**Supplementary Material**

Article title: Why dysfunctional expectations in depression persist - Results from two experimental studies investigating cognitive immunization

Journal name: Psychological Medicine

Author names: Tobias Kube, Winfried Rief, Mario Gollwitzer, Thomas Gärtner, Winfried Rief

Corresponding author: Tobias Kube, Philipps-University of Marburg, Department of Clinical Psychology and Psychotherapy, Gutenbergstraße 18, D-35032 Marburg, Germany, Phone: +0049 6421 2823341, Email: tobias.kube@staff.uni-marburg.de

**Computing the TEMINT sum score**

For each situation, participants rate between six and ten emotions of the acting person (e.g. fear, anger, sadness etc.). Participants are asked to rate a total of twelve situations with 85 emotional states. Answers are scored by comparing the participant’s answers with the acting person’s actual ratings. The TEMINT sum score reflects the overall deviations from the actual ratings, with low sum scores indicating good performance.

**Differences between experimental conditions in Study 1**

The expectation confirmation group (*M* = 4.23, *SD* = 1.40) and the expectation disconfirmation group (*M* = 4.46, *SD* = 1.26) did not significantly differ on initial task-specific expectations, *F*(1, 113) = 0.880, *p* = .350; *ɳ²p* = .008, 95% CI [0, 0.068]. Likewise, the expectation confirmation group (*M* = 4.05, *SD* = 1.24) and the expectation disconfirmation group (*M* = 4.25, *SD* = 1.30) did not significantly differ on initial generalized expectations, *F*(1, 113) = 0.723, *p* = .397; *ɳ²p* = .006, 95% CI [0, 0.064]. Further, the two groups did not differ on age, *F*(1, 113) = 0.908, *p* = .343; *ɳ²p* = .008, 95% CI [0, 0.069], depressive symptoms, *F*(1, 113) < 0.001, *p* = .993; *ɳ²p* < .001, 95% CI [0, 0.001], or TEMINT performance, *F*(1, 113) = 1.029, *p* = .313; *ɳ²p* = .009, 95% CI [0, 0.072].

**Differences between participants from the inpatient clinics and the day-care clinic in Study 1**

With regard to the clinical sample, participants from the inpatient clinics and the day-care clinic did not significantly differ on initial task-specific expectations, *F*(1, 56) = 1.737, *p* = .193; *ɳ²p* = .030, 95% CI [0, 0.157], initial generalized expectations, *F*(1, 56) = 0.577, *p* = .451; *ɳ²p* = .010, 95% CI [0, 0.113], age, *F*(1, 56) = 0.642, *p* = .426; *ɳ²p* = .011, 95% CI [0, 0.116], depressive symptoms, *F*(1, 56) = 1.445, *p* = .234; *ɳ²p* = .235, 95% CI [0, 0.148], or TEMINT performance, *F*(1, 56) = 0.154, *p* = .697; *ɳ²p* = .003, 95% CI [0, 0.083].

**Change in task-specific expectations in Study 2**

**Results.** The Time by Condition ANOVA indicated a significant main effect of Time, *F*(2,56) = 19.659; *p* < .001; *ɳ²p* = 0.260, 95% CI [0.202, 0.548], while the main effect of Condition was not significant, *F*(2,56) = 0.779; *p* = .464; *ɳ²p* = 0.027, 95% CI [0, 0.129]. The Time by Condition interaction was significant, *F*(2,56) = 3.204; *p* = .048; *ɳ²p* = 0.101, 95% CI [0, 0.246].

To further examine group differences in expectation change, we computed independent samples *t*-tests with Bonferroni-Holm adjustments. With α1 = 1.7%, the difference in expectation change between the immunization-enhancing and -inhibiting condition was not significant, *t*(36) = -2.454; *p* = .019; *d* = 0.795, 95% CI [0.167, 1.452]. The differences between the control condition and the immunization-enhancing condition, α2 = 2.5%, *t*(36) = 1.688; *p* = .100; *d* = 0.544, 95% CI [-0.105, 1.199], and the immunization-inhibiting condition and the control condition were also not significant, α3 = 5%, *t*(40) = -0.886; *p* = .381; *d* = 0.273, 95% CI [0, 0.859]. Paired-samples *t*-tests indicated significant change in expectations among participants in the immunization-inhibiting condition, *t*(20) = -4.285, *p* < .001, *d* = 0.933, 95% CI [0.420, 1.438] and the control condition, *t*(20) = -3.211, *p* = .004, *d* = 0.757, 95% CI [0.236, 1.168], while no significant change in expectations was found among participants from the immunization-enhancing condition, *t*(16) = -0.496, *p* = .627, *d* = 0.127, 95% CI [0, 0.573].

**Discussion.** Results indicate that, though beyond the main goal of the manipulation, the experimental variation of immunization tendencies also impacted change of task-specific expectations. Expectation change occurred in the immunization-inhibiting condition and in the control group, while participants from the immunization-enhancing condition did not change their task-specific expectations despite expectation-disconfirming feedback. Thus, these results are in line with the results regarding change of generalized expectations. However, the effect size for the group differences was smaller compared to the analysis regarding generalized expectations.

**Considering diagnosis as a covariate in Study 2**

In Study 2, we examined both people reporting elevated levels of depression and people meeting criteria of MDD. To examine whether this distinction influences the results, we performed an additional analysis of covariance (ANCOVA) considering the diagnosis as a covariate.

The ANCOVA indicated no significant main effect of Time, *F*(2,56) = 1.336; *p* = .253; *ɳ²p* = 0.024, or Condition, *F*(2,56) = 0.681; *p* = .510; *ɳ²p* = 0.024, or Diagnosis, *F*(2,56) = 1.214; *p* = .275; *ɳ²p* = 0.022. However, there was a significant Time by Condition interaction, *F*(2,56) = 5.090; *p* = .009; *ɳ²p* = 0.156. The Time by Diagnosis interaction was not significant, *F*(2,56) = 0.357; *p* = .553; *ɳ²p* = 0.006.

Thus, the diagnosis did not have unique effects on the dependent variable, and its inclusion did not change the significance of any of the other main or interaction effects. Effect size in the ANCOVA was similar to the one in the ANOVA for the effect of most interest, that is, the two-way interaction effect. Therefore, it can be concluded that considering the distinction between MDD diagnosis and elevated levels of depression does not significantly influence the results.

**Credibility of the cover story**

**Study 1**. As described in the main text, eleven participants (one depressed individual, ten healthy participants) had serious doubts concerning the credibility of the cover story and suspected the actual purpose of the study. Therefore, these participants were excluded so that analyses were based on data from 117 participants. Twenty-eight out of the remaining participants also suspected that the study could have another aim than the one indicated in the study information, but they did not guess the real purpose of the study. Among these 28 participants, the following reasons of suspicions were mentioned: “the study aimed to examine the participants' self-esteem” (thirteen participants), “the study aimed to examine the relationship between self-esteem and the ability to empathize with others” (five participants), “the study aimed to examine empathy in depression” (five participants), “the study aimed to examine the participants’ own emotions” (one participant), “the study aimed to examine the influence of language on the evaluation of feelings” (one participant). Another three participants suspected another aim although they did not explicitly mention one. No participant mentioned the suspicion that the study could be about participants' expectations.

**Study 2.** Similar to Study 1, two participants were excluded because they had serious doubts concerning the credibility of the cover story and suspected the actual purpose of the study. Sixteen out of the remaining 59 participants suspected that the study could have another aim than the one indicated in the study information, but did not guess the real purpose of the study. The following reasons of suspicions were mentioned: “the study aimed to examine the participants' self-esteem” (six participants), “the study aimed to examine the relationship between self-esteem and the ability to empathize with others” (two participants), “the study aimed to examine empathy in depression” (one participant), “the study aimed to examine the participants’ own emotions” (one participant), “the study was about the participants attitudes” (one participant), “the study aimed to examine how I would have reacted in the given situations” (one participant), “the study aimed to assess my emotional intelligence” (one participant), “the study aimed to examine how individuals cope with ambiguous feedback” (one participant). Another two participants suspected another aim although they did not explicitly mention one. No participant mentioned the suspicion that the study could be about participants' expectations.

# Figure legends of supplementary figures

**Figure A.** Flow-chart diagram for participants from Study 1. *Note*. MDD = major depressive disorder.

**Figure B.** Flow-chart diagram for participants from Study 2. *Note*. MDD = major depressive disorder.

**Figure C.** Illustration of one task from the Test for the Measure of Emotional Intelligence (TEMINT). Participants’ task is to empathize with the person described, and to evaluate his/her feelings. In total, participants are presented twelve situations, and they have to rate 85 emotional states.