Cognitive Therapy and Research (2024) 48:225–241
https://doi.org/10.1007/s10608-023-10454-w

ORIGINAL ARTICLE

Reduced Belief Updating in the Context of Depressive Symptoms: An Investigation of the Associations with Interpretation Biases and Self-Evaluation

Felix Würtz[1] · Tobias Kube[2,3] · Marcella L. Woud[1,3] · Jürgen Margraf[1] · Simon E. Blackwell[1,3]

Accepted: 15 November 2023 / Published online: 12 January 2024

Abstract

Background Depressive symptoms are associated with negative expectations and reduced belief updating by positive information. Cognitive immunization, the devaluation of positive information, has been argued to be central in this relationship and predictive processing models suggest that more positive information is associated with greater cognitive immunization.
Methods In an online experiment, $N = 347$ healthy participants took part in a performance task with standardized feedback of varying levels of positivity (mild, moderate, extreme). Effects of feedback positivity on cognitive immunization were investigated. Further, depressive symptoms, interpretation bias and participant's self-evaluation were examined as potential correlates of belief updating.
Results As expected, participants receiving mildly positive feedback reported a greater amount of cognitive immunization than those receiving moderately positive feedback. However, neither group differed from those receiving extremely positive feedback. Although depressive symptoms did not show the hypothesized association with cognitive immunization, they were associated with a weaker increase in positive expectations following feedback. Exploratory analyses showed associations between self-evaluation and belief updating.
Conclusions The results suggest that healthy participants engaged in cognitive immunization when feedback was less positive than expected. Depressive symptoms were associated with reduced belief updating, but not with cognitive immunization. Self-evaluation may be a promising factor for future research.

Keywords Predictive processing · Interpretation bias · Depression · Belief updating

Introduction

Negative expectations of future situations are a core component of cognitive models of depression (e.g., Beck & Haigh, 2014). That is, these models propose that individuals with depression have a tendency to expect future situations to turn out relatively negatively compared to non-depressed individuals. Supporting this assumption, there is substantial evidence that the presence of negative expectations and a lack of positive expectations are strongly linked with depressive symptoms and their relapse (Cane & Gotlib, 1985; Gopinath et al., 2007; Horwitz et al., 2017; Kube et al., 2017; Zetsche et al., 2019). Further, negative expectancies appear to persist even in the face of contradictory positive

✉ Felix Würtz
felix.wuertz@rub.de

Tobias Kube
kube@uni-landau.de

Marcella L. Woud
marcella.woud@uni-goettingen.de

Jürgen Margraf
juergen.margraf@rub.de

Simon E. Blackwell
simon.blackwell@uni-goettingen.de

[1] Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum and German Center for Mental Health (DZPG), Partner Site Bochum/Marburg, Massenbergstrasse 9-13, 44787 Bochum, Germany

[2] Department of Clinical Psychology and Psychotherapy, RPTU Kaiserslautern-Landau, Landau, Germany

[3] Present Address: Department of Clinical Psychology and Experimental Psychopathology, Georg-Elias-Mueller-Institute of Psychology, University of Göttingen, Göttingen, Germany

information (e.g., Korn et al., 2014). This persistence, in turn, could plausibly be an important factor underlying the maintenance of depressive symptoms, and as such advancing our understanding of their underlying mechanisms have gained increasing scientific attention (Rief & Joormann, 2019).

Recent advances in depression theory formulation have adopted a *predictive processing* perspective, taking into account that people's expectations are strongly intertwined with the perception of novel information (Barrett et al., 2016; Clark et al., 2018; Kube et al., 2020). Appealing to the hypothesis of a "Bayesian brain" (Adams et al., 2015; Friston et al., 2014; Paulus et al., 2019), these accounts regard expectations as prior predictions to which newly incoming information is compared. In depression, these priors are postulated to be profoundly negative and to be held with certainty (e.g., *I am certain that I will fail*). Unexpected positive information, in contrast, is assigned little precision (i.e., confidence), resulting in little or no updating of an expectation, despite the discrepancy between the prior prediction and the new information (i.e., the prediction error). Psychologically, this process has been referred to as 'cognitive immunization' (Rief et al., 2015), meaning that people cognitively devalue disconfirming evidence and thus uphold their initial expectations. For example, a depressed individual having the expectation that they will fail at whatever they do may devalue unexpectedly positive feedback from a performance test by thinking, "*This time I was good in this test, but this was just luck. Next time I will fail*".

Evidence for these proposals comes from several strains of research. For instance, it has been shown that in depression negative expectations persist despite positive feedback (Kube & Glombiewski, 2021; Kube et al., 2019a, 2019b), with congruent findings for the persistence of negativity in other (cognitive) constructs, such as negative interpretation biases (Everaert et al., 2018, 2020), pessimism (Korn et al., 2014), or blunted emotional response to social acceptance (Caouette & Guyer, 2016). Further, there is emerging evidence that cognitive immunization towards positive information might play a central role in limited belief updating in both clinically (Kube et al., 2019a, 2019b) and non-clinically depressed samples (Kube et al., 2022a, 2022b).

However, other recent studies have provided a more mixed picture. For instance, Kube (2023) did not find a significant association between depressive symptoms and reduced belief updating in the context of positive social interactions. Similarly, in a non-clinical sample, Gagne et al. (2022) found depression to be associated only with more negative prior beliefs, but not with any abnormality in belief updating. Feldmann et al. (2022), did find depressive symptoms to be related to more negative belief updating than a statistical solution balancing negative expectations and the positivity level of incoming information would suggest. However, in contrast to the predictive processing account of depression (Kube et al., 2020), they found no evidence that higher levels of depression were associated with assigning inflated precision to prior predictions.

Another aspect that has recently been considered in relation to belief updating and depression is the relationship between the magnitude of the prediction error and updating. In particular, Kube et al. (2022a, 2022b) proposed that this relationship may not be linear, as assumed in traditional learning models (e.g., Rescorla & Wagner, 1972), but may involve a "tipping point", above which the discrepancy between prior expectations and extremely positive feedback is suspiciously large (e.g., receiving the feedback, "*You are the most intelligent person in the world*"), so that the engagement in cognitive immunization increases and the degree of expectation change decreases. To test this, Kube et al. (2022a, 2022b) presented healthy participants with a standardized performance test in which they had to anticipate people's emotions in hypothetical scenarios. Participants were then randomly assigned to one of three standardized feedback conditions saying that they were either among the best 50% (termed mildly positive), the best 10% (termed moderately positive), and the best 1% (termed extremely positive). In line with their hypothesis, Kube et al. (2022a, 2022b) found that extremely positive feedback was perceived as less credible than mildly and moderately positive feedback, with the greatest credibility being attributed to the moderately positive feedback. However, surprisingly, the greatest cognitive immunization was reported in response to the mildly positive feedback. Further, feedback credibility did not fully translate to belief updating as there was only a significant difference between the mildly and the extremely positive feedback condition, meaning that the moderately positive feedback which was perceived as most credible did not lead to the largest belief updating. In the context of unexpectedly positive social feedback, however, some evidence has been found that moderately positive feedback leads to the greatest amount of belief updating, and this finding was replicated in another independent study (Kube, 2023).

Altogether, the extant literature has found some associations between depressive symptoms and belief updating, but it is unclear how robust these associations are across samples and task material. Furthermore, although there is some supporting evidence for cognitive immunization to be an important mechanism underlying the lack of belief updating (e.g., Kube, 2023), other research has provided inconclusive findings (e.g., Kube & Glombiewski, 2021; Kube et al., 2022a, 2022b). One way to explain these mixed findings may be the presence of various factors relevant to the relationship that have not yet been the focus of the research conducted so far, but which could play an important role in explaining the relationship between belief updating, cognitive immunization, and depression.

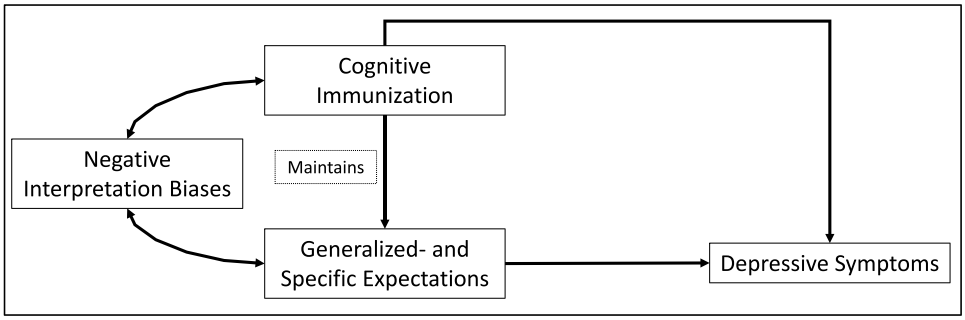

Fig. 1 Hypothetical schematic illustration of the potential relationships between interpretation biases, cognitive immunization, belief updating, and depressive symptoms. Adapted from Kube et al. (2020). Cognitive immunization is proposed to maintain depressive symptoms through a consistent devaluation of positive information that leads to dysfunctional behavior such as social withdrawal. Further, it is proposed that cognitive immunization maintains the presence of negative expectations and the absence of positive expectations through a devaluation of disconfirmatory experiences. These expectations, in turn, are postulated to maintain depressive symptoms. Both, cognitive immunization and expectations can be seen as either a specific instance of negative interpretation biases or processes that might be influenced by a general tendency to interpret information negatively

A useful starting point for investigating these is to examine well-established cognitive factors associated with depression, and that have ecological validity in terms of their operating properties. One factor for which research has delivered compelling evidence for a central role in the development and maintenance of depression is the tendency to consistently interpret ambiguity in a negative manner, namely a negative interpretation bias (IB; O'Connor et al., 2021; Rude et al., 2010; for a meta-analysis, see Everaert et al., 2017a, 2017b). In terms of the combined cognitive bias hypothesis (Everaert & Koster, 2020; Everaert et al., 2014) it has been postulated that IBs do not act in isolation, but lead to depression through the interplay with other cognitive biases as moderators (cf. Gadassi Polack et al., 2023). As of yet, the focus of combined cognitive bias research has largely been on the association between IBs and biases in attention and memory, with a neglect of other biased cognitive processes relevant to depression, such as overtly negative expectations and their limited updating (Everaert & Koster, 2020). Investigating the association between IBs and belief updating is promising as a potential link between negative expectations, IBs, and cognitive immunization can be proposed based on the overlap in their definitions and operating properties. To illustrate, as negative expectations are directed towards the future, whether they will later be confirmed or disconfirmed is to some degree uncertain. For instance, taking an upcoming exam about which one has only limited information is at first an ambiguous situation that is subject to the individual's interpretation (e.g., as an opportunity for failure or success). Based on this conceptualization, negative expectations and IBs both involve making inferences about uncertain information, albeit negative expectations are specifically future-oriented.

Further, as cognitive immunization is defined as the tendency to evaluate novel incoming information in a negative manner to fit with the initially negative expectation, and thus as a form of reappraisal, an association between IBs with cognitive immunization appears plausible, given that IBs have been found to be associated with reduced positive and increased negative reappraisal (Blanco et al., 2021; Everaert et al., 2017a, 2017b, 2020). To date, the association between limited belief updating, cognitive immunization and IBs in the context of depression is unexplored. However, the investigation of this association could shed light on the question of how these cognitive factors in concert are associated with depressive symptoms and if the persistence of both IBs and negative expectations in this context might be attributable to the same underlying processes (see Fig. 1 for a graphical depiction).

The goal of the present study was therefore two-fold: First, to conceptually replicate previous findings by Kube et al. (2022a, 2022b) concerning the relationship between symptoms of depression, belief updating, and cognitive immunization to shed light on the robustness and replicability of results. Second, to investigate the relationships between IBs, negative expectations, belief updating, and cognitive immunization. To follow these aims we applied the study design by Kube et al. (2022a, 2022b) and conducted an online study in which we informed participants that they would solve a difficult performance task. Afterwards, we assessed symptoms of depression, initial negative expectations towards the performance task, and negative IBs. This was followed by a performance task for which it is difficult to judge the level of one's own success, after which participants were randomized to receive either mildly, moderately, or extremely positive standardized feedback, independently of their actual performance. Subsequently, participants' self-rated performance, the extent to which participants adjusted their initial expectations towards the task, as well as the extent to which

participants engaged in cognitive immunization strategies concerning the feedback were assessed.

Based on our first aim to conceptually replicate the study by Kube et al. (2022a, 2022b), we hypothesized that participants receiving mildly positive feedback would report the greatest engagement in cognitive immunization strategies, followed by the extremely positive condition, and then by the moderately positive feedback condition. We further explored whether depressive symptoms would moderate this effect, namely whether depressive symptoms would be associated with the strength of the conditions' effect. Next, we expected to find that irrespective of the level of depressive symptoms, participants' performance expectations would become more positive in response to the feedback. Further, we expected to find this effect irrespective of the feedback condition. This hypothesis was formulated partly in contrast with findings in the original study, as the previous results (Kube et al., 2022a, 2022b) concerning the effects of the feedback manipulation on belief updating had been heterogenous and thus we decided to hypothesize in line with the more conservative prediction that there would be no relationship. Regarding our second aim to extend previous findings by investigating the role of IBs, we hypothesized that depressive symptoms would be associated with more negative expectations. In addition, we expected a positive association between these variables (depressive symptoms and negative expectations) and a negative IB. Further, we expected that a negative IB would explain unique variance of cognitive immunization following feedback, after taking the different conditions and depressive symptoms into account.

Methods

Pre-Registration, Power Analysis, and Open Material

Prior to the start of data collection, the study was pre-registered by uploading the study protocol to the Open Science Framework (OSF; https://osf.io/6bx4v/) and all deviations from the pre-registration are labelled as such. Sample size was determined a-priori using G*Power (Faul et al., 2009) and aiming for 80% power to detect a small to medium effect size ($f^2 = 0.03$) of a single regression coefficient in a multiple-regression model with eight predictors in total at $p < 0.05$, resulting in a sample size of $N = 264$. To account for incomplete or odd data given the online setting we aimed to overrecruit ~ 30% resulting in a final sample size of $N = 350$. Anonymized data generated during the study and used analysis scripts are openly available on the OSF (https://osf.io/6bx4v/).

Participants

The study was conducted as an online survey and participants were recruited via social media, the flyer walls of the Faculty of Psychology at Ruhr-University Bochum and at the University Koblenz-Landau, and by distributing the study link to patient representation organizations. All participants with a minimum age of 18 years were eligible for participation and there were no other in- or exclusion criteria. The study was approved by the local ethics committee of the Faculty of Psychology at Ruhr-University Bochum (approval No. 742) and all participants provided written informed consent by agreeing to an online form. As a reimbursement, participants could take part in a voucher lottery for a 30€ multi-purpose voucher. In total, $N = 350$ participants were recruited between December 2021 and June 2022. Of those, three participants had to be excluded as they reported being under 18. The final sample on which all analyses were based was therefore $N = 347$.

Materials

Performance Task with Feedback Manipulation

The Test of Emotional Intelligence (TEMINT; Schmidt-Atzert & Bühner, 2002), which assesses participants' ability to predict people's emotional response to a variety of situations, was used as a performance task. In this task, participants are presented with 12 different situations (e.g., a thirty-year-old student working in tech-support has to solve a computer problem in front of a costumer) and participants are instructed to rate the strength of certain emotions that the person might be experiencing. In line with previous research applying the TEMINT in the context of belief updating in depression (Kube et al., 2018, 2019a, 2019b, 2022a, 2022b) participants' initial expectations of their success in the task were lowered, by introducing the TEMINT as a really difficult task that can only be solved by few people. Following the TEMINT, participants were first asked for their self-evaluation of how well they did in the task on a 7-point scale ranging from 1 (*very bad*) to 7 (*very good*). After that, they were randomly allocated to one of three feedback conditions: Mildly positive, moderately positive, or extremely positive feedback, receiving the information that they were among the 50%, 10% or 1% best participants, respectively. Importantly, previous research has shown that participants' actual performance on the TEMINT is uncorrelated with their self-rated performance, indicating that participants are unaware of their actual task performance (Kube et al., 2018). Furthermore, elevated levels of depression are unrelated to performance deficits in the TEMINT (Kube et al., 2022a, 2022b). Allocation to the feedback conditions was stratified for depressive symptoms (low [QIDS < = 5] vs.

middle [5 < QIDS > = 15] vs. high [QIDS > 15] symptom level) and gender (female vs. non-female including male, non-binary and no answer). To illustrate, participants with a specific symptom level and gender were automatically and evenly distributed to the three feedback conditions, i.e., the first female participant with low depressive symptoms was allocated to the mild feedback condition, the second to the moderate, and the third to the extreme condition.

Questionnaires

Demographic Information

Participants were asked to indicate their age, gender, national identity, whether they were a native German speaker, their language fluency rated on a 5-point scale ranging from 1 (*very insecure*) to 5 (*very secure*), their educational and occupational background, whether they were currently diagnosed with a mental disorder, and whether they were currently in psychotherapy or on psychotropic medication.

Interpretation Biases

IBs were assessed via the German translation of version A of the Ambiguous Scenario Task for Depressive Mood—II (AST; Rohrbacher & Reinecke, 2014). The AST consists of 15 ambiguous everyday life scenarios and participants are instructed to briefly imagine being in each scenario and then rate its pleasantness on a scale ranging from − 5 (*very unpleasant*) to 5 (*very pleasant*). Internal consistency in the present study was α [95%-CI] = 0.83 [0.80, 0.86].

Depressive Symptoms

Depressive symptoms were assessed via the German self-report version of the Quick Inventory for Depressive Symptomatology–16 (QIDS; Roniger et al., 2015; Rush et al., 2003). The QIDS consists of 16 items assessing nine different symptom groups relevant for depression. Each symptom group is assessed via one item except for sleep, appetite, and psychomotor agitation and restlessness which consist of two to four items. For each item participants are instructed to rate how severely they were affected by the respective symptom in the past week on a 4-point scale with higher scores indicating more severe symptoms. The QIDS is scored by calculating a sum score across all items selecting the highest rated item for symptom groups assessed by more than one item. A QIDS score of ≤ 5 is indicative of no depression, 6–10 of mild, 11–15 of moderate, 16–20 of severe, and ≥ 21 of very severe depression (Rush et al., 2003). Internal consistency of the QIDS was α = 0.85 [0.83, 0.88].

Performance Expectations

The Positive Expectations Scale (PES; Kube et al., 2018) was used to assess participants' expectations towards their performance in the performance task of the study. The PES consists of four items of which two assess generalized expectations (*Working on unknown tasks will generally be easy for me* and *Solving unknown tasks will generally be hard for me*) and the remaining two assess task-specific expectations (*Working on the task will be easy for me* and *Solving the task will be hard for me*). In the current study the TEMINT was the target of the task-specific expectations., hence, the PES was relating to this task. Participants are instructed to rate each item on a 7-point scale ranging from 1 (*fully disagree*) to 7 (*fully agree*). By asking participants to complete the PES twice, before and after a performance task (in this case, the TEMINT), it is therefore possible to assess potentially task-induced changes in participants' expectations about future performance both on the specific task (specific expectations) and unspecified unknown future tasks (generalized expectations). After inverting negatively formulated items, internal consistency was: Pre-task specific α = 0.72 [0.65, 0.77], Post-task specific α = 0.42 [0.28, 0.53]; Pre-task generalized α = 0.74 [0.68, 0.79], Post-task generalized α = 0.48 [0.36, 0.58].

Cognitive Immunization

Cognitive immunization in response to performance feedback was assessed using the Cognitive Immunization after Performance Feedback Scale (CIPF; Kube et al., 2019a, 2019b). The CIPF consists of six statements resembling cognitive immunization strategies such as the tendency to question or devalue performance-related feedback (e.g., *the results of the test were an exception*). Participants are instructed to rate each item on a 7-point scale ranging from 1 (*fully disagree*) to 7 (*fully agree*). To make the score of the CIPF more intuitively interpretable, positively formulated items were inverted so higher scores on the CIPF indicate stronger cognitive immunization. Internal consistency was α = 0.69 [0.64, 0.74].

Anxiety and Stress

Symptoms of anxiety and stress in the past week were assessed using the anxiety and stress subscales of the German translation of the Depression, Anxiety and Stress Scale–21 (DASS-21; Lovibond & Lovibond, 1995) with each consisting of seven statements about typical anxiety- or stress-related symptoms. On each item participants indicate how much each statement applied to them in the past week on a 4-point scale ranging from 0 (*did not apply at all*) to 3 (*applied very much or most of the time*). Internal consistency

was α = 0.86 [0.84, 0.88] for anxiety and α = 0.89 [0.87, 0.90] for stress.

Positive and Negative Mood

State positive and negative mood was assessed using the German translation of the Positive and Negative Affect Schedule (PANAS; Breyer & Bluemke, 2016; Watson et al., 1988). The PANAS consists of 20 items of which ten assess positive and negative mood, respectively. Each item is an adjective describing a certain emotion and participants are instructed to rate how strongly they are feeling this particular emotion in the present moment on a 5-point scale ranging from 1 (*not at all*) to 5 (*extremely*). Internal consistency was α = 0.90 [0.89, 0.92] for positive and α = 0.92 [0.91, 0.94] for negative mood.

Feedback Questionnaire

At the end of the study, participants were invited to give feedback about the study including their guess on the purpose of the study via a free-text answering option.

Procedure

After providing informed consent, participants were directed to the baseline questionnaires assessing demographic data followed by the AST, the QIDS, and the DASS-21. Next, the pre-task PES was applied, followed by the TEMINT after which participants rated how well they thought they performed. Next, participants received standardized feedback depending on their allocated condition and then the post-task PES was applied together with the CIPF. Finally, participants were asked to provide feedback on the study and were debriefed afterwards and could leave their e-mail address to take part in the voucher lottery.

Statistical Approach

For the conceptual replication part of the study, the following analyses were carried out: To investigate the effects of the feedback conditions on cognitive immunization, we calculated a 3 × 1 (Condition: Mildly, moderately, extremely positive) ANCOVA controlling for depressive symptoms and including the interactions between Condition and depressive symptoms. As found by Kube et al. (2022a, 2022b), we expected to find a significant main effect of Condition, driven by a significant difference between all three conditions with the highest cognitive immunization in the mildly positive condition, followed by the extremely and then the moderate condition. We had no specific hypotheses regarding the interaction between Condition x depressive symptoms.[1] Further, the effects of Condition on belief updating were analyzed via 3 × 2 (Condition: Mildly, moderately, extremely positive; Time: Pre-, post-feedback) linear mixed models with depression as a continuous predictor and a random intercept for individual participants and including all two- and three-way interactions between depression, Condition, and Time.[2] Here, we expected to find only a significant main effect of Time, showing an increase in positive expectations from pre- to post-feedback, yet irrespective of depression or Condition. Importantly, this analysis was run twice, once with generalized expectations as dependent variable and once with task-specific expectations as dependent variable. To investigate the associations between depressive symptoms, baseline negative expectations, and IB, as well as with self-rated performance after the tasks but before the feedback, we calculated Spearman's rank correlations with bootstrapped 95% confidence intervals to account for the skewed distribution of these variables. We expected to find that more severe symptoms of depression would be associated with less positive general and specific expectations towards the performance task, as well as with a more negative IB, and a worse self-rated performance. Further, we expected that a more negative IB would be associated with less positive general and task-specific expectations, as well as a more negatively self-rated performance. Finally, for our goal to extend previous research by investigating IB's role in these associations, the following analyses were run: We investigated the unique association between IB and cognitive immunization via multiple regression with depression and IB as continuous predictors of cognitive immunization and Condition (mildly, moderately, extremely positive) as categorical predictors expecting that IB would be a significant predictor, after taking depression and group allocation into account.

To shed further light on our pattern of results through exploratory analyses, we investigated the role of baseline expectations in the association between depression and belief updating. This was done using multiple regression with the change score of expectations from pre- to post-feedback as the outcome, with Condition (mildly, moderately,

[1] The inclusion of the interaction was not pre-specified in the pre-registration as it was not included in the original study by Kube, Kirchner et al. (2022a, 2022b). However, it was added based on reviewer feedback that the inclusion of the interaction would be necessary to investigate the role of depressive symptoms in the reaction to different strengths of positive feedback. Due to the post-hoc nature of the inclusion, no hypotheses for confirmatory analyses were formulated.

[2] The inclusion of the interactions with depressive symptoms were not specified in the pre-registration due to a misreading of the original Kube, Kirchner et al. (2022a, 2022b) study. However, upon further consideration interactions were included to replicate the interactions described in the results of the original study.

extremely positive), depression, and the respective baseline expectations as predictors. Multiple regression models were calculated separately for generalized and task-specific expectations. In further exploratory analyses, the calculation of these multiple regression models was repeated, separately adding self-rated performance and IB to the model to examine whether these cognitive factors would explain variance in belief updating beyond depressive symptoms.

All analyses were run in RStudio version 2022.7.2.576 (Posit Team, 2023) using R version 4.2.1 (R Core Team, 2021). Spearman correlations were calculated using the packages 'stats' (R Core Team, 2021) and 'rcompanion' (Mangiafico, 2021) with 10,000 iterations for bootstrapped confidence intervals. ANCOVAs were calculated using the package 'afex' (Singmann et al., 2020) with Type III sums of squares and the package 'emmeans' (Lenth, 2020) for post-hoc comparisons with Bonferroni correction. Confidence intervals for effect sizes were calculated using the package 'effectsize' (Ben-Shachar et al., 2020). Linear mixed models were computed using the package 'nlme' (Pinheiro et al., 2020) with maximum likelihood estimation. Effect sizes for linear mixed models were calculated using the package 'effectsize'.[3] Post-hoc comparisons with Bonferroni correction and associated effect sizes were computed using the package 'emmeans' (Lenth, 2020) using the pooled SD as the denominator for effect sizes. Multiple regression was computed using the package 'stats' and bootstrapped confidence intervals (10,000 iterations) were calculated using the package 'car' (Fox & Weisberg, 2019). All continuous covariates and predictors were standardized before inclusion in the analysis. For linear mixed models and multiple regression, the between-subject factor Condition was effect coded with the moderately positive feedback condition as the reference category, thus significant effects of Condition can be interpreted as a significant deviation from the sample's grand mean. The within-subject factor Time was dummy coded with pre-feedback as the reference category.

Results

Sample

Comparable to Kube et al. (2022a, 2022b), participants were on average 26.29 years old ($SD = 8.31$, range = 18–67), and were primarily university students 71.47% ($N = 248$). A majority of the sample, 74.64% ($N = 259$), identified themselves as female, and 90.78% ($N = 315$) identified themselves as German. On average, participants showed mild levels of depressive symptoms on the QIDS, $M = 8.76$ ($SD = 5.80$, range = 0–26), with 63.98% of the participants ($N = 222$) reporting an elevated level of depressive symptoms (QIDS total score > 5). Comparisons of the three conditions at baseline showed that participants in the extreme condition were slightly younger than participants in the mild condition, $t(132) = 3.10$, $p = 0.007$, d [95%-CI] = 0.40 [0.14, 0.67], with no difference between the remaining conditions, and that participants in the extreme condition identified more often as non-German compared to the other conditions.[4] For a full sample description, see Table 1.

Cognitive Immunization

As expected, there was a significant main effect of Condition on cognitive immunization, $F(2, 341) = 3.77$, $p = 0.024$, η_G^2 [95%-CI] = 0.02 [0.00, 0.06]. Post-hoc pairwise comparisons revealed the expected significantly higher level of cognitive immunization in the mildly than in the moderately positive feedback condition, $t(232) = 2.75$, $p = 0.020$, d [95%-CI] = 0.36 [0.09, 0.62]. In contrast to our hypotheses, however, the comparison between the mild and the extreme condition, $t(229) = 1.54$, $p = 0.375$, d [95%-CI] = 0.21 [− 0.06, 0.48], as well as the comparison between the moderate and the extreme condition, $t(227) = 1.18$, $p = 0.723$, d [95%-CI] = 0.15 [− 0.11, 0.42] were non-significant. Finally, there was no significant association between depression and cognitive immunization, as indicated by a non-significant main effect of depression, $F(1, 343) = 1.18$, $p = 0.277$, η_G^2 [95%-CI] = 0.00 [0.00, 0.03], and depressive symptoms were not associated with different levels of cognitive immunization in reaction to the different feedback conditions, as the interaction between depressive symptoms and the condition was non-significant, $F(2, 342) = 0.01$, $p = 0.990$, η_G^2 [95%-CI] = 0.00 [0.00, 0.00]. For descriptive statistics, see Table 2 and for a graphical depiction, see Fig. 2.

Generalized and Task-Specific Expectations

Investigating generalized expectations, we found the expected main effect of Time, $t(341) = 9.04$, $p < 0.001$, d [95%-CI] = 0.49 [0.38, 0.60], indicating that participants' general expectations became more positive from pre- to post-feedback. Further, Depression was significantly and negatively associated with general expectations,

[3] Effect size calculation for linear mixed models differed from the pre-registration, as this caused only minor deviations from effect sizes obtained through the pre-specified procedure and allowed calculation of confidence intervals.

[4] Re-running analyses with national identity and age as covariates did not alter the pattern of results. Thus, the pre-registered analyses excluding these covariates are presented throughout the Results section.

Table 1 Sample description at baseline, separately for the full sample and the three feedback conditions

Measure, *M* (*SD*), range	Full sample (*N* = 347)	Level of feedback positivity received			Statistics*
		Mild (*N* = 118)	Moderate (*N* = 116)	Extreme (*N* = 113)	
Age	26.29 (8.31), 18–67	27.76 (10.20), 18–67	26.62 (8.25), 18–61	24.41 (5.35), 18–50	$F(2, 344) = 4.96$, $p = .008$
Gender (female/male/non-binary/no answer)	259/82/5/1	29/88/1/0	27/86/3/0	26/85/1/1	$\chi^2(6) = 3.74$, $p = .712$
National Identity (German/Non-German)	315/32	111/7	108/8	96/17	$\chi^2(2) = 6.85$, $p = .033$
Language fluency	4.75 (.81), 1–5	4.76 (.81), 1–5	4.74 (.80), 1–5	4.74 (.81), 1–5	$F(2, 344) = .02$, $p = .976$
QIDS	8.76 (5.80), 0–26	8.74 (5.81), 0–24	8.65 (5.80), 0–26	8.89 (5.84), 0–25	$F(2, 344) = .05$, $p = .949$
AST	15.27 (21.93), -75–75	16.92 (21.94), -56–75	16.10 (20.94), -75–58	12.68 (22.85), -75–62	$F(2, 344) = 1.20$, $p = .301$
DASS-anxiety	9.29 (9.08), 0–42	9.53 (9.19), 0–32	9.69 (9.03), 0–36	8.64 (9.05), 0–42	$F(2, 344) = .44$, $p = .643$
DASS-stress	14.93 (10.10), 0–42	15.44 (10.17), 0–38	15.07 (9.14), 0–42	14.27 (10.99), 0–42)	$F(2, 344) = .40$, $p = .667$
PANAS-positive	26.81 (8.04), 10–48	27.60 (8.51), 10–48	26.87 (7.60), 10–43	25.91 (7.96), 10–48	$F(2, 344) = 1.28$, $p = .279$
PANAS-Negative	17.39 (7.89, 10–47	16.77 (6.85), 10–40	17.19 (8.05), 10–47	18.26 (8.70), 10–47	$F(2, 344) = 1.08$, $p = .340$
Educational background					$\chi^2(14) = 6.80$, $p = .942$
Primary or no formal education	6	1	2	3	
Secondary education	255	88	82	85	
University degree	86	29	32	25	
Occupational background					$\chi^2(14) = 4.26$, $p = .994$
Pupil	4	1	1	2	
Part-time job	29	11	11	7	
Full-time job	36	13	12	11	
Vocational training	12	4	4	4	
Student (university)	248	80	84	84	
Job searching	10	5	2	3	
In pension	8	4	2	2	
Current therapy (Yes/No)	50/297	15/103	16/100	19/94	$\chi^2(2) = .84$, $p = .657$
Current psychoactive medication (Yes/No)	38/309	15/103	9/107	14/99	$\chi^2(2) = 1.83$, $p = .401$

QIDS Quick Inventory of Depressive Symptomatology, Self-Report, *AST* Ambiguous Scenarios Task, *DASS* Depression, Anxiety, Stress Scale-21, *PANAS* Positive and Negative Affect Scale

*Statistics refer to results of a 3 × 1 (Condition: Mild, Moderate, Extreme) ANOVA

$t(341) = 8.09$, $p < 0.001$, d [95%-CI] = 0.44 [0.33, 0.55], indicating that higher levels of Depression were associated with less positive expectations. However, unexpectedly, there was a significant and positive Depression x Time interaction, $t(341) = 3.92$, $p < 0.001$, d [95%-CI] = 0.21 [0.10, 0.32], indicating that more severe depressive symptoms were associated with a greater increase in positive general expectations from pre- to post-feedback, yet with a small effect size. The remaining main effects and interactions were non-significant $t(341) < 1.72$, $p > 0.088$, $d < 0.10$, indicating that there were no effects of the experimental condition on belief updating, both as a main effect or in interaction with depressive symptoms.

For task-specific expectations, there was also a significant main effect of Time, $t(341) = 10.22$, $p < 0.001$, d [95%-CI] = 0.55 [0.44, 0.67], indicating that participants' task specific expectations became more positive from pre- to post-feedback. Unexpectedly, there was also a significant main effect of the extreme Condition, $t(341) = 2.14$, $p = 0.033$, d [95%-CI] = 0.12 [0.01, 0.22], indicating that

Table 2 Model derived descriptive statistics

Measure, *EMM* (*SE*)	Level of feedback positivity received		
	Mild ($N=118$)	Moderate ($N=116$)	Extreme ($N=113$)
Cognitive immunization	26.90 (.59)	24.60 (.59)	25.60 (.60)
General expectations, pre-feedback	9.28 (.22)	8.87 (.22)	8.75 (.23)
General expectations, post-feedback	10.18 (.22)	10.24 (.22)	10.01 (.23)
Task-specific expectations, pre-feedback	9.22 (.22)	9.31 (.22)	8.67 (.23)
Task-specific expectations, post-feedback	10.32 (.22)	10.76 (.22)	10.99 (.23)

EMM Estimated Marginal Mean, *SE* Standard Error

All values are derived from the 3×2 (Condition: Mildly, moderately, extremely positive; Time: Pre-, post-feedback) linear mixed model with depressive symptoms as a continuous predictor and including all two- and three-way interactions between Time, Condition, and depressive symptoms

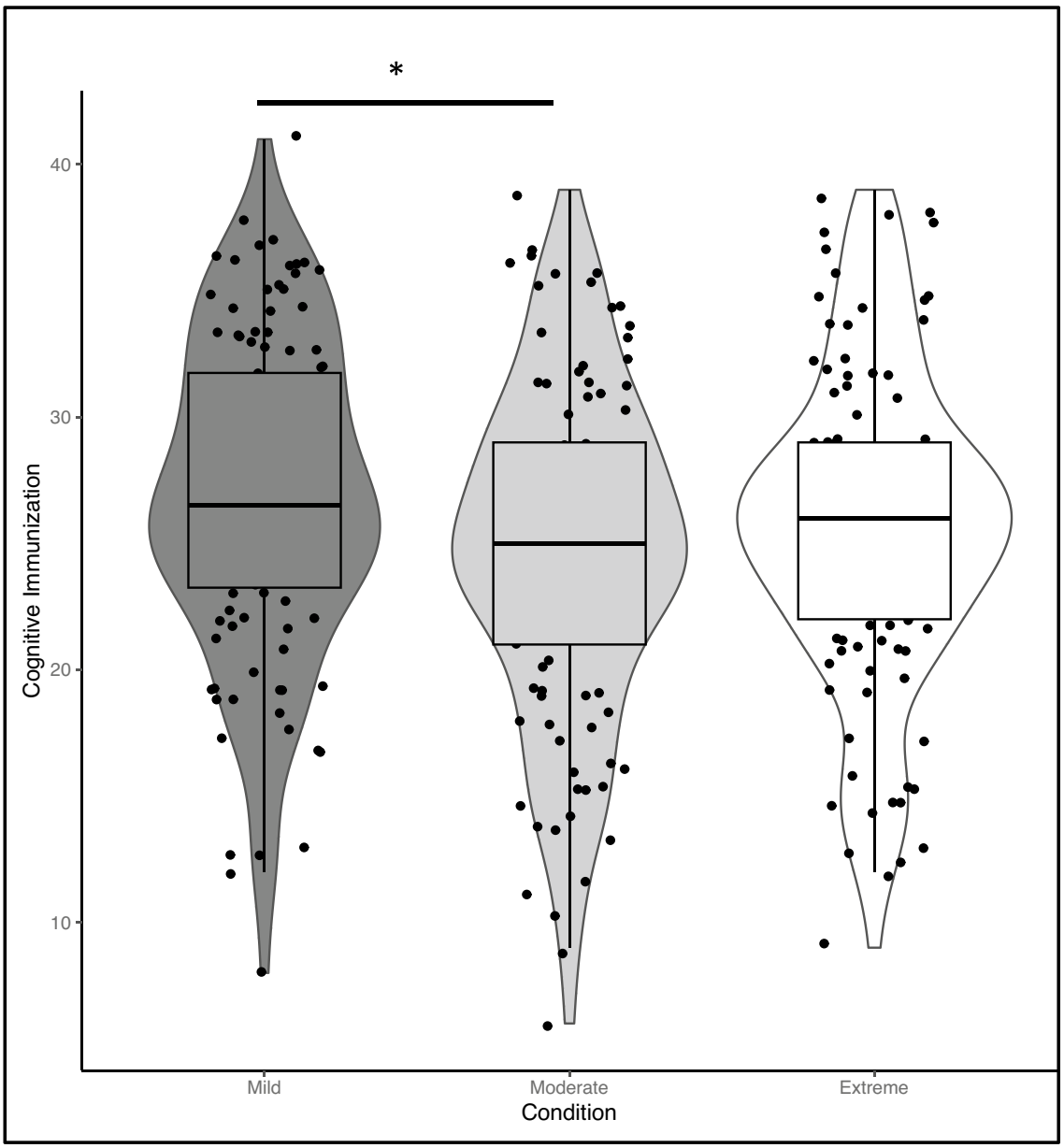

Fig. 2 *$p<.05$. Cognitive immunization in the three feedback conditions after the standardized feedback

Table 3 Zero order correlation between variables before feedback

Measure	*M* (*SD*), range	1	2	3	4
1. QIDS	8.76 (5.80), 0–26	–	–	–	–
2. AST	15.27 (21.93), − 75–75	− .43 [− .52, − .33]	–	–	–
3. Positive generalized expectations	8.97 (2.71), 2–14	− .36 [− .45, − .26]	.32 [.21, .41]	–	–
4. Positive task-specific expectations	9.07 (2.70), 2–14	− .31 [− .41, − .22]	.36 [.25, .45]	.72 [.65, .78]	–
5. Performance self-evaluation	4.63 (1.22), 1–7	− .20 [− .31, − .09]	.25 [.15, .36]	.32 [.22, .42]	.35 [.24, .45]

QIDS Quick Inventory of Depressive Symptomatology, Self-Report. *AST* Ambiguous Scenarios Task

All correlations are significant at $p < .001$. Spearman's rank correlations with bootstrapped 95%-CIs are reported

expectations at pre-feedback were less positive in the extremely positive condition as compared to the grand mean. Further, there was a significant negative main effect of Depression, $t(341) = 6.96$, $p < 0.001$, d [95%-CI] = 0.38 [0.27, 0.49], indicating that higher levels of depression were associated with less positive expectations. As with generalized expectations, however, a significant Depression x Time interaction, $t(341) = 2.74$, $p = 0.007$, d [95%-CI] = 0.15 [0.04, 0.25], indicated that higher depression scores were surprisingly associated with a greater increase in task-specific expectations from pre- to post-feedback. Finally, there was a significant negative mild Condition x Time interaction, $t(341) = 2.33$, $p = 0.020$, d [95%-CI] = 0.13 [0.02, 0.23], as well as a significant positive extreme condition x Time interaction, $t(341) = 3.07$, $p = 0.002$, d [95%-CI] = 0.17 [0.06, 0.27], indicating that the increase in positive expectations was significantly lower than average in the mild condition, and significantly higher than average in the extreme condition. The remaining main effects and interactions were non-significant, $t(341) < 1.41$, $p > 0.160$, $d < 0.08$, indicating that these interactions did not differ in dependence of depression. Further exploratory analyses on the mediating role of cognitive immunization on belief updating can be found in the Supplements, suggesting that cognitive immunization did not mediate the effects of Condition on belief updating.

Correlational Results

As expected, symptoms of depression were moderately and negatively associated with positive task-specific and generalized expectations, strongly and negatively with a positive IB, and weakly and negatively with positive self-rated performance after the task. Congruently, a more positive IB was moderately and positively associated with general and task-specific expectations, as well as with a more positive self-rated performance. For details, see Table 3.

Interpretation Bias and Cognitive Immunization

Contrary to our expectations, we found no significant association between IB and cognitive immunization after taking Condition and Depression into account, β [95%-CI] = − 0.11 [− 0.24, 0.05], $p = 0.078$, $t(342) = 1.77$. The remaining results resembled the results without including IB as a predictor, thus there was no significant main effect of Depression, β [95%-CI] = − 0.11 [− 0.22, 0.00], $p = 0.073$, $t(342) = 1.80$, yet a significant main effect of the mild feedback condition which showed a significantly higher level of cognitive immunization compared to the grand mean, β [95%-CI] = 0.19 [0.05, 0.34], $p = 0.010$, $t(342) = 2.59$, while the extremely positive feedback condition did not differ from the grand mean, β [95%-CI] = − 0.03 [− 0.18, 0.12], $p = 0.723$, $t(342) = 0.35$. Although the full model was statistically significant, $F(4, 342) = 2.98$, $p = 0.019$, it only accounted for $R^2_{adjusted} = 2.24\%$ of the total variance in cognitive immunization (Table 4).

Exploratory Analyses

Baseline Expectations, Depressive Symptoms and Belief Updating

Due to the finding that depressive symptoms were associated with less positive expectations, we ran exploratory analyses to investigate the association between depression and belief updating when controlling for baseline expectations; potentially our finding that higher levels of depressive symptoms were associated with greater increases in positive expectancies could be a result of more depressed participants having lower expectancies to start with, and hence more scope for an increase. For generalized belief updating, we found that baseline generalized expectations were a significant negative predictor of the generalized expectation change score, β [95%-CI] = − 0.60 [− 0.71, − 0.50], $p < 0.001$, $t(342) = 12.79$, indicating that more positive baseline expectations were associated with a smaller increase in positive generalized expectations. However, neither depressive symptoms nor group allocation were significant predictors, $|\beta| < 0.04$, $p > 0.524$.

For task-specific expectations, baseline expectations were a significant negative predictor of the change score, β [95%-CI] = − 0.70 [− 0.79, − 0.61], $p < 0.001$, $t(342) = 16.60$,

Table 4 Exploratory regression models predicting expectation change

Predictor	Change in generalized expectations			Change in specific expectations		
	β [95%-CI]	*t*	*p*	β [95%-CI]	*t*	*p*
Model 1						
Baseline expectations	− .60 [− .71, − .50]	12.79	< .001	− .70 [− .79, − .61]	16.60	< .001
QIDS	− .03 [− .12, .07]	.57	.569	− .09 [− .18, .00]	2.12	.035
Moderate condition	.09 [− .11, .30]	.88	.381	.14 [− .05, 3.31]	1.45	.145
Extreme condition	.02 [− .18, .23]	.23	.821	.26 [.07, .45]	2.73	.007
Model 2						
Baseline expectations	− .64 [− .75, − .53]	13.28	< .001	− .77 [− .85, − .67]	18.69	< .001
QIDS	− .01 [− .11, .10]	.12	.906	− .05 [− .13, .04]	1.18	.238
Performance Self-Evaluation	.14 [.03, .24]	3.00	.003	.26 [.17, .35]	6.51	< .001
Moderate condition	.10 [− .09, .30]	.98	.330	.17 [− .01, .36]	1.87	.063
Extreme condition	.03 [− .18, .23]	.26	.798	.27 [.09, .45]	2.92	.004
Model 3						
Baseline expectations	− .61 [− .72, − .51]	12.65	< .001	− .69 [− .78, − .59]	16.00	< .001
QIDS	.02 [− .11, .13]	.33	.743	− .05 [− .14, .05]	1.01	.315
AST	− .02 [.13, .09]	.39	.696	− .11 [− .20, − .02]	2.35	.020
Moderate condition	.09 [− .11, .30]	.88	.380	.14 [− .06, .32]	1.42	.155
Extreme condition	.03 [− .17, .24]	.25	.803	.26 [.06, .45]	2.66	.008

QIDS Quick Inventory of Depressive Symptomatology. *AST* Ambiguous Scenarios Task

95%-CIs are bootstrapped. Baseline expectations refer to generalized and specific baseline expectations in the respectivemodel

indicating that more positive task specific expectations at baseline were associated with a smaller increase in positive expectations. Further, depressive symptoms were a significant negative predictor of the change score, β [95%-CI] = − 0.09 [− 0.18, 0.00], $p = 0.035$, $t(342) = 2.12$, showing that more severe depressive symptoms were associated with a smaller increase in positive expectations after taking baseline expectations into account. Finally, the expectation change score in the mildly positive feedback condition was smaller compared to the grand mean, β [95%-CI] = − 0.13 [− 0.25, − 0.03], $p = 0.016$, $t(342) = 2.42$, while the extremely positive condition showed a larger change score compared to the grand mean, β [95%-CI] = 0.13 [0.02, 0.24], $p = 0.022$, $t(342) = 2.31$.

Cognitive Factors Involved in Belief Updating

To shed further light on cognitive factors potentially associated with limited belief updating, we repeated the main analyses for the association between depression and belief updating, with participants' self-rated performance and IB as additional predictors of belief updating after accounting for depressive symptoms. Due to the significant correlations between both IB and self-rated performance with baseline expectations, we investigated IB and self-rated performance as predictors of expectation change, controlling for baseline expectations.

Self-Rated Performance

For generalized expectations, self-rated performance was a significant predictor of belief updating, β [95%-CI] = 0.14 [0.03, 0.24], $p = 0.003$, $t(341) = 3.00$, indicating that a more positive self-rated performance was associated with a greater increase in positive expectations. Further, baseline expectations were a significant negative predictor of belief updating β [95%-CI] = − 0.64 [− 0.75, -− 0.53], $p < 0.001$, $t(342) = 13.28$, indicating that more positive baseline expectations were associated with a smaller increase in positive expectations from pre- to post feedback. However, neither depressive symptoms, β [95%-CI] = − 0.01 [− 0.11, 0.10], $p = 0.906$, $t(342) = 0.12$, nor the allocated Condition, $|\beta| < 0.05$, $p > 0.477$, were significant predictors of belief updating.

For task-specific expectations, results partly mirrored those for generalized expectations, such that baseline expectations were also a significant negative predictor of belief updating, β [95%-CI] = − 0.77 [− 0.86, − 0.68], $p < 0.001$, $t(342) = 18.69$, and more positive self-rated performance was associated with a greater increase in positive expectations, β [95%-CI] = 0.26 [0.17, 0.34], $p < 0.001$, $t(342) = 6.51$, while depressive symptoms were no significant predictor of belief updating, β [95%-CI] = − 0.05 [− 0.13, 0.04], $p = 0.238$, $t(342) = 1.18$. However, in contrast to the results for generalized belief updating, Condition was a significant predictor, as such that the mildly positive condition showed

a significantly smaller increase in positive expectations compared to the grand mean, β [95%-CI] = − 0.14 [− 0.25, − 0.04], $p = 0.006$, $t(342) = 2.78$, while the extremely positive condition showed a significantly greater increase in positive expectations, β [95%-CI] = 0.12 [0.02, 0.22], $p = 0.023$, $t(342) = 2.29$.

Interpretation Bias

IB was neither a predictor of generalized belief updating, β [95%-CI] = 0.02 [− 0.10, 0.13], $p = 0.743$, $t(341) = 0.33$, nor of task-specific updating, β [95%-CI] = − 0.05 [− 0.14, 0.05], $p = 0.315$, $t(341) = 1.01$. This illustrates, that for both generalized and task-specific expectations, including IB as an additional predictor together with depressive symptoms and baseline expectations did not change the pattern of results obtained through the model without IB.

Discussion

The main aims of the present study were (a) to conceptually replicate previous findings by Kube et al. (2022a, 2022b) on the relationship between symptoms of depression, belief updating, and cognitive immunization, as well as the effects of different levels of positive feedback on these variables, and (b) to investigate the relationships between IBs, negative expectations, belief updating, and cognitive immunization.

Regarding our first aim, our results were only partly in line with those of Kube et al. (2022a, 2022b). Specifically, we found that the mildly positive feedback condition reported the highest level of cognitive immunization, in accordance with the original study. However, the remaining results did not fully mirror those of the original study as there were no further significant group differences in cognitive immunization, and there was no association between cognitive immunization and depression, irrespective of the strength of positive feedback. Further, similar to Kube et al. (2022a, 2022b), we found that all participants updated their general and task-specific expectations after feedback, irrespective of condition. However, unexpectedly, and in contrast to the original study, we found that more severe depressive symptoms were associated with a greater increase in positive expectations. Due to the unexpected nature of this result, we followed it up in exploratory analyses. We found that depression was associated with reduced task-specific belief updating when baseline expectations were taken into account, in line with previous studies (Kube & Glombiewski, 2021; Kube et al., 2019a, 2019b). For generalized expectations, too, the surprising positive association between expectation update and depression severity diminished when controlling for the effects of baseline expectations; instead, depression was unrelated to change in generalized expectations in this analysis.

For our second aim, we found the expected associations, meaning that depressive symptoms were associated with a more negative IB, and both were associated with less positive performance expectations, and a less positive performance self-evaluation. However, there was no association between IB and cognitive immunization. In exploratory analyses, we found that the performance self-evaluation was a significant predictor of belief updating, and notably, depression no longer predicted reduced belief updating when the performance self-evaluation was taken into account. Finally, IB was no significant predictor of belief updating.

Put into context, our findings suggest that, irrespective of depressive symptoms, mildly positive feedback leads to a greater amount of cognitive immunization than moderately or extremely positive feedback in a non-clinical sample. This finding is a replication of the results by Kube et al. (2022a, 2022b) and provides some additional evidence regarding the potential robustness of the effect. While we descriptively found that the least amount of cognitive immunization could be found in the moderate condition, the comparison between the moderate and the extreme condition were non-significant. One interpretation of these results, also proposed by Kube et al. (2022a, 2022b), is that participants generally expected to have performed better than the average participant and thus evaluated the feedback that they were better than 50% of the participants as invalid. Consequently, the cognitive immunization in the mildly positive condition might represent an immunization against unexpectedly negative feedback instead of positive feedback as in the other two conditions. This is in line with previous research suggesting that healthy participants and participants across a spectrum of depressive symptoms show a positivity bias (Moore & Fresco, 2012), as well as with our finding that participants in our sample rated their performance on average above the midpoint of the scale. Another reason for the consistent finding that the mildly positive feedback condition leads to the highest cognitive immunization could be associated with the precision of that information which is argued to be central in models of predictive processing. While the mildly positive feedback condition informs participants that they are within an interval with 50% of other participants, the other conditions give substantially more precise information, namely that participants are within a 10% or a 1% interval. The feedback in the mildly positive condition might consequently appear as the most arbitrary leading to cognitive immunization in form of a devaluation of the task's validity. Future research applying this paradigm could therefore include a more exact description of participants' reached performance including the upper bound of the interval to control for objective feedback precision.

Further, in contrast to the original findings by Kube et al. (2022a, 2022b) and a recent study testing the effects of different levels of feedback positivity following a social situation (Kube, 2023), our findings suggest that cognitive immunization is not associated with depressive symptoms. These diverging findings may be, however, attributable to a non-robust association in this particular paradigm. To illustrate, the original study using the same paradigm (Kube et al., 2022a, 2022b) reported an effect size within the limits of our effect size confidence interval ($\eta_p^2 = 0.014$, in comparison to our 95%-CI [0.00, 0.03]) and is therefore comparable to our finding, while another paradigm applied by Kube et al. (2022a, 2022b; Study 2) led to a considerably larger effect size for cognitive immunization and depressive symptoms that could be replicated (Kube, 2023). Consequently, whether cognitive immunization towards positive feedback is associated with depressive symptoms remains a matter of debate that is replicable for social- but not for performance situations. Importantly, our finding that the level of feedback positivity does not moderate the relationship between depressive symptoms and cognitive immunization is in line with findings by Kube (2023). Taking together these results on cognitive immunization, it can be concluded that depressive symptoms may be associated with cognitive immunization, yet the discrepancy between the feedback positivity and the initial expectation, suggested by predictive processing models, may not be the driving factor.

For the update of generalized and task-specific performance expectations, we replicated that participants' performance expectations became more positive after feedback. We further replicated the finding that there were no group differences in the increase of generalized positive expectations based on the level of feedback positivity. In contrast to our own hypotheses, yet in line with findings by Kube et al. (2022a, 2022b), we found that for task-specific expectations the mild feedback condition led to the lowest increase in positive expectations while the extreme condition led to the greatest increase. These findings are in contrast with the assumption of a tipping point at which the prediction error is large enough to change prior expectations, yet not too large to be credible, put forward by predictive processing models of depression (Kube et al., 2020). Instead, our findings are rather in line with the Rescorla-Wagner learning model, suggesting that more positive feedback would lead to a greater updating of prior expectations (Rescorla & Wagner, 1972). Comparable to our discussion of the effects of the feedback manipulation on cognitive immunization, a central factor in these results might be methodological considerations. To illustrate, comparing the design of those studies that did not find the tipping point of feedback positivity (the present study and the original study by Kube et al., 2022a, 2022b) to those that did find evidence for it (Kube, 2023; Kube et al., 2022a, 2022b study 2), one central difference is the paradigm with which expectation violation was investigated. While the application of the TEMINT paradigm did not produce a tipping point, differing levels of feedback positivity regarding a social situation did. One reason for that finding might be that for the TEMINT it is particularly difficult to estimate one's performance (e.g., Kube et al., 2018), and thus participants might put more trust in external feedback and consequently not estimate their own prior expectations to be particularly precise. On the other hand, people are familiar with social situations and might be more experienced in forming expectations about their successful application of social skills. Consequently, they might form their post-feedback expectations more strongly through a comparison of the feedback to their own expectations. One option to investigate this assumption in future research might be to include an assessment in participants' confidence in their own expectations. Another possible reason for the discrepancy between the findings obtained through the TEMINT and the social situation is that the different levels of feedback positivity in the social situation were intuitively described (i.e., the social situation was increasingly pleasant) while the percent ranks provided in the TEMINT were to an extent arbitrary (for different percent rank choices with a comparable goal, see Zell & Lesick, 2021). Consequently, the application of a social situation and the non-reliance on percent ranks as a manipulation of feedback positivity may present a more valid option for the investigation of belief updating. This potential superiority of a social situation to the TEMINT is further supported by our findings on the internal consistency of the outcome scale. Our study found a considerable drop in internal consistency from pre- to post-feedback on the PES which limits the validity of the findings. One interpretation might be that participants were unconfident in the conclusions they should draw based on the performance feedback they just received and thus did not provide coherent answers on the post-training PES. Relying on feedback regarding a social situation instead on the TEMINT including the accompanying performance questionnaire (Kube, 2023; Kube et al., 2022a, 2022b study 2) might provide a more reliable outcome measure.

Taking the role of depressive symptoms in belief updating into account, the pattern of results becomes more nuanced. Our initial results surprisingly indicated that more severe symptoms were associated with greater belief updating. However, our exploratory analyses suggested that this finding was likely attributable to regression to the mean and depressive symptoms being associated with less positive baseline expectations, leaving more room for an increase in positive expectations. Specifically, results of our exploratory analyses of task-specific belief updating controlling for baseline expectations showed that depressive symptoms were associated with thwarted belief updating. Generally in line with predictive processing models of depression

(Kube et al., 2020), and previous empirical findings (e.g., Kube et al., 2022a, 2022b), this underlines that depressive symptoms are associated with reduced belief updating in light of positive feedback, proposing it as a potential factor associated with depression maintenance. However, results of our further exploratory analyses suggested that participants' self-evaluation might be a potentially central cognitive factor here, as we found that a more positive self-evaluation of participants' performance predicted greater belief updating, and when taking this into account, depressive symptoms were no longer a significant predictor. The exploratory nature of these findings warrants cautious interpretation, yet given replication in further studies, it might propose participants' self-evaluation as a potential cognitive factor explaining the lack of belief updating in depression. Consequently, self-evaluation may be a critical target for interventions to improve the ability to update expectations in a positive manner, and ultimately affect depressive symptoms. This would be in line with results from studies investigating the role of self-esteem in depression, showing an important association between low self-esteem and depressive symptoms (Phillips & Hine, 2013; Phillips et al., 2012; Porter et al., 2019), self-esteem being associated with more frequent sudden gains during treatment (Kelly et al., 2007), and self-esteem change during treatment being a potentially central mechanism in symptom improvement (Dinger et al., 2017). Our exploratory findings might therefore carefully suggest that future research testing whether an improvement in self-esteem is associated with an improvement in depressive symptoms through increasing participants' susceptibility to positive feedback is promising.

Moving the focus to our investigation of the role of IBs in positive expectations, their updating, and cognitive immunization, correlational evidence suggested that positive expectations and a more positive IB are associated with each other, yet amplitudes of these correlations suggested relatively distinct cognitive constructs. This finding could be attributable to the instructions participants received regarding the performance task, as the information that the task is very difficult might have rendered the expectations regarding the task to rather resemble optimism in light of negative information instead of interpretation of ambiguity, while the former has been found to be associated with cognitive immunization (Supplementary Material of Kube et al., 2022a, 2022b). Further, IB was not associated with cognitive immunization, contrasting earlier studies that found a small, yet significant association between IBs and the application of reappraisal strategies in the context of depressive symptoms (For positive reappraisal, see Blanco et al., 2021; Everaert et al., 2017a, 2017b; For negative reappraisal, see Everaert et al., 2020). An important difference between this study and earlier studies concerns the match between IB and the object of reappraisal. Namely, other studies found a significant association between a general IB, resembling the interpretation of broad range of situations or stimuli, and general reappraisal, for example, the dampening of positivity in general. Our finding might therefore be attributable to a mismatch between our broad conceptualization of IB through the AST on the one hand, assessing the interpretation of various everyday situations, and the narrow conceptualization of cognitive immunization on the other, that was conceptualized as the negative reappraisal of the positive performance feedback. Future studies investigating the association between IB and cognitive immunization might therefore aim for a comparable object of interpretation and for cognitive immunization, i.e., by using a social performance task as Kube et al. (2022a, 2022b) due to the AST's focus on ambiguous social situations.

Limitations

Results of this study need to be interpreted in light of some limitations. First, the inclusion of different levels of feedback as a between-group factor does not allow the investigation of the unmanipulated association between IB and cognitive immunization. The division of our sample into three groups, which might have slightly different association between the two variables, may have therefore canceled out a potential association or may have limited power to the point where this association became non-significant. In particular, the mild feedback condition poses an important limitation here, as our results suggest that participants interpreted this feedback rather as slightly negative than as slightly positive. Consequently, this condition only allowed the investigation of cognitive immunization and belief updating in light of slightly negative feedback. Another limitation that has been discussed throughout the manuscript concerns the internal consistency of the positive expectation assessment at post-feedback. While the scales showed sufficient internal consistency in previous studies (e.g., Kube et al., 2022a, 2022b) we found only an inacceptable level of internal consistency which in turn limits the generalizability of our findings. This limitation becomes more apparent considering that there is currently no research on the retest-reliability of the scale, raising the question whether the repeated application of the PES is valid in the experimental context. Further, while our sample on average showed above-cut-off levels of depression, the sample must still be considered non-clinical and thus our findings are potentially not generalizable to patient populations with higher levels of depressive symptoms. Finally, our sample consisted to the greatest extent of female, German students. Our results are therefore not necessarily generalizable to a more diverse sample.

Conclusion

The present study had the aim to replicate previous findings on the association between depressive symptoms, cognitive immunization, and belief updating. Further, the aim was to shed light on the role of IBs in these associations. Our results suggest that independently of depression, participants engage in the greatest amount of cognitive immunization when presented with feedback that suggests that they performed as well as the average participant. Taking our finding into account that participants rated their own performance on average above the mid-point of the assessment scale, this might suggest that the feedback being average conflicted with a more positive self-evaluation and thus led to cognitive immunization. Further, depressive symptoms were associated with thwarted updating of positive, task-specific expectations. However, as this association disappeared after taking participants' self-evaluation into account, and our exploratory results propose participants' self-evaluation as a potentially important factor in targeting thwarted belief updating, and ultimately depression. Finally, IB was moderately associated with positive expectations, suggesting different cognitive constructs, while the remaining results did not suggest a central role of IB in belief updating or cognitive immunization. Due to the mixed nature of the results of the present study, future studies should optimize the TEMINT framework for the investigation of belief updating or adopt a social scenario framework that may produce more valid and reliable findings.

Author Contributions FW, TK, MLW, JM, and SEB: contributed to the study design. FW: contributed the data analysis and wrote the first version of the manuscript. All authors commented on previous versions of the manuscript and approved the final version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Felix Würtz is supported by a doctoral scholarship of the Studienstiftung des deutschen Volkes. Marcella L. Woud and Jürgen Margraf are Principal Investigators within the SFB 1280 Extinction Learning, Deutsche Forschungsgemeinschaft (DFG). Project title: The impact of modification of stimulus-outcome contingencies on extinction and exposure in anxiety disorders (project number: 316803389). Marcella L. Woud is further supported by the Emmy Noether Program of the Deutsche Forschungsgemeinschaft (DFG), project title: Mechanisms of interpretation biases in anxiety–bridging experimental investigation and clinical translation (WO 2018/3-1, project number: 442163275). The funders had no role in the study design, collection, analysis, or interpretation of data, writing the manuscript, or the decision to submit the paper for publication.

Data Availability The analysis scripts and the anonymized dataset are openly available on the Open Science Framework (OSF; https://osf.io/6bx4v/).

Declarations

Conflict of Interest Felix Würtz, Tobias Kube, Marcella L. Woud, Jürgen Margraf, and Simon E. Blackwell declare that they have no conflict of interest.

Ethical Approval The study was approved by the local ethics committee of the Faculty of Psychology at Ruhr-University Bochum (approval no. 742).

Informed Consent All participants provided written informed consent by agreeing to an online form.

Research Involving Human and Animal Rights No animal studies were carried out by the authors for this article.

References

Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2015). Computational psychiatry: Towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery & Psychiatry.* https://doi.org/10.1136/jnnp-2015-310737

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society b: Biological Sciences, 371*(1708), 20160011. https://doi.org/10.1098/rstb.2016.0011

Beck, A. T., & Haigh, E. A. P. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual Review of Clinical Psychology, 10*(1), 1–24. https://doi.org/10.1146/annurev-clinpsy-032813-153734

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software, 5*(56), 2815. https://doi.org/10.21105/joss.02815

Blanco, I., Boemo, T., & Sanchez-Lopez, A. (2021). An online assessment to evaluate the role of cognitive biases and emotion regulation strategies for mental health during the COVID-19 lockdown of 2020: Structural equation modeling study. *JMIR Mental Health, 8*(11), e30961. https://doi.org/10.2196/30961

Breyer, B., & Bluemke, M. (2016). Deutsche version der positive and negative affect schedule PANAS (GESIS Panel). *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen (ZIS)*. https://doi.org/10.6102/ZIS242

Cane, D. B., & Gotlib, I. H. (1985). Depression and the effects of positive and negative feedback on expectations, evaluations, and performance. *Cognitive Therapy and Research, 9*(2), 145–160. https://doi.org/10.1007/BF01204846

Caouette, J. D., & Guyer, A. E. (2016). Cognitive distortions mediate depression and affective response to social acceptance and

rejection. *Journal of Affective Disorders, 190*, 792–799. https://doi.org/10.1016/j.jad.2015.11.015

Clark, J. E., Watson, S., & Friston, K. J. (2018). What is mood? *A Computational Perspective. Psychological Medicine, 48*(14), 2277–2284. https://doi.org/10.1017/S0033291718000430

Dinger, U., Ehrenthal, J. C., Nikendei, C., & Schauenburg, H. (2017). Change in self-esteem predicts depressive symptoms at follow-up after intensive multimodal psychotherapy for major depression. *Clinical Psychology & Psychotherapy, 24*(5), 1040–1046. https://doi.org/10.1002/cpp.2067

Everaert, J., & Koster, E. H. W. (2020). The interplay among attention, interpretation, and memory biases in depression: Revisiting the combined cognitive bias hypothesis. In T. Aue & H. Okon-Singer (Eds.), *Cognitive biases in health and psychiatric disorders: Neurophysiological foundations* (pp. 193–213). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-816660-4.00009-X

Everaert, J., Duyck, W., & Koster, E. H. W. (2014). Attention, interpretation, and memory biases in subclinical depression: A proof-of-principle test of the combined cognitive biases hypothesis. *Emotion, 14*(2), 331–340. https://doi.org/10.1037/a0035250

Everaert, J., Grahek, I., Duyck, W., Buelens, J., Van den Bergh, N., & Koster, E. H. W. (2017a). Mapping the interplay among cognitive biases, emotion regulation, and depressive symptoms. *Cognition & Emotion, 31*(4), 726–735. https://doi.org/10.1080/02699931.2016.1144561

Everaert, J., Podina, I. R., & Koster, E. H. W. (2017b). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review, 58*, 33–48. https://doi.org/10.1016/j.cpr.2017.09.005

Everaert, J., Bronstein, M. V., Cannon, T. D., & Joormann, J. (2018). Looking through tinted glasses: Depression and social anxiety are related to both interpretation biases and inflexible negative interpretations. *Clinical Psychological Science, 6*(4), 517–528. https://doi.org/10.1177/2167702617747968

Everaert, J., Bronstein, M. V., Castro, A. A., Cannon, T. D., & Joormann, J. (2020). When negative interpretations persist, positive emotions don't! Inflexible negative interpretations encourage depression and social anxiety by dampening positive emotions. *Behaviour Research and Therapy, 124*, 103510. https://doi.org/10.1016/j.brat.2019.103510

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Feldmann, M., Kube, T., Rief, W., & Brakemeier, E. (2022). Testing bayesian models of belief updating in the context of depressive symptomatology. *International Journal of Methods in Psychiatric Research*. https://doi.org/10.1002/mpr.1946

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage.

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry, 1*(2), 148–158. https://doi.org/10.1016/S2215-0366(14)70275-5

Gadassi Polack, R., Davis, A. L., & Joormann, J. (2023). Moderators and mediators of the interpretation bias-emotional disorders link. In M. L. Woud (Ed.), *Interpretational processing biases in emotional psychopathology* (pp. 55–77). Springer International Publishing. https://doi.org/10.1007/978-3-031-23650-1_4

Gagne, C., Agai, S., Ramiro, C., Dayan, P., & Bishop, S. (2022). Biased belief priors versus biased belief updating: Differential correlates of depression and anxiety. *PLOS Computational Biology, 18*(8), e1010176. https://doi.org/10.1371/journal.pcbi.1010176

Gopinath, S., Katon, W. J., Russo, J. E., & Ludman, E. J. (2007). Clinical factors associated with relapse in primary care patients with chronic or recurrent depression. *Journal of Affective Disorders, 101*(1–3), 57–63. https://doi.org/10.1016/j.jad.2006.10.023

Horwitz, A. G., Berona, J., Czyz, E. K., Yeguez, C. E., & King, C. A. (2017). Positive and negative expectations of hopelessness as longitudinal predictors of depression, suicidal ideation, and suicidal behavior in high-risk adolescents. *Suicide and Life-Threatening Behavior, 47*(2), 168–176. https://doi.org/10.1111/sltb.12273

Kelly, M. A. R., Roberts, J. E., & Bottonari, K. A. (2007). Non-treatment-related sudden gains in depression: The role of self-evaluation. *Behaviour Research and Therapy, 45*(4), 737–747. https://doi.org/10.1016/j.brat.2006.06.008

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine, 44*(3), 579–592. https://doi.org/10.1017/S0033291713001074

Kube, T. (2023). If the discrepancy between expectations and actual information is too large, expectation change decreases – A replication study. *Journal of Behavior Therapy and Experimental Psychiatry, 79*, 101831. https://doi.org/10.1016/j.jbtep.2022.101831

Kube, T., & Glombiewski, J. A. (2021). How depressive symptoms hinder positive information processing: An experimental study on the interplay of cognitive immunisation and negative mood in the context of expectation adjustment. *Cognitive Therapy and Research, 45*(3), 517–528. https://doi.org/10.1007/s10608-020-10191-4

Kube, T., D'Astolfo, L., Glombiewski, J. A., Doering, B. K., & Rief, W. (2017). Focusing on situation-specific expectations in major depression as basis for behavioural experiments—Development of the depressive expectations scale. *Psychology and Psychotherapy: Theory, Research and Practice, 90*(3), 336–352. https://doi.org/10.1111/papt.12114

Kube, T., Rief, W., Gollwitzer, M., & Glombiewski, J. A. (2018). Introducing an EXperimental Paradigm to investigate Expectation Change (EXPEC). *Journal of Behavior Therapy and Experimental Psychiatry, 59*, 92–99. https://doi.org/10.1016/j.jbtep.2017.12.002

Kube, T., Glombiewski, J. A., Gall, J., Touissant, L., Gärtner, T., & Rief, W. (2019a). How to modify persisting negative expectations in major depression? An experimental study comparing three strategies to inhibit cognitive immunization against novel positive experiences. *Journal of Affective Disorders, 250*, 231–240. https://doi.org/10.1016/j.jad.2019.03.027

Kube, T., Rief, W., Gollwitzer, M., Gärtner, T., & Glombiewski, J. A. (2019b). Why dysfunctional expectations in depression persist—Results from two experimental studies investigating cognitive immunization. *Psychological Medicine, 49*(09), 1532–1544. https://doi.org/10.1017/S0033291718002106

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry, 87*(5), 388–398. https://doi.org/10.1016/j.biopsych.2019.07.017

Kube, T., Friehs, T., Glombiewski, J. A., & Gollwitzer, M. (2022a). Depression is not related to deficits in emotional reasoning skills. *Personality and Individual Differences, 196*, 111728. https://doi.org/10.1016/j.paid.2022.111728

Kube, T., Kirchner, L., Lemmer, G., & Glombiewski, J. A. (2022b). How the discrepancy between prior expectations and new information influences expectation updating in depression—The greater, the better? *Clinical Psychological Science, 10*(3), 430–449. https://doi.org/10.1177/21677026211024644

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Retrieved, from https://CRAN.R-project.org/package=emmeans

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales* (2nd ed.). Psychology Foundation.

Mangiafico, S. (2021). *rcompanion: Functions to Support Extension Education Program Evaluation*. Retrieved, from https://CRAN.R-project.org/package=rcompanion

Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical Psychology Review, 32*(6), 496–509. https://doi.org/10.1016/j.cpr.2012.05.004

O'Connor, C. E., Everaert, J., & Fitzgerald, A. (2021). Interpreting ambiguous emotional information: Convergence among interpretation bias measures and unique relations with depression severity. *Journal of Clinical Psychology*. https://doi.org/10.1002/jclp.23186

Paulus, M. P., Feinstein, J. S., & Khalsa, S. S. (2019). An active inference approach to interoceptive psychopathology. *Annual Review of Clinical Psychology, 15*(1), 97–122. https://doi.org/10.1146/annurev-clinpsy-050718-095617

Phillips, W. J., & Hine, D. W. (2013). Exploring the factor structure of implicit and explicit cognitions associated with depression. *Assessment, 20*(4), 474–483. https://doi.org/10.1177/1073191112437595

Phillips, W. J., Hine, D. W., & Bhullar, N. (2012). A latent profile analysis of implicit and explicit cognitions associated with depression. *Cognitive Therapy and Research, 36*(5), 458–473. https://doi.org/10.1007/s10608-011-9381-z

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. Retrieved, from https://CRAN.R-project.org/package=nlme

Porter, A. C., Zelkowitz, R. L., Gist, D. C., & Cole, D. A. (2019). Self-evaluation and depressive symptoms: A latent variable analysis of self-esteem, shame-proneness, and self-criticism. *Journal of Psychopathology and Behavioral Assessment, 41*(2), 257–270. https://doi.org/10.1007/s10862-019-09734-1

Posit Team. (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. Retrieved , from http://www.posit.co/

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Retrieved , from https://www.R-project.org/

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (Vol. 2, pp. 64–99). Appleton-Century-Crofts.

Rief, W., & Joormann, J. (2019). Revisiting the cognitive model of depression: The role of expectations. *Clinical Psychology in Europe, 1*(1), e32605. https://doi.org/10.32872/cpe.v1i1.32605

Rief, W., Glombiewski, J. A., Gollwitzer, M., Schubö, A., Schwarting, R., & Thorwart, A. (2015). Expectancies as core features of mental disorders. *Current Opinion in Psychiatry, 28*(5), 378–385. https://doi.org/10.1097/YCO.0000000000000184

Rohrbacher, H., & Reinecke, A. (2014). Measuring change in depression-related interpretation bias: Development and validation of a parallel ambiguous scenarios test. *Cognitive Behaviour Therapy, 43*(3), 239–250. https://doi.org/10.1080/16506073.2014.919605

Roniger, A., Späth, C., Schweiger, U., & Klein, J. (2015). A Psychometric evaluation of the German version of the quick inventory of depressive symptomatology (QIDS-SR16) in outpatients with depression. *Fortschritte Der Neurologie Psychiatrie, 83*(12), e17–e22. https://doi.org/10.1055/s-0041-110203

Rude, S. S., Durham-Fowler, J. A., Baum, E. S., Rooney, S. B., & Maestas, K. L. (2010). Self-report and cognitive processing measures of depressive thinking predict subsequent major depressive disorder. *Cognitive Therapy and Research, 34*(2), 107–115. https://doi.org/10.1007/s10608-009-9237-y

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Schmidt-Atzert, L., & Bühner, M. (2002). *Development of a Performance Measure of Emotional Intelligence*. In 43rd Congress of the German Psychological Society, Berlin, Germany

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). *afex: Analysis of Factorial Experiments* [Computer software]. Retrieved from https://CRAN.R-project.org/package=afex

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063

Zell, E., & Lesick, T. L. (2021). Taking social comparison to the extremes: The huge-fish-tiny-pond effect in self-evaluations. *Social Psychological and Personality Science, 12*(6), 1030–1038. https://doi.org/10.1177/1948550620956535

Zetsche, U., Bürkner, P. C., & Renneberg, B. (2019). Future expectations in clinical depression: Biased or realistic? *Journal of Abnormal Psychology, 128*(7), 678–688. https://doi.org/10.1037/abn0000452