Confidence judgment in depression and dysphoria: The depressive realism vs. negativity hypotheses

Tiffany Szu-Ting Fu [a,*], Wilma Koutstaal [b], Lucia Poon [c], Anthony J. Cleare [c,d]

[a] Institute of Biomedical Sciences, Academia Sinica, 6F, No. 16, Alley 10, Lane 437, Pa-The Rd Sec 2, Taipei 10552, Taiwan
[b] Department of Psychology, University of Minnesota, Minneapolis, MN, USA
[c] National Affective Disorders Unit, Bethlem Royal Hospital, Beckenham, Kent, UK
[d] Department of Psychological Medicine, Institute of Psychiatry, King's College London, UK

ARTICLE INFO

ABSTRACT

Background and objectives: According to the negativity hypothesis, depressed individuals are over-pessimistic due to negative self-concepts. In contrast, depressive realism suggests that depressed persons are realistic compared to their nondepressed controls. However, evidence supporting depressive realism predominantly comes from judgment comparisons between controls and nonclinical dysphoric samples when the controls showed overconfident bias. This study aimed to test the validity of the two accounts in clinical depression and dysphoria.
Methods: Sixty-eight participants, including healthy controls ($n = 32$), patients with DSM-IV major depression ($n = 20$), and dysphoric participants with CDC-defined chronic fatigue syndrome ($n = 16$) performed an adjective recognition task and reported their item-by-item confidence judgments and post-test performance estimate (PTPE).
Results: Compared to realistic PTPE made by the controls, patients with major depression showed significant underconfidence. The PTPE of the dysphoric participants was relatively accurate. Both the depressed and dysphoric participants displayed less item-by-item overconfidence as opposed to significant item-by-item overconfidence shown by the controls.
Limitations: The judgment-accuracy patterns of the three groups need to be replicated with larger samples using non-memory task domains.
Conclusion: The present study confirms depressive realism in dysphoric individuals. However, toward a more severe depressive emotional state, the findings did not support depressive realism but are in line with the prediction of the negativity hypothesis. It is not possible to determine the validity of the two hypotheses when the controls are overconfident. Dissociation between item-by-item and retrospective confidence judgments is discussed.

1. Introduction

The cognitive model of depression forwarded by Beck (Beck, 1967, 1987) proposes that depressed individuals view themselves as defective, inadequate, diseased and deprived. As a result of these negative self-concepts, depressed persons believe that they are undesirable and worthless and tend to underestimate or criticize themselves. The "negativity hypothesis" (Clark, Beck, & Alford, 1999; Gilboa-Schechtman, Erhard-Weiss, & Jeczemien, 2002) thus leads to the prediction that depressed persons will be overly pessimistic in their self-referent evaluations (Dunn, Dalgleish, Lawrence, & Ogilvie, 2007; Stone, Dodrill, & Johnson, 2001; Whitton, Larson, & Hauser, 2008). However, several studies (Alloy & Abramson, 1979; Alloy & Ahrens, 1987; Crocker, Alloy, & Kayne, 1988) appear to support the contrary view of "depressive realism". According to depressive realism, depressed persons are neither over-optimistic nor over-pessimistic but rather realistic.

Nevertheless, most studies supporting depressive realism have included only dysphoric/mildly depressed individuals —– as defined by Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) scores between 10 and 19 — and often have used decision tasks for which there is no objectively correct answer (e.g., the contingency judgment paradigm). The results obtained from a mildly depressed population may be different from those demonstrated by severely depressed individuals (Clark et al., 1999) and decision tasks without objectively

* Corresponding author. Tel./fax: +886 2 27219650.
E-mail address: tiffanyfu0821@gmail.com (T. Szu-Ting Fu).

correct answers cannot determine the extent to which an individual is over-optimistic, over-pessimistic or realistic (Haaga & Beck, 1995; Stone et al., 2001).

A further shortcoming of many previous studies involves the inclusion of only one outcome measure or task, on which the healthy controls showed overconfidence. It is not possible to differentially test the validity of the two hypotheses if one only compares the judgment accuracy of depressed or dysphoric versus nondepressed individuals when healthy controls show positive bias in their confidence judgments because both hypotheses make the same prediction under this experimental condition (Dobson & Franche, 1989; Fu, Koutstaal, Fu, Poon, & Cleare, 2005). The two hypotheses make competing predictions only when healthy controls show either accurate judgments or underconfidence (Stone et al., 2001). Under these conditions, the negativity hypothesis predicts that depressed individuals will demonstrate a self-deprecating bias, whereas depressive realism predicts that depressed individuals still will be realistic.

The current investigation sought to address each of these methodological shortcomings. First, a clinically depressed group, a dysphoric group, and matched controls were tested. The dysphoric group was comprised of individuals with chronic fatigue syndrome (CFS). Previous research has shown that the BDI scores of these chronic fatigue patients often indicate dysphoria (Johnson, DeLuca, & Natelson, 1996; Moss-Morris & Petrie, 2001). Inclusion of the dysphoric CFS group was to further test the validity of depressive realism in individuals with mild depressive symptoms in the absence of clinical depression.

Second, we used a recognition memory task for which there were objectively correct answers thereby facilitating the measuring of the degree of judgment accuracy. Thirdly, according to our pilot study, healthy controls ($n = 45$) demonstrated the required experimental condition to contrast the validity of the two hypotheses on this recognition task. Specifically, healthy controls showed overconfidence when judgments were made at an item-by-item level, but underconfidence on a retrospective judgment, that is, a post-test performance estimate (PTPE). Because the healthy controls showed differential judgment-accuracy patterns on the two confidence assessments, both types of judgments were included.

In summary, we used a recognition task in which the healthy controls showed distinct patterns of judgment accuracy to test the validity of the depressive realism versus the negativity hypotheses in two patient groups: individuals experiencing a current episode of major depression, and dysphoric individuals with CFS. Under the situations where the healthy controls showed realistic or underconfident judgment, we hypothesized that (1) the confidence assessment of the individuals with depression and dysphoria should be realistic if the depressive realism account is correct; On the other hand, (2) if the negativity hypothesis is valid, then the depressed patients should demonstrate greater underconfidence compared to the healthy controls.

2. Materials and methods

2.1. Sample

Twenty-three depressed patients who fulfilled DSM-IV-TR criteria for a current episode of major depression, unipolar subtype were invited to participate in the study. Of these, twenty depressed patients gave consent to the study. The depressed patients (inpatients) were recruited from the National Affective Disorders Unit, Bethlem Royal Hospital, Kent. A total of forty-five individuals who fulfilled the Centres for Disease Control (CDC) criteria for CFS (Fukuda et al., 1994) were contacted and twenty-two agreed to take part in the study. Of these, sixteen met the

criterion of dysphoria (a BDI score between 10 and 19). They were recruited from the Chronic Fatigue Research and Treatment Unit, King's College Hospital, London. The healthy controls included 32 volunteers recruited amongst staff and their relatives at South London and Maudsley NHS Foundation Trust.

Exclusion criteria for the depressed, and CFS participants were: history of electroconvulsive treatment in the last year; the presence of neurological disorders, for example, stroke, seizure disorder, etc.; history of head injury with loss of consciousness; and concurrent diagnosis of any of: bipolar disorder, psychosis, current drug or alcohol abuse or dependence, or history of drug or alcohol abuse or dependence within the past 6 months as defined in DSM-IV-TR. Exclusion criteria for the healthy controls were: the presence of psychiatric disorder or drug or alcohol abuse or dependence according to the Patient Health Questionnaire (PHQ) (Spitzer, Kroenke, & Williams, 1999); the presence of neurological disorders, for example, stroke, seizure disorder, etc.; and history of head injury with loss of consciousness. Most of the depressed (19/20) patients were on antidepressants. However, few of the CFS patients (3/22) were on medication for the treatment of CFS.

2.2. Measures

2.2.1. Instruments

The Hamilton Depression Rating Scale (HAM-D) (Hamilton, 1967) and BDI (Beck et al., 1961) were included to assess the emotional state of the participants. The HAM-D is the most frequently used clinician rating scale for measuring symptom severity in patients who have been diagnosed as depressed (Iannuzzo, Jaeger, Goldberg, Kafantaris, & Sublette, 2006). It contains 21 items rated on either a 5-point (0–4) or a 3-point (0–2) scale. The BDI is a 21-item self-report instrument designed to assess the severity of depressive symptoms in the previous two weeks (Beck, Steer, & Brown, 1996). It is rated on a 4-point scale (0–3). Higher scores on the HAM-D and BDI indicate more severe symptoms of depressive disorder.

2.2.2. Adjective recognition task

The trait adjectives for this task were taken from adjectives compiled by Alloy, Greenberg, Clements, and Kolden (1983). The adjectives were based on theoretical accounts of the cognitive schemata concerning the self, the world and the future that are held by persons who are depressed, who are anxious, or neither depressed nor anxious. We combined adjectives from the first two groups to obtain 36 items representing the cognitions held by depressed and anxious individuals versus another 36 items regarding the cognitions held by nondepressed non-anxious (control) individuals.

At encoding, participants were presented 36 pseudo-randomly intermixed items, including 18 control adjectives and 18 depression-anxiety relevant adjectives regarding schemata for the self (12 items), the world (12 items) and the future (12 items). Participants were instructed to evaluate whether or not each item applied to themselves (e.g., "worthy"), their world (e.g., "gloomy") or their future (e.g., "optimistic") and to indicate their answers (i.e., yes/no) using pre-designated keys on the computer keyboard. Thereafter, participants were given the adjective recognition task (see Fig. 1). This old/new recognition test included the 36 previously presented or "old" items and 36 "new" items not presented previously. Assignment of adjectives to new versus old study status was counterbalanced across participants.

2.3. Procedure

This study was approved by the Joint South London and Maudsley and the Institute of Psychiatry NHS Research Ethics
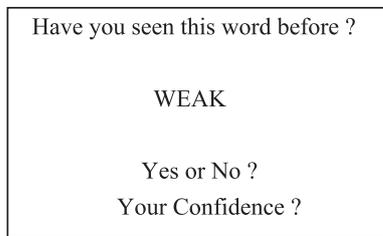
Fig. 1. Adjective recognition task.

Committee. Participants were tested individually during one experimental session of about 40 min. After giving written informed consent, participants completed the BDI and were interviewed with the HAM-D. Then they were verbally instructed as to how to perform the adjective recognition task. This task was presented on a Macintosh laptop and participants pressed pre-designated keys on the computer keyboard to indicate their answers and item-by-item confidence judgment (on a scale from 1 to 6, with 1 indicating guessing and 6 indicating 100% confidence). After the participants completed the task, they were asked verbally to estimate the percentage of questions they thought they answered correctly. Given that one account of the lower levels of confidence that healthy controls show on the retrospective confidence judgment is that participants fail to take into consideration the likelihood of correctly answering by guessing (Liberman, 2004), following the procedure we used in our previous study (Fu et al., 2005), participants were reminded that by simply guessing alone they could expect to correctly answer approximately 50% of the two-alternative forced-choice questions. Lastly, participants were paid a small honorarium.

2.4. Outcome variables

There were four key outcome variables (shown in italics) in the present analysis. We used *proportion correct* to represent average task performance. The abbreviations of "IbI" and "PTPE" (i.e., post-test performance estimate) are used for the item-by-item and retrospective confidence judgments, respectively. The terms *O/U IbI* (i.e., item-by-item over/underconfidence) and *O/U PTPE* (i.e., PTPE over/underconfidence) are the signed difference between confidence and performance. That is, O/U IbI and O/U PTPE measure the degree to which an individual is overconfident, underconfident or realistic.

To further explore the precision of the item-by-item confidence judgments of the three participant groups, the absolute value of the difference between confidence and performance was calculated for each person using the following formula:

$$\text{Absolute accuracy} = \frac{1}{n}\sum_{i=1}^{n}(C_i - P_i)^2$$

Where n = number of items; C_i = confidence judgment about item i; and P_i = performance on item i (Schraw, 2009).

For O/U IbI and O/U PTPE, a value of 0 indicates perfect accuracy, a value greater than 0 indicates overconfidence, and a value less than 0 indicates underconfidence. For item-level absolute accuracy, scores of 0 correspond to perfect accuracy, whereas scores of 1 correspond to total inaccuracy. That is, the closer to 0 the signed and absolute difference scores are, the better judgment accuracy a person has.

2.5. Statistical analyses

The demographic and clinical information of participants were summarized by descriptive statistics and one-way between-groups ANOVAs were used to assess group differences. To determine if there were significant differences in the mean scores on the four key outcome variables across the three groups, four separate one-way between-groups ANOVAs were performed. These overall analyses were followed by planned comparisons to explore where the differences occurred. Next one-sample t-tests against a hypothesized mean of zero were used to determine if there was significant overconfidence or underconfidence in the judgment-accuracy relation for the measures of O/U IbI and O/U PTPE. Finally, given that the adjective recognition task contained both emotional and control stimuli, a mixed between-within subjects ANOVA was performed to determine if there was a valence-dependent effect on the degree of realism shown for the item-by-item confidence. All of the statistics were calculated using SPSS, version 18 (SPSS, Chicago, IL, USA) with an alpha level of 0.05.

3. Results

3.1. Participants

Table 1 shows the mean scores of the demographic and clinical information of the three groups. The three groups did not differ in age ($F[2, 67] = 2.18$, $p = .13$, $\eta^2 = 0.06$), or in years of education ($F[2, 67] = 1.45$, $p = .24$, $\eta^2 = 0.04$). As expected, group differences occurred with regard to HAM-D ratings ($F[2, 67] = 252.79$, $p < .001$, $\eta^2 = 0.89$) and BDI scores ($F[2, 67] = 141.66$, $p < .001$, $\eta^2 = 0.81$). Planned comparison showed that each of the groups significantly differed from each of the others for both the HAM-D (smallest $t = 8.65$, $p < .001$) and the BDI (smallest $t = 7.31$, $p < .001$).

3.2. Task performance

A statistically significant difference was found in proportion correct across the three groups ($F[2, 67] = 6.75$, $p < .01$, $\eta^2 = 0.17$). Planned comparison indicated that depressed patients recognized significantly fewer adjectives than did the healthy controls (mean difference = 0.08, $p < .01$, Cohen's $d = 0.84$), and the dysphoric CFS patients (mean difference = 0.10, $p < .001$, Cohen's $d = 1.24$) (see Table 2).

3.3. Signed difference between confidence and performance

No significant group differences were found for O/U IbI ($F[2, 67] = 2.12$, $p = .13$, $\eta^2 = 0.06$). In contrast, the three groups significantly differed on O/U PTPE ($F[2, 67] = 3.21$, $p < .05$, $\eta^2 = 0.09$). Planned comparison indicated that depressed patients showed significantly greater global underconfidence than did the healthy controls (mean difference = 0.08, $p < .01$, Cohen's $d = 0.61$) (see Table 2).

Table 1
Demographic and clinical characteristics of the healthy control (HC), depressive disorder (DD), and chronic fatigue patient (CFS) groups.

	HC ($n = 32$)	DD ($n = 20$)	CFS ($n = 16$)
Gender			
Male	12 (37.5%)	5 (25.0%)	10 (62.5%)
Female	20 (62.5%)	15 (75.0%)	6 (37.5%)
Age	47.4 (14.1)	54.7 (11.6)	47.1 (13.1)
Education	13.4 (4.0)	12.0 (4.0)	13.8 (2.5)
HAM-D	0 (0.0%)	19.1 (4.4)	7.9 (3.7)
BDI	1.8 (2.3)	30.0 (10.2)	15.5 (2.9)

Note. Standard deviations and percentages are presented in parentheses. HAM-D = Hamilton Depression Rating Scale. BDI = Beck Depression Inventory.

Table 2
Performance and realism of confidence judgment for the healthy control (HC), depressed patient (DD), and chronic fatigue patient (CFS) groups.

	HC ($n = 32$)	DD ($n = 20$)	CFS ($n = 16$)
Proportion correct	0.79 (0.10)	0.71 (0.09)	0.81 (0.07)
Signed difference between confidence and performance			
O/U IbI	0.11 (0.11)	0.06 (0.14)	0.04 (0.10)
O/U PTPE	0.01 (0.12)	−0.07 (0.14)	−0.03 (0.09)
Absolute accuracy	0.17 (0.08)	0.22 (0.06)	0.15 (0.05)
Valence and item-by-item confidence			
O/U IbI for emotional adjectives	0.12 (0.11)	0.06 (0.14)	0.05 (0.10)
O/U IbI for control adjectives	0.11 (0.11)	0.05 (0.15)	0.04 (0.10)

Note. Standard deviations are presented in parentheses. O/U IbI = Item-by-item over/underconfidence. O/U PTPE = over/underconfidence based on post-test performance estimate.

3.4. Absolute accuracy

For absolute accuracy, a statistically significant difference was found (F [2, 67] = 5.66, $p < .01$, $\eta^2 = 0.15$) (see Table 2). Planned comparison indicated that the depressed patients were significantly less accurate than were the healthy controls (mean difference = 0.05, $p < .05$, Cohen's $d = 0.71$) and the dysphoric CFS patients (mean difference = 0.07, $p < .01$, Cohen's $d = 1.27$).

3.5. Judgment-accuracy relation

3.5.1. Healthy controls versus depressed patients

One-sample t-tests against a hypothesized mean of zero showed that, on the O/U IbI, the healthy controls were significantly overconfident ($M = 0.11$, $SD = 0.11$) (t[31] = 5.79, $p < .001$) and the depressed patients were numerically overconfident ($M = 0.06$, $SD = 0.14$) (t[19] = 1.82, $p = .08$). On the O/U PTPE, the healthy controls were relatively realistic ($M = 0.01$, $SD = 0.12$) (t[31] = 0.54, $p = .60$) but the depressed patients were significantly underconfident ($M = −0.07$, $SD = 0.14$) (t[19] = −2.4, $p < .05$).

3.5.2. Healthy controls versus dysphoric CFS participants

On the O/U IbI where the healthy controls showed significant overconfidence, the dysphoric CFS participants demonstrated overconfidence at trend level ($M = 0.04$, $SD = 0.10$) (t[15] = 1.85, $p = .08$). On the O/U PTPE where the healthy controls were neither overconfident nor underconfident, the dysphoric CFS participants displayed relatively realistic judgment ($M = −0.03$, $SD = 0.09$) (t[15] = −1.37, $p = .19$).

3.6. Valence and realism of item-by-item confidence

According to the mixed between-within subjects ANOVA, the interaction effect between type of experimental stimuli (emotional versus control adjectives) and participant group was not significant (F [2, 65] = 0.02, $p = .98$, $\eta_p^2 = 0.001$). The main effect comparing the two types of experimental stimuli was not significant (F [1, 65] = 0.70, $p = .41$, $\eta_p^2 = 0.01$), suggesting there was no valence-dependent effect on the realism of item-by-item confidence. No main effect for group was found (F [2, 65] = 2.19, $p = .12$, $\eta_p^2 = 0.06$) (see Table 2).

4. Discussion

The primary aim of this study was to test the validity of the depressive realism versus negativity hypotheses with regard to the judgment-accuracy relation in an adjective recognition task in three groups: healthy controls, clinically depressed and dysphoric CFS patients. Under the essential experimental precondition where the healthy controls showed realistic retrospective confidence judgments (Fu et al., 2005; Stone et al., 2001), we found that the depressed patients demonstrated significant PTPE underconfidence, which was against the prediction of depressive realism, but consistent with the prediction of the negativity hypothesis. Relatedly, the reduction from item-by-item overconfidence to PTPE underconfidence of the depressed patients compared with the healthy controls is another piece of evidence challenging the validity of depressive realism. If the depressed patients were realistic, the reduction in the level of confidence observed for the item-by-item judgments to that observed for the PTPE judgment would be expected to be smaller for this group than for the healthy controls. However, we found a greater reduction in the depressed patients than the controls. Furthermore, in terms of absolute accuracy for the item-by-item judgments, the depressed patients were not more precise than were the healthy controls.

As stated earlier, the inclusion of a dysphoric group provided an additional opportunity to examine the validity of depressive realism in a group of participants having milder levels of depressive symptoms. In contrast to the negatively biased retrospective judgment displayed by the depressed patients, the dysphoric CFS patients, as did the healthy controls, provided relatively realistic PTPE judgment. In addition, not only was the reduction from item-by-item overconfidence to PTPE judgment of the dysphoric participants smaller than for the controls, but also the absolute accuracy of this group on the item-by-item judgment did not differ from that of the controls. Taken together, the retrospective judgment-accuracy relation observed for the controls compared with the dysphoric CFS patients seems to imply that the depressive realism account may be more suited as a description of the judgment patterns found in individuals with milder levels of depressive symptomatology, rather than clinical depression. Furthermore, based solely on the item-by-item judgment where the healthy controls were significantly overconfident, the present study showed the differential validity testing of the negativity versus depressive realism hypotheses is not possible because the decreased overconfidence pattern demonstrated by the depressed and dysphoric participants for this single measure is predicted by both hypotheses. It is only through the additional information provided by contrasting PTPE judgment of the controls to that of the depressed and dysphoric participants that it could be seen that the depressed patient group showed a negativity bias, whereas the dysphoric group demonstrated comparatively realistic confidence assessments.

On the one hand, the negativity hypothesis is supported by the PTPE findings of the healthy controls versus the depressed patients. On the other hand, depressive realism seems valid for individuals experiencing mild depressive symptoms. However, the item-by-item overconfidence demonstrated by the three groups of participants[1] highlights an issue: Why do confidence judgments differ when they are provided on a post-test basis (in aggregate form) versus on an item-by-item basis?

According to Kahneman and Tversky (1979, 1982), the distinction between item-by-item and aggregate confidence assessments lies in the mode of judgment under uncertainty and the forms of evidence considered. These researchers propose that when making a judgment or decision, individuals can take one of two approaches. On the one hand, an individual can focus predominantly on the immediate case at hand, treating the case as unique and singular, and concentrating on the particular characteristics and considerations that apply to this case. On the other hand, an individual could take a broader view in which the current case is only one instance

---

[1] This finding is not due to the use of emotional words.

of a larger class of similar cases. If an individual adopts this latter, broader view, this person may consider the present case in comparison with other comparable cases and form judgments based on the distributional information about the class of cases. The perspective taken in the first case has been termed an "inside view" (or singular mode) as opposed to an "outside view" (or distributional mode) in the second case.

Because an inside view involves less complex decision-making processes (Sniezek & Buckley, 1991; Sniezek, Paese, & Switzer III, 1990), overconfidence is common when judgment is made on an item-by-item basis. On the contrary, when making aggregate judgments, the consideration of distributional information, which involves greater personal and situational factors, such as one's past experiences with scenarios that are similar to the current one, typically produces a more accurate judgment of performance (Stone, Rittmayer, Murray, & McNiel, 2011; Treadwell & Nelson, 1996).

Previous research has demonstrated that the episodic autobiographical memory of depressed individuals is characterized by a lack of positive memories (Lemogne et al., 2006). The scores of the positive memories of the depressed were significantly lower than those of the controls. When asked to recall important specific memories, in addition, the percentage of negative life events generated by the depressed was significantly higher than that of the controls (Habermas, Ott, Schubert, Schneider, & Pate, 2008). Regarding semantic autobiographical memory, Cox, McWilliams, Enns, and Clara (2004) found that major depression is strongly associated with excessive self-criticism (e.g., "I am no good at all" or "I am incompetent").

Therefore, it is likely that the informational source drawn upon in making the PTPE judgments by the depressed patients is influenced and biased by pervasive negativity. Since self-concept is a valid cue when making the retrospective confidence judgment (Kröner & Biermann, 2007) and aggregate confidence is based on a more general view of one's ability (Stone et al., 2011), the PTPE judgments of the depressed may thus shift away from comparatively accurate judgments that are made by the healthy controls to significant underconfidence. This account can also explain the realistic PTPE judgments provided by the dysphoric CFS group. Whereas the depressed patients possess negative self-concepts, the dysphoric individuals with CFS reported significantly less self-reproach symptoms (e.g., feeling like a failure, disappointment with self, and self-blame, etc.) (Hawk, Jason, & Torres-Harding, 2006). Presumably due to their relatively well-preserved self-concept, PTPE over/underconfidence of the dysphoric CFS individuals was more similar to that shown by the healthy controls, thereby producing a "dysphoric" effect on the realism of judgment.

This study has some limitations that warrant consideration when the findings are interpreted. First, our findings are limited by the fact of the inclusion of an episodic memory task. It remains to be determined whether the judgment-accuracy patterns for the groups investigated here will be observed for tasks from different content domains. Second, given that men are sometimes more confident than are women (e.g., Beyer, 1990, 1999; Weiss, Kemmler, Deisenhammer, Fleischhacker, & Delazer, 2003), further research is needed to validate the current findings with a design matched on gender. Third, the current evidence pointing to depressive realism in the dysphoric CFS group should be interpreted cautiously given that these findings were obtained with a relatively small number of dysphoric participants and with a response rate of about 50%. It is also possible that dysphoric states differ between patient groups, and that CFS patients with mild depressive symptoms have a different set of underlying cognitive correlates than other patients with similar levels of dysphoria. Therefore, the validity of depressive realism needs to be replicated in studies comparing healthy controls and larger samples of individuals with dysphoria. As stated in the introduction, such further studies must attempt to meet the following experimental preconditions to provide a valid test of the two hypotheses, including: (a) a clinically depressed group, (b) a decision task or paradigm with objectively correct answers, on which (c) healthy controls are either underconfident or realistic on at least one outcome measure.

In conclusion, the present study found some evidence consistent with the depressive realism account. However, the validity of depressive realism appears to be limited to individuals with mild depressive symptoms. Toward a more severe depressive emotional state, the findings were contrary to the prediction of depressive realism but in line with the negativity hypothesis. The current investigation also demonstrated distinct evaluation patterns between item-by-item and global-level confidence judgments and provided a theoretical explanation for these differences.

References

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: sadder but wiser? *Journal of Experimental Psychology: General, 108*, 441–485.

Alloy, L. B., & Ahrens, A. H. (1987). Depression and pessimism for the future: biased use of statistically relevant information in predictions for self versus others. *Journal of Personality and Social Psychology, 52*, 366–378.

Alloy, L.B., Greenberg, M.S., Clements, C., & Kolden, G. (1983). Depression, anxiety, and self schemata: a test of Beck's theory. Paper presented at the meeting of the American Psychological Association, (Anaheim, CA).

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Harper & Row.

Beck, A. T. (1987). Cognitive models of depression. *Journal of Cognitive Psychotherapy, 1*, 5–37.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression Inventory-II*. San Antonio, SA: TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluation of performance. *Journal of Personality and Social Psychology, 59*, 960–970.

Beyer, S. (1999). Gender differences in the accuracy of grade expectations and evaluations. *Sex Roles, 41*, 279–296.

Clark, D. A., Beck, A. T., & Alford, B. A. (1999). *Scientific foundations of cognitive therapy and therapy of depression*. New York, NY: Guilford Press.

Cox, B. J., McWilliams, L. A., Enns, M. W., & Clara, I. P. (2004). Broad and specific personality dimensions associated with major depression in a nationally representative sample. *Comprehensive Psychiatry, 45*, 246–253.

Crocker, J., Alloy, L. B., & Kayne, N. T. (1988). Attributional style, depression, and perceptions of consensus for event. *Journal of Personality and Social Psychology, 54*, 840–846.

Dobson, K., & Franche, R. L. (1989). A conceptual and empirical review of the depressive realism hypothesis. *Canadian Journal of Behavioural Science, 21*, 419–433.

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*, 1–15.

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*, 243–252.

Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., & Komaroff, A. (1994). The chronic fatigue syndrome: a comprehensive approach to its definition and study. *Annals of Internal Medicine, 121*, 953–959.

Gilboa-Schechtman, E., Erhard-Weiss, D., & Jeczemien, P. (2002). Interpersonal deficits meet cognitive biases: memory for facial expressions in depressed and anxious men and women. *Psychiatry Research, 113*, 279–293.

Haaga, D. A. F., & Beck, A. T. (1995). Perspectives on depressive realism: implications for cognitive theory of depression. *Behaviour Research and Therapy, 33*, 41–48.

Habermas, T., Ott, L. M., Schubert, M., Schneider, B., & Pate, A. (2008). Stuck in the past: negative bias, explanatory style, temporal order, and evaluative perspectives in life narratives of clinically depressed individuals. *Depression and Anxiety, 25*, E121–E132.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6*, 278–296.

Hawk, C., Jason, L. A., & Torres-Harding, S. (2006). Differential diagnosis of chronic fatigue syndrome and major depressive disorder. *International Journal of Behavioral Medicine, 13*, 244–251.

Iannuzzo, R. W., Jaeger, J., Goldberg, J. F., Kafantaris, V., & Sublette, M. E. (2006). Development and reliability of the HAM-D/MADRS Interview: an integrated depression symptom rating scale. *Psychiatry Research, 145*, 21–37.

Johnson, S. K., DeLuca, J., & Natelson, B. H. (1996). Depression in fatiguing illness: comparing patients with chronic fatigue syndrome, multiple sclerosis and depression. *Journal of Affective Disorders, 39*, 21–30.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: biases and corrective procedures. *Management Science, 12*, 313–327.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition, 11*, 143–157.

Kröner, S., & Biermann, A. (2007). The relationship between confidence and self-concept: towards a model of response confidence. *Intelligence, 35*, 580–590.

Lemogne, C., Piolino, P., Friszer, S., Claret, A., Girault, N., Jouvent, R., et al. (2006). Episodic autobiographical memory in depression: specificity, autonoetic consciousness, and self-perspective. *Consciousness and Cognition, 15*, 258–268.

Liberman, V. (2004). Local and global judgments of confidence. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 30*, 729–732.

Moss-Morris, R., & Petrie, K. J. (2001). Discriminating between chronic fatigue syndrome and depression: a cognitive analysis. *Psychological Medicine, 31*, 469–479.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition & Learning, 4*, 33–45.

Sniezek, J. A., & Buckley, T. (1991). Confidence depends on level of aggregation. *Journal of Behavioral Decision Making, 4*, 263–272.

Sniezek, J. A., Paese, P. W., & Switzer, F. S., III (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46*, 264–282.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & , the Patient Health Questionnaire Primary Care Study Group. (1999). Validity and utility of a self-report version of Prime-MD. *Journal of American Medical Association, 282*, 1737–1744.

Stone, E. R., Dodrill, C. L., & Johnson, N. (2001). Depressive cognition: a test of depressive realism versus negativity using general knowledge questions. *Journal of Psychology, 135*, 583–602.

Stone, E. R., Rittmayer, A. D., Murray, N. T., & McNiel, J. M. (2011). Demonstrating and eliminating the aggregation effect in physical skill tasks. *Journal of Behavioral Decision Making, 24*, 1–22.

Treadwell, J. R., & Nelson, T. O. (1996). Availability of information and the aggregation of confidence in prior decisions. *Organizational Behavior and Human Decision Processes, 68*, 13–27.

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences, 35*, 863–871.

Whitton, S. W., Larson, J. J., & Hauser, S. T. (2008). Depressive symptoms and bias in perceived social competence among young adults. *Journal of Clinical Psychology, 64*, 791–805.