Empirical Article

How to Change Negative Outcome Expectations in Psychotherapy? The Role of the Therapist's Warmth and Competence

Anna Seewald iD and Winfried Rief

Department of Clinical Psychology and Psychotherapy, Philipps-University of Marburg

Abstract

Negative outcome expectations of psychological treatments predict unfavorable treatment outcomes. Therefore, therapists should approach negative outcome expectations and ideally transform them into more positive outcome expectations. In this study, we investigated the therapist's interpersonal behavior to optimize the modification of negative outcome expectations. After inducing negative expectations in an online experiment, we presented different videos of therapist–patient interactions to violate the induced negative outcome expectations. We kept the expectation-violating information constant and manipulated the therapist's warmth and competence. Results confirmed a significant influence of the therapist's warmth and competence on expectation violation, which led to the most positive outcome expectations when the therapist was both warm and competent. In contrast to former correlational analyses, our experimental study confirms the causal role of the therapist's interpersonal behavior and its impact on changing patients' negative outcome expectations. On the basis of these findings, more powerful approaches to optimize critical outcome expectations can be developed.

"I can't believe that psychological treatments are helpful for my problem" illustrates pessimistic prognostic beliefs about the therapy effectiveness—so-called negative *outcome expectations* (OE). They could have resulted from direct negative experiences with psychological treatments or other sources of information (Ladwig et al., 2014; Ten Have et al., 2010). Negative OE are particularly problematic because they can impair the success of psychological treatments (Constantino et al., 2011, 2020; Dew & Bickman, 2005; Greenberg et al., 2006). Therefore, negative OE should be targeted within the therapeutic process by providing new experiences that violate the negative treatment expectations. However, even though the relationship between OE and clinical outcome was widely investigated, there were no studies on improving the expectation violation of negative OE. Accordingly, we attempted to close this research gap using an experimental design based on a theoretical framework of expectation violation (ViolEx Model; Rief et al., 2015; Rief & Joormann, 2019).

The Role of Warmth and Competence for Treatment Outcomes and Expectations

Several factors, for instance, the therapeutic alliance, empathy, and expectations, play a significant role in psychological treatments, which were broadly summarized under the term "common factors" (Grencavage & Norcross, 1990; Wampold, 2015). One common factor, the "therapist effect" (Castonguay & Hill, 2017; Crits-Christoph et al., 1991; Johns et al., 2019), including the

Corresponding Author:
Anna Seewald, Department of Clinical Psychology and Psychotherapy, Philipps-University of Marburg
Email: anna.seewald@uni-marburg.de

therapist's characteristics and behavior, moved even more into the spotlight during the past few years. Thereby, different aspects of the therapist's influences need to be disentangled.

Even though the factual information transmitted by diverse therapists can be identical, interpersonal behavior can differ excessively between therapists and can affect the information received by patients (Schulz von Thun et al., 2014). Importantly, the therapist's warmth and competence were considered the two critical dimensions of interpersonal behavior (Howe et al., 2019). The therapist can display either personal engagement and care (warmth) or efficiency, expertise, and knowledge (competence). Furthermore, both dimensions can interact and be present simultaneously (i.e., the therapist explains using a speed tailored to the individual patient). In addition, some studies have discussed a positive correlation between warmth and competence. Therefore, the perception of high warmth could also lead to a higher perception of competence (Howe et al., 2019; Judd et al., 2005; Kraft-Todd et al., 2017).

Using the example of sham acupuncture, Kaptchuk et al. (2008) showed that treatment outcome was much more efficient if the therapist showed warm and empathetic behavior compared with a more technical style of therapeutic interaction. Furthermore, physicians' warmth and competence were associated not only with a faster recovery in the treatment of common colds (Rakel et al., 2011) but also with higher pain tolerances (Czerniak et al., 2016). In short, meta-analyses including both qualitative and quantitative studies emphasized warmth and competence as moderate to strong predictors for positive clinical outcomes (Derksen et al., 2013; Di Blasi et al., 2001; Elliott et al., 2018; Howe et al., 2019; Howick et al., 2018). However, the underlying mechanisms of these effects remained unclear because assumptions about therapists' effects were typically based on correlational results, which limited any causal interpretations.

Some authors hypothesized that the therapist's warmth and competence might have enhanced positive expectations and that these positive expectations were decisive for improved outcomes (Blasini et al., 2018; Howe et al., 2019; Westra et al., 2011; Zion & Crum, 2018). Indeed, competence in conducting cognitive behavioral therapy was associated with better midtreatment OE, which was further connected to better treatment outcomes (Westra et al., 2011). Moreover, a therapist demonstrating warm behavior resulted in more positive OE after an initial therapy session (Ahmed et al., 2012). In addition, the placebo effect of an antiallergy skin cream was enlarged by a warm and competent practitioner (Howe et al., 2017). This result suggested that the widely investigated influence of positive expectations on clinical outcomes

was strengthened by warmth and competence. To summarize, a variety of research underpins the relationship between warmth and competence and outcomes, and some authors have assumed the enhancement of positive expectations as an underlying mechanism for this relation.

Expectation Change in Psychological Treatments

Expectations have played important roles not only in medical studies, but also in psychological treatments in which expectations predicted positive outcomes and were necessary to understand and treat psychological disorders (Rief & Glombiewski, 2016). Psychological interventions were more and more considered processes that violated patients' dysfunctional expectations. Using a theoretical framework for expectation violation, we found that the credibility of the source providing expectation violation—in psychological treatments, usually the therapist—was assumed as one of the crucial factors determining expectation violations (ViolEx model; Rief et al., 2015; Rief & Joormann, 2019). This model is also relevant for investigating the change from critical treatment OE into more positive attitudes. A very understanding and experienced therapist may violate negative expectations and increase patients' positive expectations about treatments' effectiveness and thereby improve clinical outcomes.

Critically, to our knowledge, no study to date has investigated the influence of warmth and competence on expectation violation directly. However, the violation of negative OE to create positive OE is crucial for positive therapeutic outcomes in many clinical interactions (i.e., in situations with increased risk for treatment discontinuation or in clinical conditions with low motivation for psychological treatment, e.g., substance use disorder or chronic pain; Raylu & Kaur, 2012; Tse et al., 2013). This was further underpinned by the findings that patients with a generalized anxiety disorder had more positive treatment outcomes when they experienced more pleasant surprises and disconfirmation of initial negative OE and a higher expectation change in the first therapy sessions (Newman & Fisher, 2010; Westra et al., 2010). Expectation changes during initial therapy sessions varied notably between different therapists (Ahmed et al., 2012; Vîslă et al., 2019). Therefore, ways to optimize expectation violation should be investigated by taking warmth and competence into account.

Research Question

In detail, in this research, we aimed to investigate the specific role of a therapist's interpersonal behavior to

influence and optimize expectation violation of negative OE. Because previous studies had limited standardization, we used an experimental study design. We first manipulated negative OE before violating them via videos of therapist–patient interactions (see Moors & Zech, 2017). We kept the expectation-violating information constant and manipulated the interpersonal behavior of the therapist within the videos, such as being warm or competent. We predicted a causal effect of the therapist's warmth and competence on changing negative OE and that these effects would be amplified when the therapist showed both features. Therefore, we expected the most positive OE in the group in which the therapist was both warm and competent and no significant difference between the high-competence and high-warmth groups (high competence/high warmth > high competence/low warmth = high warmth/low competence > low warmth/low competence). Finally, we employed Bayesian analyses to quantify the strength of our evidence and conquer the shortcomings of conventional analyses.

Method

Participants

For our 2 × 2 between-subjects design, our prior recruitment goal was set to 180 participants (45 per cell) because this would provide us with adequate power (1 − β = 0.9, α = .05) to detect a medium effect size (Howe et al., 2017) calculated using G*Power (Faul et al., 2007, 2009). We recruited online via university mailing lists and social media from May 11 to 14, 2021. For our experiment, 202 participants volunteered; 11 did not meet our inclusion criteria, and four had to be excluded because they fast-forwarded the therapist video. A total sample of 187 participants remained (M = 24.76 years, SD = 5.20; range = 18–55 years), including 67 men and 120 women (nobody identified as having a nonbinary gender; Table 1). Inclusion criteria were (a) being at least 18 years old, (b) speaking German as a native language or at a native-language level, (c) having no visual or hearing impairment that would affect video viewing, (d) having no neurological disorders, (e) having no diagnosed major mental disorder, and (f) receiving no current treatment for any psychological disorder.

Study design

This study was a double-blind 2 × 2 factorial between-subjects design: Warmth (low vs. high) × Competence (low vs. high). OE were measured as the main dependent variable three times: before the expectation manipulation (baseline OE); after the negative manipulation via an auditory scenario that included negative

information about the therapy effectiveness (negative OE); and after the expectation violation via four different videos (counterbalanced warmth and competence of the displayed therapist), which included positive information about the therapy effectiveness (positive OE). The study was approved by the local ethics committee of the University of Marburg (Reference 2020-85k) and was preregistered at OSF (https://doi.org/10.17605/osf.io/qc8vf).

Material and procedure

Step 1. Baseline assessment. This online study was implemented in SoSci Survey (Leiner, 2019). Participants were told that this study was intended to examine the perceived effectiveness of stress-management methods in psychotherapy. First, participants gave informed consent, completed a technical check, provided demographic data, and filled out initial questionnaires (Patient Health Questionnaire [PHQ-9], Kroenke & Spitzer, 2002; Generalized Anxiety Disorder Screener [GAD-7], Spitzer et al., 2006; Big Five Inventory [BFI-10], Rammstedt & John, 2007; generic rating scale for previous treatment experiences, treatment expectations, and treatment effects [GEEE] Rief et al., 2021; and Perceived Stress Scale [PSS-10], Cohen et al., 1983). After that, they were instructed to listen carefully to the first part of a scenario (4 min, 26 s) and put themselves in the patient's shoes as much as possible. In this part, the clinical problem was introduced by describing a person who was very stressed at work (e.g., "Thoughts are racing through your head. You wonder how you are going to manage it all"). The scenario was spoken by a male professional speaker and used simple and short sentences in the present tense (Kirn et al., 2015). We chose this topic because 46% of the German population feel stressed by school, university, or their jobs (Techniker Krankenkasse, 2016). Thus, we assumed a high relevance of this topic and a high self-involvement because most people could identify with the described person. After assessing baseline OE, the second step of the study followed.

Step 2. Negative OE manipulation. Negative OE were manipulated via the second part of the scenario (1 min, 54 s). In this part, the stressed person from the first part recalled a past initial consultation in a psychotherapy practice. During this consultation, the person received negative information about the effectiveness, treatment success, and treatment rationale of psychotherapy to treat stress problems (Constantino, 2012; Kazdin & Krouse, 1983). For instance, "You learned that there are only a few studies on the effectiveness of stress management" (effectiveness). After negative OE was assessed, the main step of the study followed.

Table 1. Sociodemographics Dependent on Group

Characteristic	Low competence/ low warmth	High competence/ low warmth	High warmth/ low competence	High competence/ high warmth
N total	47 (25.13)	45 (24.06)	47 (25.13)	48 (25.67)
Gender				
Male	18 (38.3)	16 (35.6)	16 (34.0)	17 (35.4)
Female	29 (61.7)	29 (64.4)	31 (66.0)	31 (64.6)
Nationality				
German	38 (80.9)	40 (88.9)	42 (89.4)	42 (87.5)
German and other	6 (12.8)	5 (11.1)	4 (8.5)	3 (6.3)
Other	3 (6.4)	0 (0.0)	1 (2.1)	3 (6.3)
Education				
Primary/secondary level	2 (4.3)	3 (6.7)	4 (8.5)	2 (4.2)
Higher level[a]	45 (95.7)	42 (93.3)	43 (91.5)	46 (95.8)
GEEE-psychotherapy				
No experience	35 (74.5)	40 (88.9)	43 (91.5)	37 (77.1)
Some experience	12 (25.5)	5 (11.1)	4 (8.5)	11 (22.9)
GEEE-stress management				
No experience	32 (68.1)	34 (75.6)	37 (78.7)	36 (75.0)
Some experience	15 (31.9)	11 (24.4)	10 (21.3)	12 (25.0)

Note: Values are *n*s with percentages in parentheses. GEEE = generic rating scale for previous treatment experiences, treatment expectations, and treatment effects (Rief et al., 2021).
[a]Higher education level includes currently studying at a university or having already obtained a university degree.

Step 3. Expectation violation: positive OE manipulation. Participants were randomized (stratified by gender) to one of four therapist–patient-interaction videos (low competence/low warmth vs. high competence/low warmth vs. high warmth/low competence vs. high competence/high warmth). All videos had comparable durations (low competence/low warmth: 4 min, 27 s; high competence/low warmth: 4 min, 18 s; high warmth/low competence: 4 min, 24 s; high competence/high warmth: 4 min, 16 s) and used the same setting and a male therapist (White man, middle-aged). To increase self-involvement, the videos were filmed over a woman's shoulder (kept gender neutral), in an attempt to induce in the patients the feeling that they were directly talking to the therapist. We tried to violate the previously induced negative OE by targeting the same categories as in the negative manipulation (effectiveness, treatment success, and treatment rationale) and giving more promising information about the treatment of stress problems (e.g., "Recent study results show that the effectiveness of current treatment methods is very high," effectiveness). Although we kept the expectation-violating information transmitted constant in all four groups, the interpersonal behavior of the therapist was modulated by counterbalancing warm (e.g., no smiling vs. smiling) and competent (e.g., low level of expertise vs. high level of expertise) behavior across the four videos (adapted from Howe et al., 2017, 2019). The assessment of positive OE and final questionnaires followed (therapist's warmth and competence,

therapeutic alliance, therapy motivation, and agreement with the therapist's arguments). The experiment lasted approximately 30 min, and participants were reimbursed €10. The study procedure is displayed in Figure 1.

Pilot studies of the stimulus material. We tested the negative OE manipulation and warmth and competence manipulations in two pilot studies with participants who did not take part in the present study. In the first pilot study ($N = 14$), we compared one version of the scenario spoken by a female speaker with another version spoken by a male speaker. Only the version with the male speaker led to a significant reduction of OE, which is why we used this version in the present study; main effect timing, female speaker: $\chi^2(1) = 0.36$, $p = .56$, Bayes's factor (BF) $BF_{10} = 0.50$; male speaker: $\chi^2(1) = 4.41$, $p = .036$, $BF_{10} = 1.65$.

In the second pilot study ($N = 23$), participants watched the videos and rated the therapist's warmth and competence. Warmth scores were significantly predicted by the warmth manipulation, $\chi^2(1) = 11.79$, $p < .001$, $BF_{10} = 12.49$, and competence scores were significantly predicted by the competence manipulation, $\chi^2(1) = 4.70$, $p = .030$, $BF_{10} = 1.67$, which indicated a sufficient manipulation. Table S1 in the Supplemental Material available online presents the content of the expectation violation, and Table S2 presents the manipulation of warmth and competence; Figure S1 displays a video screenshot.
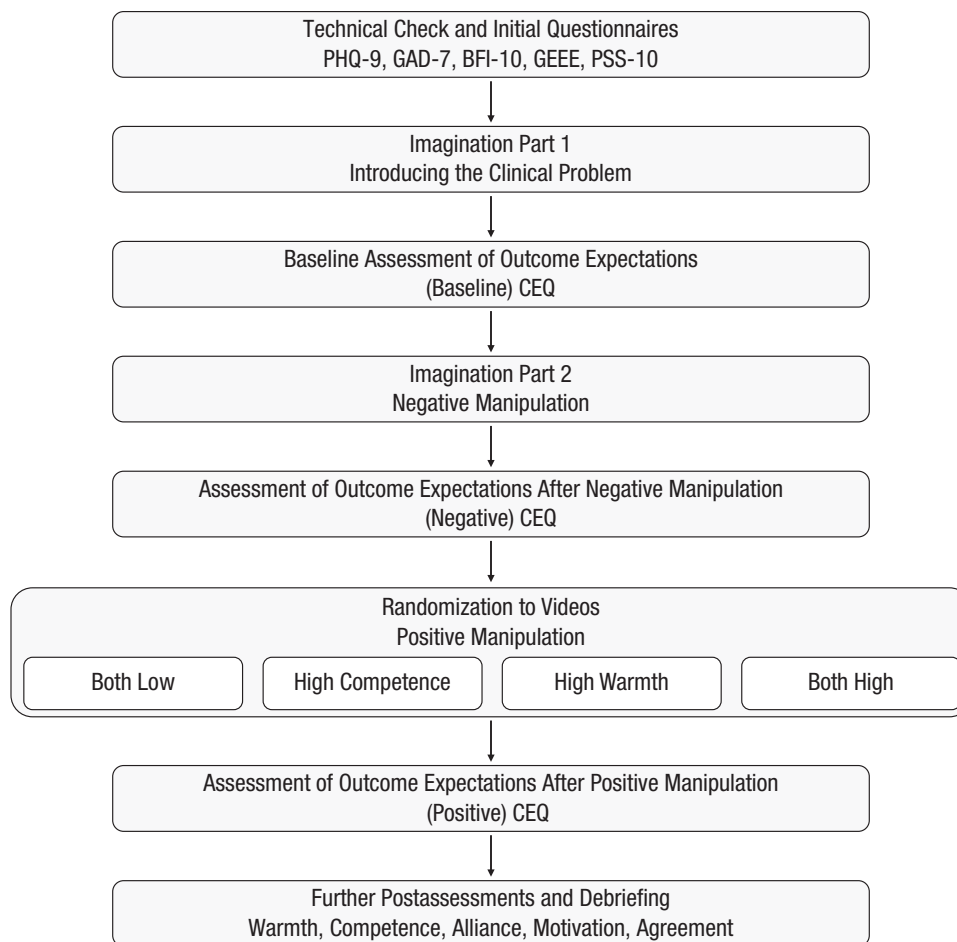
Fig. 1. Experimental procedure.

Questionnaires

Control variables. Previous studies indicated that depression, anxiety, or neuroticism could influence expectation change (Barlow et al., 2014; Grupe & Nitschke, 2013; Kube et al., 2019; Rychlak & Lerner, 1965). Furthermore, stress and experiences with psychological treatments may have influenced expectations in this experiment. Therefore, we assessed those variables to control for differences across our four groups.

PHQ-9. Participants completed the German version of the PHQ-9 depression module (Gräfe et al., 2004) to assess the severity of depressive symptoms in the past 14 days. They rated nine questions on a 4-point scale from 0 (*not at all*) to 3 (*nearly every day*), which resulted in sum scores from 0 to 27 and a clinical screening cutoff point of 10. The PHQ-9 demonstrated high internal consistency (α = .88) and good criterion and discriminant validity in a German sample (Gräfe et al., 2004).

GAD-7. Participants completed the German version of the GAD-7 (Löwe et al., 2008) to detect symptoms of

generalized anxiety disorder. They rated seven questions on a 4-point scale from 0 (*never*) to 3 (*nearly every day*), which resulted in sum scores from 0 to 21 and a clinical screening cutoff point of 10. The GAD-7 demonstrated high internal consistency (αs = .89–.92), high test–retest reliability (r_{tt} = .83), and good construct validity (Löwe et al., 2008; Spitzer et al., 2006).

PSS-10. Participants completed the German version of the PSS-10 (E. E. Schneider et al., 2020) to measure stress symptoms in the past month. They rated 10 questions on a 5-point scale from 1 (*never*) to 5 (*very often*). The PSS-10 total sum scores ranged from 10 to 50 and demonstrated high internal consistency (α = .89) and good construct validity in a nonclinical German sample (E. E. Schneider et al., 2020).

GEEE. Participants completed the GEEE (Rief et al., 2021) to measure (a) previous experiences with psychotherapy (GEEE-T) and (b) stress-management methods (GEEE-S). Participants had to rate whether they experienced (a) psychotherapy or (b) stress-management methods *never*, *daily*, *more than 10 days*, *5 to 10 days*, *1 to*

4 days, or *not during the last 12 months, but before*. We transformed the scale into a binary variable (0 = no experience, 1 = some experience).

BFI-10. Participants completed the German version of the BFI-10 (Rammstedt & John, 2007) because it reflects a widely used model to assess personality. They rated 10 questions on a 5-point scale from 1 (*does not apply at all*) to 5 (*fully applies*). For the Neuroticism factor, we used a mean score of the two neuroticism items (range = 1–5), which reflect how emotionally labile a person is. The two neuroticism items demonstrated sufficient test–retest reliability (r_{tt} = .49) and good construct validity in a German sample.

Variables for main analysis—OE, warmth, and competence.

Credibility Expectancy Questionnaire. Participants completed a German version (Devilly & Borkovec, 2000) of the Credibility Expectancy Questionnaire (CEQ; Koch et al., 2016), the most widely used questionnaire to measure OE. Four items were answered on a 9-point scale from 1 to 9, and two items were answered on an 11-point scale from 0% to 100% and transformed into a 9-point scale for analysis. Four items asked how participants think (e.g., "At this point, how logical does a therapy offered to you seem?"), and two items asked how participants feel (e.g., "At this point, how much do you really feel that therapy will help you to reduce your stress symptoms?") about psychotherapy to reduce their stress symptoms. We used a total mean score (range = 1–9), which demonstrated high internal consistency (αs = .84–.85) and good construct validity in three different samples (Devilly & Borkovec, 2000).

Warmth and competence. Participants rated the therapist's warmth and competence with six items for warmth and six items for competence from the Stereotype Content Model (SCM; Fiske et al., 2002). They answered the question "How did the therapist seem to you?" by rating six items for warmth (English: friendly, well-intentioned, trustworthy, warm, good-natured, sincere; German: *freundlich, wohlwollend, vertrauenswürdig, warmherzig, gutmütig, aufrichtig*) and six items for competence (English: competent, confident, capable, efficient, intelligent, skillful; German: *kompetent, selbstbewusst, fähig, effizient, intelligent, qualifiziert*). Each item was rated on a 5-point scale from 1 (*not at all*) to 5 (*extremely*); row means (range = 1–5) reflected warmth and competence scores. These items demonstrated good internal consistency for the warmth (α = .90) and competence scales (α = .94) in stereotype research (Fiske et al., 2002).

Exploratory variables. We measured working alliance, therapy motivation, and agreement with the therapist's

statements for an exploratory test of the influence of warmth and competence on these variables.

Working Alliance Inventory, short version. Participants completed the German version (Wilmers et al., 2008) of the Working Alliance Inventory, short version (WAI-S; Horvath & Greenberg, 1989) to assess the quality of therapeutic alliance. They rated 12 questions on a 5-point scale from 1 (*not at all*) to 5 (*very*), and a total sum score (range = 12–60) was calculated. The WAI-S demonstrated high internal consistency (αs = .90–.93) for the total score and good convergent and criterion validity.

Therapy motivation. Participants rated their motivation to do psychotherapy on one self-generated item ("How motivated are you to do psychotherapy to work on your stress symptoms?") on a visual analogue scale from 0 (*not at all motivated* or 0%) to 10 (*fully motivated* or 100%).

Agreement. We measured agreement with the therapist's statements used for the expectation violation. Participants rated the therapist's statements (e.g., "Recent study results show that the effectiveness of current treatment methods is very high"; see Table S1 in the Supplemental Material) using four self-generated items. We used a 5-point scale from 1 (*not at all*) to 5 (*very*), and sum scores (range = 5–20) were calculated. Therefore, higher scores indicated higher agreement with the therapist's arguments used for expectation violation.

Statistical analysis

We conducted all analyses using the R software environment (Version 4.1.0; R Core Team, 2021). The primary confirmatory analysis was fully preregistered in every detail. We preregistered that we would conduct likelihood-ratio tests for all other analyses (control analyses, manipulation checks, and exploratory analyses). For our main variables, we checked the internal consistency by calculating Cronbach's α. Furthermore, we analyzed our data with likelihood-ratio tests, which compares two nested models: one model including the considered effect compared with a restricted model without the considered effect (*lmtest* package, *lrtest* function, Version 0.9-38; Zeileis & Hothorn, 2002). Unless described otherwise, we first calculated two model comparisons to test the main effects of warmth and competence against the null model. Furthermore, we tested the model that included the interaction of warmth and competence against the model without the interaction. A significant result indicated that the model that included the considered effect significantly better described the data. Some shortcomings of these

Table 2. Mean Age and Questionnaire Scores by Group

Characteristic	Group				$F(3, 183)$	p	BF_{01}
	Low competence/ low warmth	High competence/ low warmth	High warmth/low competence	High competence/ high warmth			
Age (years)	25.70 (6.40)	23.98 (4.32)	24.64 (4.22)	24.69 (5.52)	0.86	.46	13.03
PHQ-9	6.30 (4.12)	6.93 (5.35)	5.70 (4.38)	6.48 (4.06)	0.59	.62	18.07
GAD-7	4.21 (3.35)	5.67 (4.85)	4.91 (3.51)	4.69 (4.30)	1.03	.38	10.63
Neuroticism	3.03 (0.78)	3.06 (0.98)	2.89 (0.88)	3.03 (0.90)	0.32	.81	25.03
PSS-10	26.49 (6.04)	29.02 (7.41)	26.87 (6.73)	26.85 (7.83)	1.23	.30	8.39

Note: Values under the Group heading are mean scores with standard deviations in parentheses unless otherwise specified. PHQ-9 = Patient Health Questionnaire (Kroenke & Spitzer, 2002); GAD-7 = Generalized Anxiety Disorder Screener (Spitzer et al., 2006); neuroticism = neuroticism score extracted from the Big Five Inventory (Rammstedt & John, 2007); PSS-10 = Perceived Stress Scale (Cohen et al., 1983); BF = Bayes's factor. BF_{01} demonstrates evidence strength in favor of the null hypothesis.

conventional null hypothesis significance tests are that we cannot gather evidence in favor of the null hypothesis and that large samples can lead to an overestimation of the evidence against the null hypothesis and therefore to an increased risk of false-positive results (Benjamin et al., 2018; Edwards et al., 1963; Rouder et al., 2009; Wagenmakers et al., 2011).

To tackle these problems, we additionally calculated BFs (*BayesFactor* package, Version 0.9.12-4.2; Morey & Rouder, 2018). We incorporated default priors because we had no a priori assumptions about the underlying effect size (increased sample number to 100,000; Rouder et al., 2012). The BF encompasses the probability (marginal likelihood) of the data given one hypothesis relative to another hypothesis (Jeffreys, 1961; Kass & Raftery, 1995). We report BF_{01}, which indicates evidence in favor of the null hypothesis, and BF_{10}, which indicates evidence in favor of the alternative hypothesis. For instance, a BF_{10} of 5 means that it is 5 times more likely that the data resulted under the alternative hypothesis compared with the null hypothesis. It is possible to also state evidence for the null hypothesis with a lower risk for false-positive results (Wagenmakers et al., 2011; Wetzels et al., 2011). BFs were interpreted according to Jeffreys (1961): Values between 1 and 3 indicate anecdotal evidence, values between 3 and 10 indicate moderate evidence, values between 10 and 30 indicate strong evidence, values between 30 and 100 indicate very strong evidence, and values higher than 100 indicate extreme evidence.

Results

Sociodemographics

Demographic variables confirmed the random assignment (stratified by gender) to the four groups (Table 1).

The results indicated no significant group differences in terms of the distribution of gender, $\chi^2(3) = 0.19$, $p = .98$, $BF_{01} = 88.59$; nationality, $\chi^2(6) = 5.16$, $p = .52$, $BF_{01} > 100$; education level, $\chi^2(3) = 1.12$, $p = .77$, $BF_{01} > 100$; and previous experiences with psychotherapy, $\chi^2(3) = 7.08$, $p = .07$, $BF_{01} = 6.25$, or stress management, $\chi^2(3) = 1.48$, $p = .69$, $BF_{01} = 63.29$. BFs demonstrated moderate to extreme evidence for the null effect. Exploratory results indicate that in the baseline assessment of OE, female participants had significantly more positive baseline OE compared with male participants (female: $M = 6.34$, $SD = 1.37$; male: $M = 5.80$, $SD = 1.33$), $t(140.03) = 2.66$, $p = .009$, $BF_{10} = 4.06$. After our negative and positive manipulations, no significant gender effects remained, which indicates a sufficient manipulation. No other effects on baseline OE were significant.

Group comparisons of control variables

The results indicated no significant group differences in all investigated control variables. BFs demonstrated moderate to strong evidence for the null effect (Table 2). All mean scores were well below the corresponding clinical cutoff criteria.

Manipulation check

Manipulation of negative OE. We used a total mean score of the CEQ ($\alpha s = .85–.97$) for the assessment of OE. To examine whether the manipulation of negative OE worked and was similar in all groups, we tested nested models including the factors timing (baseline, negative) and group (low competence/low warmth, high competence/low warmth, high warmth/low competence, high competence/high warmth). Timing significantly predicted negative OE, and there was no significant group or interaction effect: timing: $\chi^2(1) = 177.26$, $p < .001$;
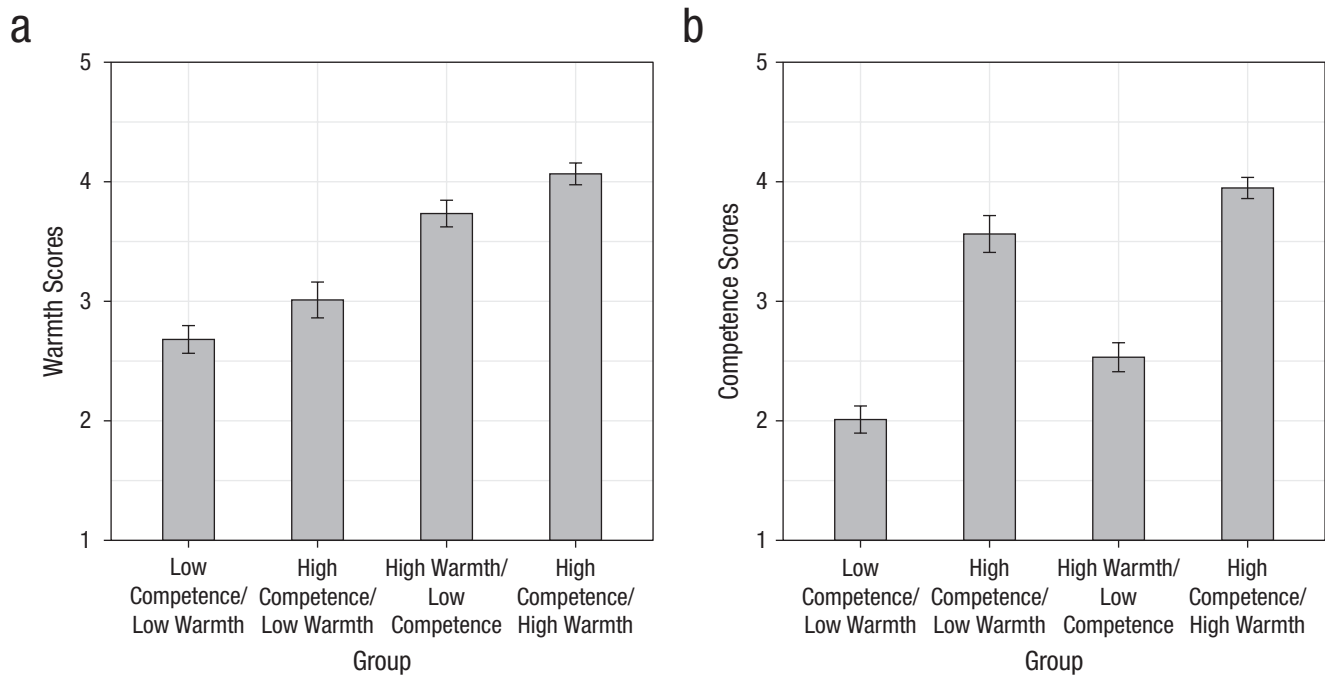
a

b

Fig. 2. Mean scores for (a) warmth and (b) competence for each group. Error bars indicate ± 1 *SE*.

group: $\chi^2(3) = 2.26$, $p = .52$; interaction: $\chi^2(3) = 0.39$, $p = .94$. BFs demonstrated extreme evidence for the timing effect ($BF_{10} > 100$) and very strong evidence for no group or interaction effect (group: $BF_{01} = 32.79$; interaction: $BF_{01} = 32.26$). Independent of the group, all participants had more negative OE after the negative manipulation (base: $M = 6.15$, $SD = 1.38$; negative: $M = 3.59$, $SD = 1.88$). All mean OE scores dependent on time and group are displayed in Figure 3 and Table S3 in the Supplemental Material.

Manipulation of warmth. All warmth items demonstrated high internal consistency in this study ($\alpha = .92$). Warmth and competence significantly predicted warmth scores, and there was no interaction effect; warmth: $\chi^2(1) = 65.85$, $p < .001$; competence: $\chi^2(1) = 6.09$, $p = .014$; interaction: $\chi^2(1) < 0.001$, $p = .99$. BFs demonstrated extreme evidence for warmth ($BF_{10} > 100$) and only anecdotal evidence for competence ($BF_{10} = 2.68$) as predictors for warmth scores and moderate evidence for no interaction effect ($BF_{01} = 4.61$).

Manipulation of competence. All competence items demonstrated high internal consistency in this study ($\alpha = .95$). Warmth and competence significantly predicted competence scores, and there was no interaction effect; warmth: $\chi^2(1) = 8.56$, $p = .003$; competence: $\chi^2(1) = 107.18$, $p < .001$; interaction: $\chi^2(1) = 0.33$, $p = .57$. BFs demonstrated extreme evidence for competence ($BF_{10} > 100$)

and only moderate evidence for warmth ($BF_{10} = 8.51$) as predictors for competence scores and moderate evidence for no interaction effect ($BF_{01} = 3.96$). All mean warmth and competence scores dependent on time and group are displayed in Figure 2.

Main analysis: OE dependent on warmth and competence of therapist

To control for variances of negative OE, we included negative OE as a covariate in the model comparisons. Warmth and competence significantly predicted OE, and there was no interaction effect; warmth: $\chi^2(1) = 7.88$, $p = .005$; competence: $\chi^2(1) = 49.33$, $p < .001$; interaction: $\chi^2(1) = 0.015$, $p = .90$. BFs demonstrated extreme evidence for warmth ($BF_{10} > 100$) and extreme evidence for competence ($BF_{10} > 100$) as predictors for OE and moderate evidence for no interaction effect ($BF_{01} = 4.68$). Therefore, high competence/high warmth group had the most positive OE. Furthermore, a significant difference between the high warmth/low competence and high competence/low warmth groups resulted, $t(89.71) = 2.79$, $p = .006$; BFs indicated moderate evidence ($BF_{10} = 6.12$) for more positive OE in the high-competence group compared with the high-warmth group (high competence/high warmth > high competence/low warmth > high warmth/low competence > low competence/low warmth). All mean OE with error bars are displayed in Figure 3.
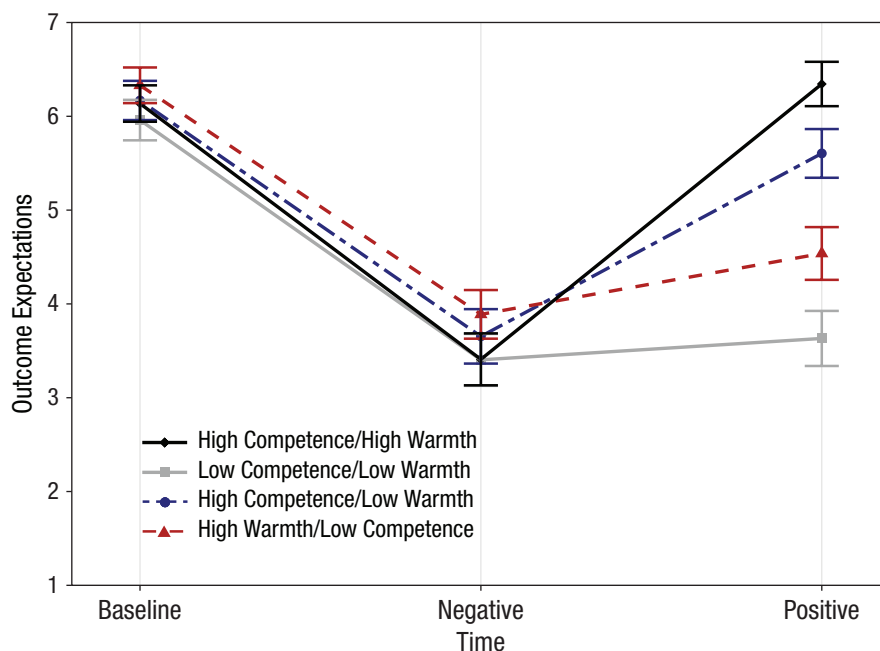
Fig. 3. Mean outcome expectations (OE) for each group at baseline, after the negative OE manipulation, and after the positive OE manipulation. Error bars indicate ±1 *SE*.

Exploratory analyses

Working alliance and therapy motivation. Warmth and competence significantly predicted working alliance scores, and there was no interaction effect; warmth: $\chi^2(1) = 34.77$, $p < .001$; competence: $\chi^2(1) = 33.43$, $p < .001$; interaction: $\chi^2(1) = 0.075$, $p = .78$. BFs demonstrated extreme evidence for warmth ($BF_{10} > 100$) and extreme evidence for competence ($BF_{10} > 100$) as predictors for working alliance scores and moderate evidence for no interaction effect ($BF_{01} = 4.45$).

Likewise, warmth and competence significantly predicted therapy motivation scores, and there was no interaction effect; warmth: $\chi^2(1) = 20.43$, $p < .001$; competence: $\chi^2(1) = 45.17$, $p < .001$; interaction: $\chi^2(1) = 0.31$, $p = .58$. BFs demonstrated extreme evidence for warmth ($BF_{10} > 100$) and extreme evidence for competence ($BF_{10} > 100$) as predictors for therapy motivation scores and moderate evidence for no interaction effect ($BF_{01} = 3.93$). All mean alliance and motivation scores with error bars are displayed in Figure 4.

Agreement. The results indicated no significant group differences regarding participants' agreement with the arguments used for the expectation violation. BFs demonstrated anecdotal evidence for the null effect, $\chi^2(3) = 7.36$, $p = .061$, $BF_{01} = 1.89$ (low competence/low warmth: $M = 14.74$, $SD = 3.07$; high competence/low warmth: $M = 15.89$, $SD = 2.42$; high warmth/low competence: $M = 15.55$, $SD = 2.84$; high competence/high warmth: $M = 16.04$, $SD = 1.54$).

Discussion

To the best of our knowledge, this is the first study to initially induce negative OE and then investigate the influence of the therapist's warmth and competence on subsequent expectation violation using a standardized experimental study design. Expectation violation was stronger, and therefore, more positive OE resulted when the therapist was both warm and competent. These results extend the previously demonstrated relevance of warmth and competence for clinical outcomes (Derksen et al., 2013; Di Blasi et al., 2001; Elliott et al., 2018; Howe et al., 2019; Howick et al., 2018) by demonstrating their causal relevance on expectation violation of negative OE into positive OE.

In contrast to previous evidence, which suggested both dimensions to be equally relevant (Howe et al., 2017), in this study, OE were more positive in the group with high competence/low warmth compared with high warmth/ low competence. We discuss three possible reasons for the stronger effect of competence in the following. First, this difference might have resulted because competence led to a stronger persuasion compared with warmth (e.g., "He/she seems confident, so he/she might be right"; Geers et al., 2018; Greenberg, 1969). However, according to our data, agreement with the therapist's statements did not differ across groups, which renders this explanation rather implausible. Instead, we speculate a direct effect of competence on OE (e.g., "He/she seems confident, so I think psychological treatments will help me").
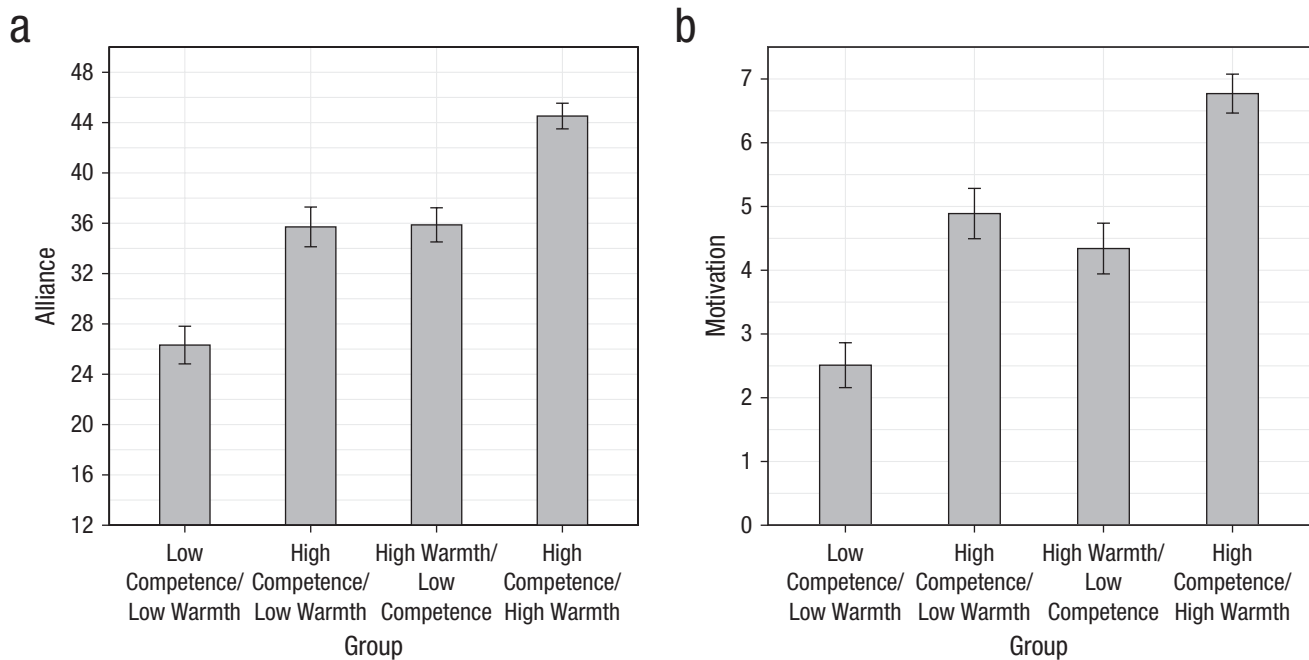
Fig. 4. Mean scores for (a) alliance and (b) motivation for each group. Error bars indicate ±1 *SE.*

Second, the intensity of our competence manipulation could have been stronger than the intensity of our warmth manipulation, which would have led to this pattern of results. This would indicate that the competence of the highly competent therapist was more pronounced than the warmth of the highly warm therapist. However, our descriptive data showed no quantitative difference between the warmth and competence ratings, which does not suggest a stronger competence manipulation.

Third, another possible explanation for the stronger effect of competence could be the lack of pathological distress in this study's healthy sample. We assume that the personal relevance in this study was relatively low, considering that we used an imaginary study design and the cover story of stress-related problems contained a more "technical" problem. This may have led to less need for warmth and a higher preference for the therapist's competence (i.e., "A therapist who understands me but does not contain the competence to help me renders my beliefs of beneficial psychological treatments outcomes rather low"). Accordingly, this might be the reason for a higher expectation change because of competence in this study. However, as suggested by the current literature, both dimensions might be equally relevant concerning personal topics (i.e., "The therapist has to understand my problem first before she/he can use her/his knowledge to help me"; Howe et al., 2019). Therefore, results should be extended by diverse topics with high personal relevance, for example, by letting participants choose the topic themselves. Moreover, warmth could be even more important in clinical samples, or the preference for warmth and competence might differ across different disorders and other clinical challenges, which should be tested in future studies. The present study outlines experimental research that could be implemented in a clinical setting to increase the external validity (see Flückiger & Grosse Holtforth, 2008).

In addition, our exploratory analyses revealed that warmth and competence also influenced therapeutic alliance and motivation. These results further support their importance in psychological treatments not only for expectation violation but also for other relevant factors. The therapeutic alliance (Ardito & Rabellino, 2011; Kelley et al., 2014; Wampold, 2018) and motivation (Rumpold et al., 2005; W. Schneider & Klauer, 2001) were reported as predictors for positive clinical outcomes, and therefore, options to improve them are helpful. Thus, therapists should use warm and competent interpersonal behavior not only to optimize OE but also to improve the therapeutic alliance and therapy motivation, which may all contribute to a better treatment outcome.

Limitations and future directions

In the following, we discuss four possible limitations of the outlined study. First, our manipulation check of warmth and competence scores revealed some

shortcomings of our design. It is difficult to increase one of the two dimensions without influencing the other, at least in the participants' subjective perception. Our data indicate that some of our manipulations increased both attributes. For example, "providing an exact explanation"—in this study manipulated as competence—could be associated with both competent and warm behavior. This behavior requires the competence of transferring knowledge but also providing explanations for the patient's interest, which could appear warm and empathic (Howe et al., 2019). Alternatively, an overall positive correlation between warmth and competence judgments may explain this overlap. This would indicate that a higher rating on one dimension automatically leads to a higher perception of the other dimension (Judd et al., 2005; Kraft-Todd et al., 2017). If corresponding ratings are used, participants might have problems disentangling the psychological concepts of warmth and competence. Even though our warmth and competence manipulations were not as clearly distinguishable as we assumed, our Bayesian analyses revealed a sufficient manipulation. Overall, warmth scores were much stronger predicted by the warmth manipulation compared with the competence manipulation and vice versa. However, the overlap of the two constructs limits a distinct interpretation of the separate influences of warmth and competence.

Second, the present study results are limited to a White, middle-aged, male therapist. We tried to tackle this shortcoming by stratifying the participant's gender to control for possible gender effects and ensuring that all sample characteristics did not differ across groups. Nevertheless, the therapist's nationality, age, and gender could have had a notable influence on warmth and competence perceptions because these aspects already include different assumptions of warmth and competence (Fiske et al., 2002, 2007). In addition, we did not ask for the participant's ethnic background, but almost every participant had German citizenship and was highly educated. According to previous studies, race and education can influence therapy expectations (Silverman et al., 2021) or the relevance of warmth and competence (Fiske et al., 2002, 2007). Therefore, our results should be expanded with therapists and participants of different genders, ages, and cultural backgrounds.

Third, we used an imaginary design to enhance standardization and manipulate negative OE. Even though we used a highly relevant topic, scenario, and video settings to keep self-involvement as high as possible, personal relevance could have been relatively low in this study. Participants' stress levels in this study ($M = 27.29$, $SD = 7.05$) were comparable with the stress levels found in a large validation study ($M = 28.33$, $SD = 6.97$) in a nonclinical sample (E. E. Schneider et al., 2020).

Given that a score of 30 would indicate that participants rated all stress items with "sometimes," it is unlikely that these scores reflect excessive stress levels. Therefore, participants might have lacked high personal relevance for the effectiveness of psychotherapy to treat stress. According to the elaboration-likelihood model (Petty & Briñol, 2012; Petty & Cacioppo, 1986), low personal relevance could have led to a lower level of processing. Low-level processing is characterized by the usage of simple heuristics such as the attractiveness of the source (i.e., using only the perception of the therapist's warmth and competence to form expectations rather than precisely evaluating the given information; Geers et al., 2018). Thus, this may be the reason that we did not find influences on agreement with the therapist's statements but, rather, a direct effect of warmth and competence on OE. In future studies using clinical samples, self-involvement might be higher, which could result in a deeper level of processing and a careful evaluation of the arguments given by the therapist. Therefore, the content may play a more important role in influencing the expectation violation of negative OE in clinical samples.

Last, to operationalize OE, we used a self-report measure. Even though this was characterized by high validity and reliability, a challenging problem of self-report measures can be the participants' bias to respond in favor of the hypothesis because of the presented study material (demand effect; Orne, 1962). In this case, the self-report measure would be more likely to reflect that participants paid attention and understood the study content. However, this would not reflect the overall goal to measure participants' expectations. To bridge these research gaps, participants' implicit expectations or behavioral tendencies need to be assessed to control for such possible biases. For this purpose, more studies should combine implicit and explicit measures to make transparent in what manner these measures relate and represent different aspects of expectation violation to aim not only for a change in reported expectations but also in behavior toward seeking psychological treatment.

Conclusion

In conclusion, this study made significant contributions to the field of modifying patients' expectations in psychotherapy. Although many assumptions about the role of psychotherapists' interpersonal behavior are based on correlational analyses with limited implications for causality, we were able to use an experimental design that first standardized the development of negative expectations and then varied therapist's interpersonal behavior along the dimensions of warmth and

competence. Warmth and competence were shown to have a causal influence on the effect of expectation violation of negative OE. These insights can give therapists helpful behavioral guidance to improve expectation violation, especially for the treatment of patients holding negative OE. Future studies should aim to investigate that improving outcome expectations by optimizing the warmth and competence of therapists will finally lead to optimized treatment outcomes.

Transparency

ORCID iD

Anna Seewald [iD] https://orcid.org/0000-0003-0992-9361

Acknowledgments

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/21677026221094331

References

Ahmed, M., Westra, H. A., & Constantino, M. J. (2012). Early therapy interpersonal process differentiating clients high and low in outcome expectations. *Psychotherapy Research*, *22*(6), 731–745. https://doi.org/10.1080/10503307.2012.724538

Ardito, R. B., & Rabellino, D. (2011). Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. *Frontiers in Psychology*, *2*, Article 270. https://doi.org/10.3389/fpsyg.2011.00270

Barlow, D. H., Ellard, K. K., Sauer-Zavala, S., Bullis, J. R., & Carl, J. R. (2014). The origins of neuroticism. *Perspectives on Psychological Science*, *9*(5), 481–496. https://doi.org/10.1177/1745691614544528

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Fehr, E., . . . Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z

Blasini, M., Peiris, N., Wright, T., & Colloca, L. (2018). The role of patient–practitioner relationships in placebo and nocebo phenomena. *International Review of Neurobiology*, *139*, 211–231. https://doi.org/10.1016/bs.irn.2018.07.033

Castonguay, L. G., & Hill, C. E. (Eds.). (2017). *How and why are some therapists better than others? Understanding therapist effects.* American Psychological Association. https://doi.org/10.1037/0000034-000

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385–396. https://doi.org/10.2307/2136404

Constantino, M. J. (2012). Believing is seeing: An evolving research program on patients' psychotherapy expectations. *Psychotherapy Research*, *22*(2), 127–138. https://doi.org/10.1080/10503307.2012.663512

Constantino, M. J., Arnkoff, D. B., Glass, C. R., Ametrano, R. M., & Smith, J. Z. (2011). Expectations. *Journal of Clinical Psychology*, *67*(2), 184–192. https://doi.org/10.1002/jclp.20754

Constantino, M. J., Coyne, A. E., Goodwin, B. J., Vîslă, A., Flückiger, C., Muir, H. J., & Gaines, A. N. (2020). Indirect effect of patient outcome expectation on improvement through alliance quality: A meta-analysis. *Psychotherapy Research*, *31*(6), 1–15. https://doi.org/10.1080/10503307.2020.1851058

Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., Luborsky, L., McLellan, A., Woody, G., Thompson, L., Gallagher, D., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, *1*(2), 81–91. https://doi.org/10.1080/10503309112331335511

Czerniak, E., Biegon, A., Ziv, A., Karnieli-Miller, O., Weiser, M., Alon, U., & Citron, A. (2016). Manipulating the placebo response in experimental pain by altering doctor's performance style. *Frontiers in Psychology*, *7*, Article 874. https://doi.org/10.3389/fpsyg.2016.00874

Derksen, F., Bensing, J., & Lagro-Janssen, A. (2013). Effectiveness of empathy in general practice: A systematic review. *British Journal of General Practice*, *63*(606), 76–84. https://doi.org/10.3399/bjgp13X660814

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, *31*(2), 73–86. https://doi.org/10.1016/S0005-7916(00)00012-4

Dew, S. E., & Bickman, L. (2005). Client expectancies about therapy. *Mental Health Services Research*, *7*(1), 21–33. https://doi.org/10.1007/s11020-005-1963-5

Di Blasi, Z., Harkness, E., Ernst, E., Georgiou, A., & Kleijnen, J. (2001). Influence of context effects on health outcomes: A systematic review. *The Lancet*, *357*(9258), 757–762. https://doi.org/10.1016/S0140-6736(00)04169-6

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*(3), 193–242. https://doi.org/10.1037/h0044139

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy*, *55*(4), 399–410. https://doi.org/10.1037/pst0000175

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

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Flückiger, C., & Grosse Holtforth, M. (2008). Focusing the therapist's attention on the patient's strengths: A preliminary study to foster a mechanism of change in outpatient psychotherapy. *Journal of Clinical Psychology*, *64*(7), 876–890. https://doi.org/10.1002/jclp.20493

Geers, A. L., Briñol, P., Vogel, E. A., Aspiras, O., Caplandies, F. C., & Petty, R. E. (2018). The application of persuasion theory to placebo effects. *International Review of Neurobiology*, *138*, 113–136. https://doi.org/10.1016/bs.irn.2018.01.004

Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)" [Screening for mental disorders using the "Patient Health Questionnaire (PHQ-D)."] *Diagnostica*, *50*(4), 171–181. https://doi.org/10.1026/0012-1924.50.4.171

Greenberg, R. P. (1969). Effects of presession information on perception of the therapist and receptivity to influence in a psychotherapy analogue. *Journal of Consulting and Clinical Psychology*, *33*(4), 425–429. https://doi.org/10.1037/h0027816

Greenberg, R. P., Constantino, M. J., & Bruce, N. (2006). Are patient expectations still relevant for psychotherapy process and outcome? *Clinical Psychology Review*, *26*(6), 657–678. https://doi.org/10.1016/j.cpr.2005.03.002

Grencavage, L. M., & Norcross, J. C. (1990). Where are the commonalities among the therapeutic common factors? *Professional Psychology: Research and Practice*, *21*(5), 372–378. https://doi.org/10.1037/0735-7028.21.5.372

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*(7), 488–501. https://doi.org/10.1038/nrn3524

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, *36*(2), 223–233. https://doi.org/10.1037/0022-0167.36.2.223

Howe, L. C., Goyer, J. P., & Crum, A. J. (2017). Harnessing the placebo effect: Exploring the influence of physician characteristics on placebo response. *Health Psychology*, *36*(11), 1–9. https://doi.org/10.1037/hea0000499

Howe, L. C., Leibowitz, K. A., & Crum, A. J. (2019). When your doctor "Gets It" and "Gets You": The critical role of competence and warmth in the patient-provider interaction. *Frontiers in Psychiatry*, *10*, Article 475. https://doi.org/10.3389/fpsyt.2019.00475

Howick, J., Moscrop, A., Mebius, A., Fanshawe, T. R., Lewith, G., Bishop, F. L., Mistiaen, P., Roberts, N. W., Dieninytė, E., Hu, X. Y., Aveyard, P., & Onakpoya, I. J. (2018). Effects of empathic and positive communication in healthcare consultations: A systematic review and meta-analysis. *Journal of the Royal Society of Medicine*, *111*(7), 240–252. https://doi.org/10.1177/0141076818769477

Jeffreys, H. (1961). *The theory of probability*. Oxford University Press.

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review*, *67*, 78–93. https://doi.org/10.1016/j.cpr.2018.08.004

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*(6), 899–913. https://doi.org/10.1037/0022-3514.89.6.899

Kaptchuk, T. J., Kelley, J. M., Conboy, L. A., Davis, R. B., Kerr, C. E., Jacobson, E. E., Kirsch, I., Schyner, R. N., Nam, B. H., Nguyen, L. T., Park, M., Rivers, A. L., McManus, C., Kokkotou, E., Drossman, D. A., Goldman, P., & Lembo, A. J. (2008). Components of placebo effect: Randomised controlled trial in patients with irritable bowel syndrome. *BMJ*, *336*(7651), 999–1003. https://doi.org/10.1136/bmj.39524.439618.25

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.2307/2291091

Kazdin, A. E., & Krouse, R. (1983). The impact of variations in treatment rationales on expectancies for therapeutic change. *Behavior Therapy*, *14*(5), 657–671. https://doi.org/10.1016/S0005-7894(83)80058-6

Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J., & Riess, H. (2014). The influence of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomized controlled trials. *PLOS ONE*, *9*(4), Article e94207. https://doi.org/10.1371/journal.pone.0094207

Kirn, T., Echelmeyer, L., & Engberding, M. (2015). *Imagination in der Verhaltenstherapie* [Imagination in behavioral therapy]. Springer.

Koch, S. C., Mergheim, K., Raeke, J., Machado, C. B., Riegner, E., Nolden, J., Diermayr, G., von Moreau, D., & Hillecke, T. K. (2016). The embodied self in Parkinson's disease: Feasibility of a single tango intervention for assessing changes in psychological health outcomes and aesthetic experience. *Frontiers in Neuroscience*, *10*, Article 287. https://doi.org/10.3389/fnins.2016.00287

Kraft-Todd, G. T., Reinero, D. A., Kelley, J. M., Heberlein, A. S., Baer, L., & Riess, H. (2017). Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PLOS ONE*, *12*(5), Article e0177758. https://doi.org/10.1371/journal.pone.0177758

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Kube, T., Rief, W., Gollwitzer, M., Gärtner, T., & Glombiewski, J. A. (2019). Why dysfunctional expectations in depression persist—Results from two experimental studies investigating cognitive immunization. *Psychological Medicine*, *49*(9), 1532–1544. https://doi.org/10.1017/S0033291718002106

Ladwig, I., Rief, W., & Nestoriuc, Y. (2014). Welche Risiken und Nebenwirkungen hat Psychotherapie? Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP) [What are the risks and side effects of psychotherapy? Development of the Inventory for Recording Negative Effects of Psychotherapy (INEP)]. *Verhaltenstherapie*, *24*(4), 252–263. https://doi.org/10.1159/000367928

Leiner, D. J. (2019). *SoSci survey* (Version 3.2.32) [Computer software]. SoSci Survey GmbH. https://www.soscisurvey.de

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, *46*(3), 266–274. https://doi.org/10.1097/mlr.0b013e318160d093

Moors, F., & Zech, E. (2017). The effects of psychotherapist's and clients' interpersonal behaviors during a first simulated session: A lab study investigating client satisfaction. *Frontiers in Psychology*, *8*, Article 1868. https://doi.org/10.3389/fpsyg.2017.01868

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs* (Version 0.9.12-4.2) [Computer software]. Comprehensive R Archive Network. https://CRAN.R-project.org/package=BayesFactor

Newman, M. G., & Fisher, A. J. (2010). Expectancy/credibility change as a mediator of cognitive behavioral therapy for generalized anxiety disorder: Mechanism of action or proxy for symptom change? *International Journal of Cognitive Therapy*, *3*(3), 245–261. https://doi.org/10.1521/ijct.2010.3.3.245

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. https://doi.org/10.1037/h0043424

Petty, R. E., & Briñol, P. (2012). The elaboration likelihood model. In P. A. van Lange (Ed.), *Theories of social psychology* (pp. 224–245). SAGE. https://doi.org/10.4135/9781446249215.n12

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In R. E. Petty & J. T. Cacioppo (Eds.), *Communication and persuasion* (pp. 1–24). Springer. https://doi.org/10.1007/978-1-4612-4964-1_1

Rakel, D., Barrett, B., Zhang, Z., Hoeft, T., Chewning, B., Marchand, L., & Scheder, J. (2011). Perception of empathy in the therapeutic encounter: Effects on the common cold. *Patient Education and Counseling*, *85*(3), 390–397. https://doi.org/10.1016/j.pec.2011.01.009

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Raylu, N., & Kaur, I. (2012). Relationships between treatment expectations and treatment outcomes among outpatients with substance use problems. *International Journal of Mental Health and Addiction*, *10*(5), 607–621. https://doi.org/10.1007/s11469-011-9358-x

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.1.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rief, W., & Glombiewski, J. A. (2016). Expectation-focused psychological interventions (EFPI). *Verhaltenstherapie*, *26*, 1–7. https://doi.org/10.1159/000442374

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

Rief, W., & Joormann, J. (2019). Revisiting the cognitive model of depression: The role of expectations. *Clinical Psychology in Europe*, *1*(1), 1–19. https://doi.org/10.32872/cpe.v1i1.32605

Rief, W., Nestoriuc, Y., Mueller, E., Hermann, C., Schmidt, K., & Bingel, U. (2021). *Generic rating scale for previous treatment experiences, treatment expectations, and treatment effects (GEEE)*. PsychArchives. https://doi.org/10.23668/psycharchives.4717

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rumpold, G., Doering, S., Smrekar, U., Schubert, C., Koza, R., Schatz, D. S., Bertl-Schuessler, A., Janecke, N., Lampe, A., & Schuessler, G. (2005). Changes in motivation and the therapeutic alliance during a pretherapy diagnostic and motivation-enhancing phase among psychotherapy outpatients. *Psychotherapy Research*, *15*(1–2), 117–127. https://doi.org/10.1080/10503300512331327092

Rychlak, J. F., & Lerner, J. J. (1965). An expectancy interpretation of manifest anxiety. *Journal of Personality and Social Psychology*, *2*(5), 677–684. https://doi.org/10.1037/h0022627

Schneider, E. E., Schönfelder, S., Domke-Wolf, M., & Wessa, M. (2020). Measuring stress in clinical and nonclinical subjects using a German adaptation of the Perceived Stress Scale. *International Journal of Clinical and Health Psychology*, *20*(2), 173–181. https://doi.org/10.1016/j.ijchp.2020.03.004

Schneider, W., & Klauer, T. (2001). Symptom level, treatment motivation, and the effects of inpatient psychotherapy. *Psychotherapy Research*, *11*(2), 153–167. https://doi.org/10.1080/713663888

Schulz von Thun, F., Ruppel, J., & Stratmann, R. (2014). *Miteinander reden. Kommunikationspsychologie für Führungskräfte* [Talk to each other: Communication psychology for executives]. Rowohlt.

Silverman, A. L., Werntz, A., Ko, T. M., & Teachman, B. A. (2021). Implicit and explicit beliefs about the effectiveness of psychotherapy vs. medication: A large-scale examination and replication. *The Journal of Nervous and Mental Disease*, *209*(11), 783–795. https://doi.org/10.1097/NMD.0000000000001384

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Techniker Krankenkasse. (2016). *Entspann dich, Deutschland. TK-Stressstudie 2016* [Relax, Germany. TK Stress Study 2016]. https://www.tk.de/resource/blob/2026630/9154e4c71766c410dc859916aa798217/tk-stressstudie-2016-data.pdf

Ten Have, M., de Graaf, R., Vilagut, G., Kovess, V., & Alonso, J. (2010). Are attitudes towards mental health help-seeking associated with service use? Results from the European study of epidemiology of mental disorders. *Social Psychiatry and Psychiatric Epidemiology*, *45*(2), 153–163. https://doi.org/10.1007/s00127-009-0050-4

Tse, M. M. Y., Vong, S. K. S., & Tang, S. K. (2013). Motivational interviewing and exercise programme for community-dwelling older persons with chronic pain: A randomised controlled study. *Journal of Clinical Nursing*, *22*(13–14), 1843–1856. https://doi.org/10.1111/j.1365-2702.2012.04317.x

Vîslă, A., Flückiger, C., Constantino, M. J., Krieger, T., & Grosse Holtforth, M. (2019). Patient characteristics and the therapist as predictors of depressed patients' outcome expectation over time: A multilevel analysis. *Psychotherapy Research*, *29*(6), 709–722. https://doi.org/10.1080/10503307.2018.1428379

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238

Wampold, B. E. (2018). The therapeutic value of the relationship for placebo effects and other healing practices. *International Review of Neurobiology*, *139*, 191–210. https://doi.org/10.1016/bs.irn.2018.07.019

Westra, H. A., Aviram, A., Barnes, M., & Angus, L. (2010). Therapy was not what I expected: A preliminary qualitative analysis of concordance between client expectations and experience of cognitive–behavioural therapy. *Psychotherapy Research*, *20*(4), 436–446. https://doi.org/10.1080/10503301003657395

Westra, H. A., Constantino, M. J., Arkowitz, H., & Dozois, D. J. A. (2011). Therapist differences in cognitive–behavioral psychotherapy for generalized anxiety disorder: A pilot study. *Psychotherapy*, *48*(3), 283–292. https://doi.org/10.1037/a0022011

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Wilmers, F., Munder, T., Leonhart, R., Herzog, T., Plassmann, R., Barth, J., & Linster, H. W. (2008). Die deutschsprachige Version des Working Alliance Inventory-short revised (WAI-SR)—Ein schulenübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz [The German-language version of the Working Alliance Inventory-short revised (WAI-SR)—A cross-school, economic and empirically validated instrument for the measurement of therapeutic alliance]. *Klinische Diagnostik Und Evaluation*, *1*(3), 343–358. https://doi.org/10.7892/boris.27962

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10. https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf

Zion, S. R., & Crum, A. J. (2018). Mindsets matter: A new framework for harnessing the placebo effect in modern medicine. *International Review of Neurobiology*, *138*, 137–160. https://doi.org/10.1016/bs.irn.2018.02.002