (Hagen et al., 2017). The results of this study demonstrated that MCT as treatment for depression

is encouraging, given the large reductions in both depressive and anxious symptoms. According to this study, MCT corresponded to significant improvements in depressive symptoms, indicating that metacognitive elements have some effect on depressive symptoms.

Overall, it is important for research to clearly establish both the impacts of depression on metacognition as well as how metacognition can be used to treat depression. This research has shown what types of judgments are impacted by depression and which are not – namely, this research suggests that prospective judgments are more impacted by depression than retrospective judgments. Still, having a better understanding about what types of judgments are affected by depression gives researchers and clinicians the ability to contemplate alternative possible treatments for depression that may focus on metacognitive elements.

Future Directions

The majority of participants in this study identified as women, and research suggests that women generally tend to be less confident in their judgments when compared with men (Beyer, 1990). Rather than relying on convenience sampling, future researchers may wish to include gender as subject variable to test for differences between women and men. Furthermore, other scales, such as the Beck Depression Inventory, offer additional opportunities to screen for clinical levels of depression (Beck, 1961). This inventory along with other assessments could be used to assess clinical levels of depression or even other psychological disorders. Additionally, it may be worth studying the impact of anxiety since anxiety was found to be correlated with JOL bias scores in Trial 2. Future studies could consider standardizing and combining the scores to examine

the effects of anxiety and depression on JOLs. Finally, although word translation pairs are commonly used in JOL tasks, different cognitive tasks could be considered when exploring these effects. Testing participants on cognitive tasks that can be considered to be more complex, such as remembering the details of a story, could be fruitful.

In conclusion, the UWP effect in this study followed the pattern of past research (e.g., Koriat, 1997). Still, the inferential statistics reported in this study make it difficult to draw firm conclusions about how depressive symptomology may affect confidence and prospective judgments. Additional research will be needed to obtain definitive support for the depressive realism hypothesis in this context. Given how common depression is, and given its potential impact on confidence and decision-making, it is important to determine how the symptoms impact various metacognitive elements.

REFERENCES

Albert, P. (2015). Why is depression more prevalent in women? *Journal of Psychiatry & Neuroscience, 40*(4), 219-221. https://doi.org/10.1503/jpn.150205

Battle, J. (1978). Relationship between depression and self-esteem. *Psychological Reports, 42*(3), 745-746. https://doi.org/10.2466/pr0.1978.42.3.745

Beck, A. T. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561-571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology, 59*(5), 960-970. https://doi.org/10.1037/0022-3514.59.5.960

Callahan, C. M., & Wolinsky, F. D. (1994). The effect of gender and race on the measurement properties of the CES-D in older adults. *Medical Care, 32*(4), 341-356. https://doi.org/10.1097/00005650-199404000-00003

Capobianco, L., Faija, C., Husain, Z., & Wells, A. (2020). Metacognitive beliefs and their relationship with anxiety and depression in physical illnesses: A systematic review. *PLOS One, 15*(9), 1-19. https://doi.org/10.1371/journal.pone.0238457

Daly, M., Sutin, A.R., Robinson, E. (2021). Depression reported by US adults in 2017-2018 and March and April 2020. *Journal of Affective Disorders, 278*, 131-135. https://doi.org/10.1016/j.jad.2020.09.065

Dobson, K., & Franche, R.-L. (1989). A conceptual and empirical review of the depressive realism hypothesis. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 21*(4), 419-433. https://doi.org/10.1037/h0079839

Dunlosky, J. & Metcalfe, J. (2008). *Metacognition*. SAGE Publishing.

Dunn, B. D., Dalgleish, T., Lawrence, A. D., & Ogilvie, A. D. (2007). The accuracy of self-monitoring and its relationship to self-focused attention in dysphoria and clinical depression. *Journal of Abnormal Psychology*, *116*(1), 1-15. https://doi.org/10.1037/0021-843x.116.1.1

Fu, T., Koutstaal, W., Fu, C. H. Y., Poon, L., & Cleare, A. J. (2005). Depression, confidence, and decision: Evidence against depressive realism. *Journal of Psychopathology and Behavioral Assessment*, *27*(4), 243-252. https://doi.org/10.1007/s10862-005-2404-x

Fu, T., Koutstaal, W., Fu, C. H. Y., Poon, L., & Cleare, A. J. (2012). Confidence judgment in depression and dysphoria: The depressive realism vs. negativity hypotheses. *Journal of Behavior Therapy and Experimental Psychology, 43*(2), 699-704. https://doi.org/10.1016/j.jbtep.2011.09.014

Hagen, R., Hjemdal, O., Solem, S., Kennair, L. E. O., Nordahl, H.M., Fisher, P., & Wells, A. (2017). Metacognitive therapy for depression in adults: A waiting list randomized controlled trial with six months follow-up. *Frontiers in Psychology*, *8*(31), 1-10. https://doi.org/10.3389/fpsyg.2017.00031

Hancock, J. A., Moffoot, A. P. R., & O'Carroll, R. E. (1996). "Depressive realism" assessed via confidence in decision-making. *Cognitive Neuropsychiatry, 1*(3), 213-220. https://doi.org/10.1080/135468096396514

Haringsma, R., Engels, G. I., Beekman, A. T. F., & Spinhoven, P. (2004). The criterion validity of the Center for Epidemiological Studies Depression Scale(CES-D) in a sample of self-referred elders with depressive symptomatology. *International Journal of Geriatric Psychiatry, 19*(6), 558-563. https://doi.org/10.1002/gps.1130

Henry, S. K, Grant, M. M, & Cropsey, K. L. (2018). Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample. *Journal of Affective Disorders, 234*, 270-275. https://doi.org/10.1016/j.jad.2018.02.071

Hirschfield, R. M. A. (2001). The comorbidity of major depression and anxiety disorders: Recognition and management in primary care. *The Primary Care Companion to the Journal of Clinical Psychiatry, 3*(6), 244-254. https://doi.org/10.4088/pcc.v03n0609

Iranpour, S., Sabour, S., Koohi, F., & Saadati, H.M. (2022). The trend and pattern of depression prevalence in the US: Data from National Health and Nutrition Examination Survey (NHANES) 2005 to 2016. *Journal of Affective Disorders, 298*(Part A), 508-515. https://doi.org/10.1016/j.jad.2021.11.027

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, J. A., Walters, E. E., Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (NSC-R). *Journal of the American Medical Association, 289*(23), 3095-3105. https://doi.org/10.1001/jama.289.23.3095

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370. https://doi.org/10.1037/0096-3445.126.4.349

Lewinsohn, P. M., Mischel, W., Chaplin, W., & Barton R. (1980). Social competence and depression: The role of illusory self-perceptions. *Journal of Abnormal Psychology, 89*(2), 203-212. https://doi.org/10.1037/0021-843X.89.2.203

Mazzoni, G. & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology, 122*(1), 47-60. https://doi.org/10.1037/0096-3445.122.1.47

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*(3), 272-277. https://doi.org/10.1093/oxfordjournals.bmb.a070866

Moore, D. A. & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes, 103*(2), 197-213. https://doi.org/10.1016/j.obhdp.2006.09.002

Myers, S.J., Rhodes, M.G., & Hausman, H.E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition, 48*(5), 745-758. https://doi.org/10.3758/s13421-020-01025-5

Nelson, T.O. & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2*(3), 325-335. https://doi.org/10.1080/09658219408258951

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "Delayed-JOL Effect." *Psychological Science, 2*(4), 267-271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition Learning, 13*, 179-212. https://doi.org/10.1007/s11409-018-9183-8

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385-401. https://doi.org/10.1177/014662167700100306

Scheck, P., & Nelson, T. O. (2005). Lack of pervasiveness of the underconfidence-with-practice effect: Boundary conditions and an explanation via anchoring. *Journal of Experimental Psychology: General, 134*(1), 124-128. https://doi.org/10.1037/0096-3445.134.1.124

Seow, T. X. F., Rouault, M., Gillan, C. M., & Fleming, S. M. (2021). How local and global metacognition shape and mental health. *Biological Psychiatry, 90*(7), 436-446. https://doi.org/10.1016/j.biopsych.2021.05.013

Spitzer, R.L., Kroenke, K., Williams, J.B.W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166,* 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Stone, E. R., Dodrill, C. L., & Johnson, N. (2001). Depressive cognition: A test of depressive realism versus negativity using general knowledge questions. *The Journal of Psychology, 135*(6), 583-602. https://doi.org/10.1080/00223980109603722

Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, *2*, 1-14. https://doi.org/10.1177/2398212818810591

Vittengl, J. R., Clark, L. A., Dunn, T. W., & Jarrett, R. B. (2007). Reducing relapse and recurrence in unipolar depression: A comparative meta-analysis of cognitive-behavioral therapy's effects. *Journal of Consulting and Clinical Psychology, 75*(3), 475-488. https://doi.org/10.1037/0022-006x.75.3.475

Wells, A., & Matthews, G. (1994). *Attention and emotion: A clinical perspective.* Lawrence Erlbaum Associates, Inc.