Effects of cognitive load during interpretation bias modification on interpretation bias and stress reactivity

Van Bockstaele, B.; Clarke, P.J.F.; Notebaert, L.; MacLeod, C.; Salemink, E.

DOI
10.1016/j.jbtep.2020.101561

Publication date
2020

Document Version
Final published version

Published in
Journal of Behavior Therapy and Experimental Psychiatry

Citation for published version (APA):
Van Bockstaele, B., Clarke, P. J. F., Notebaert, L., MacLeod, C., & Salemink, E. (2020). Effects of cognitive load during interpretation bias modification on interpretation bias and stress reactivity. *Journal of Behavior Therapy and Experimental Psychiatry*, *68*, [101561]. https://doi.org/10.1016/j.jbtep.2020.101561

Journal of Behavior Therapy and Experimental Psychiatry 68 (2020) 101561

Contents lists available at ScienceDirect

Journal of Behavior Therapy and Experimental Psychiatry

journal homepage: www.elsevier.com/locate/jbtep

Effects of cognitive load during interpretation bias modification on interpretation bias and stress reactivity

Bram Van Bockstaele[a,c,*], Patrick J.F. Clarke[b], Lies Notebaert[c], Colin MacLeod[c], Elske Salemink[d,a]

[a] Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT, Amsterdam, the Netherlands
[b] School of Psychology, Curtin University, Kent St, WA 6102, Bentley, Australia
[c] School of Psychological Science, University of Western Australia, 35 Stirling Highway, WA 6009 Crawley, Australia
[d] Department of Clinical Psychology, Utrecht University, P.O. Box 80140, 3508 TC, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Interpretation bias
Cognitive bias modification
Stress reactivity
Secondary load

ABSTRACT

Background and objectives: Interpretation bias modification can affect stress reactivity, yet results have not been consistent. This inconsistency may be partly due to variability in the degree to which training procedures alter interpretation at a more automatic, rather than strategic, level of processing, and a mismatch in available resources between the training and the stress situation. We tested this possibility by investigating whether imposing a secondary cognitive load during interpretation bias modification would strengthen training-induced effects on both interpretation bias and emotional reactivity.

Method: We trained 71 participants in a single session to interpret ambiguity either positively or negatively. Half of our participants did so while performing a cognitively demanding secondary task. We assessed the effects of these different training regimes on interpretation bias and both self-reported and physiological indices of stress reactivity.

Results: Positive and negative interpretation bias modification resulted in training-congruent changes in interpretation bias. There were no group differences in self-reported stress reactivity, but positive interpretation training did improve recovery from stress as indexed by the heart rate measurement. Countering our hypothesis, the addition of cognitive load during the training increased neither the induced interpretive change nor its emotional impact.

Limitations: Sample size was relatively small, though sufficient to detect medium sized effects.

Conclusions: Adding cognitive load to interpretation bias modification does not alter training-induced change in interpretation bias or emotional reactivity.

1. Introduction

Cognitive models propose that anxiety is characterized by information processing biases (Williams, Watts, MacLeod, & Andrews, 1997). One of these biases is Interpretation Bias (IB): Compared to non-anxious individuals, anxious individuals are more prone to interpret emotionally ambiguous stimuli and situations as threatening. Crucially, IB may causally contribute to anxiety (Hirsch, Meeten, Krahé, & Reeder, 2016). This causality hypothesis is empirically supported by Interpretation Bias Modification (IBM) studies, in which experimentally induced changes in IB are related to changes in anxiety and stress reactivity. An often-used IBM paradigm is the modified scenario completion task (Mathews & Mackintosh, 2000), in which participants read emotionally ambiguous scenarios. One word of each scenario is presented as a word fragment which participants are required to complete. In positive training groups, the correct word fragment completion disambiguates each scenario in a positive manner. In possible comparison groups, completed word fragments disambiguate the scenarios either always negatively (negative training) or equally often positively or negatively (50/50 training), scenarios are emotionally neutral, or no training is delivered.

Crucially, IBM may affects not only people's interpretations of new ambiguous materials, but also their vulnerability to experience anxiety. While some studies addressed the effects of IBM on clinical anxiety, most IBM studies assess the effects of IBM on participants' reactivity to laboratory-based stressors as a proxy for anxiety responses in daily life. For instance, Tran, Siemer, and Joormann (2011) found that positive IBM resulted in a smaller decrease in self-esteem in response to a stress task than negative IBM. However, these effects are not found consistently. Salemink, van den Hout, and Kindt (2009, see also Salemink,

---

[*] Corresponding author. Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT, Amsterdam, the Netherlands
E-mail address: bram.vanbockstaele@uwa.edu.au (B. Van Bockstaele).

https://doi.org/10.1016/j.jbtep.2020.101561
Received 19 June 2019; Received in revised form 17 November 2019; Accepted 16 February 2020

van den Hout, & Kindt, 2007) found that positive IBM reduced immediate state anxiety but not stress reactivity compared with 50/50 training. Recent meta-analyses confirmed that IBM affects IB and anxiety in both adults (Menne-Lothmann et al., 2014) and youths (Krebs et al., 2018), yet the effects of IBM on stress reactivity were only significant in youths.

The inconsistent effects of IBM on stress reactivity may result from IBM not consistently inducing a change in IB that operates at a more automatic level of processing, especially with respect to the efficiency (Teachman, Joormann, Steinman, & Gotlib, 2012) and capacity-free (McNally, 1995) features of automaticity. Booth, Mackintosh, and Sharma (2017) found that the relationship between IB and trait anxiety was stronger when IB was assessed under additional cognitive load, that is, under conditions with limited resources available. Bowler et al. (2012) found that social anxiety symptom improvement following IBM was mediated by the degree of change in IB only when IB was measured under cognitive load, suggesting that the emotional impact of IBM is carried by training-induced change in the more efficient, automatically activated interpretations alone. Hence, maximizing the impact of IBM on more efficient, automatic components of IB may be crucial.

Furthermore, less cognitive resources are available in stressful situations (Ouimet, Gawronski, & Dozois, 2009). Based on the theoretical framework of transfer appropriate processing (Morris, Bransford, & Franks, 1977), stronger training effects are expected when the test context replicates the training context (Hertel & Mathews, 2011). In many IBM studies, there may be a mismatch between the available cognitive resources during training and the stressful test situation. Adding a cognitive load during training may thus improve the match in available cognitive resources between the training and the stress test and may consequently facilitate transfer of training effects to the stressor. This is important as finding ways to establish robust far-transfer effects has been put forward as “a major goal for CBM-researchers” (p. 527, Hertel & Mathews, 2011). Imposing a cognitive load on participants during IBM may thus be a promising strategy to improve the match between training and test contexts, and may maximize the impact of IBM on the more efficient, automatic features of IB.

In the present study, we investigated whether adding cognitive load during IBM leads to greater changes in IB and more pronounced effects on emotional reactivity. Half of our participants was assigned to a single-session positive IBM group, while the other half was assigned to a single-session negative IBM group. Participants were trained either with or without added cognitive load. After the training, we assessed IB both with and without load, and we exposed participants to a stress task. In addition to subjective, self-reported indices of emotional reactivity, we also measured heart rate (HR) and heart rate variability (HRV) during the stress task, thus presenting a more comprehensive picture of the effects of IBM on emotional reactivity with measures that are less susceptible to demand effects. HR typically increases while HRV typically decreases in response to stress-inducing tasks (Kreibig, 2010). If adding cognitive load induces change in more automatic IB, then through state-dependent learning the increased match between training and assessment contexts will lead to training-induced changes in IB being disproportionately evident when IB is also assessed under secondary load. In addition, if increased changes in more automatic IB and the improved match between training and stressor contexts strengthen the effects of IBM on stress reactivity, then positive IBM with load is expected to reduce stress reactivity relative to both positive IBM without load and negative IBM.

2. Method

2.1. Participants

Participants were mainly students of the University of Amsterdam who scored within the middle six deciles (i.e., scores between 31 and 48) of the trait version of the State and Trait Anxiety Inventory (STAI: van der Ploeg, Defares, & Spielberger, 1980) during a screening at the start of the year. This was done to avoid testing participants who had a strong bias in either direction before the training and to avoid exposing high-anxious individuals to the negative training. Walk-in participants who met the same inclusion criterion were also allowed to participate. A total of 74 participants started the experiment. The data of 3 participants were (partially) lost due to computer crashes and were not included in the analyses.

2.2. Design

We used a 2 × 2 between-subjects design. Participants were randomly allocated to one of the possible combinations of training valence (positive/negative) and cognitive load (with/without load). Our main dependent variables were IB, derived from reaction times (RTs) during the scenario completion training and from similarity ratings in the recognition task, and stress reactivity, measured using both negative mood ratings and physiological indices (HR and HRV).

2.3. Questionnaires

To assess baseline anxiety we used the Dutch version of the STAI. Both the state (STAI-S) and the trait version (STAI-T) consist of 20 items scored on 4-point Likert scales. Cronbach's alphas were .91 and .90, respectively.

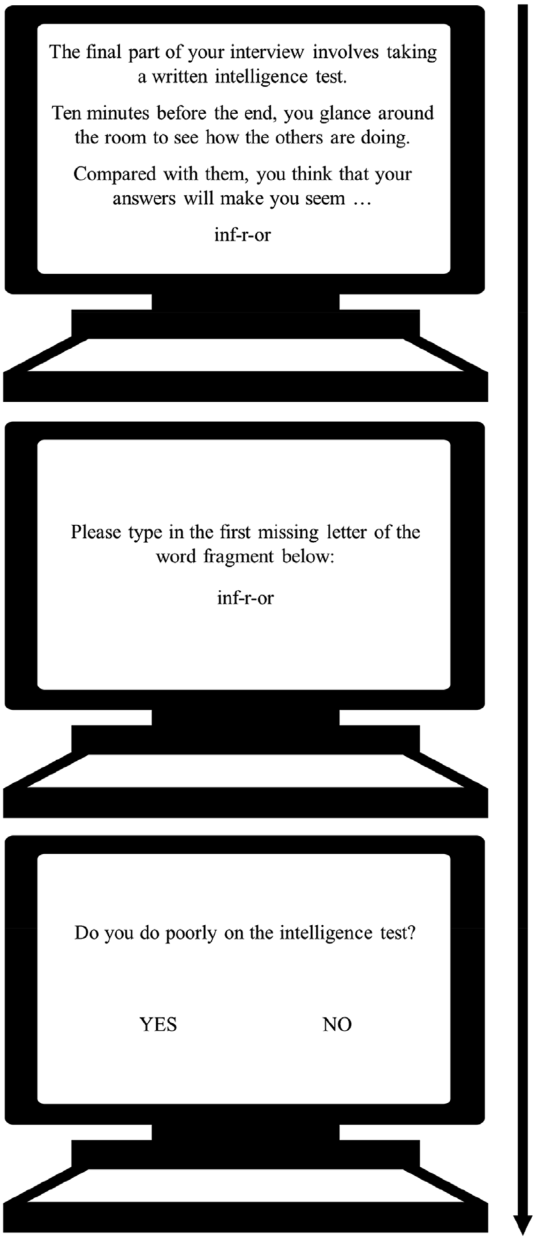
Three separate Visual Analogue Scales (VAS) were used to track mood throughout the experiment, respectively assessing self-confidence (insecure-confident), stress (stressed-relaxed), and mood (negative-positive). The VAS were completed 5 times: At the very start (Baseline), after the scenario completion training (Post-Training), before the stressor (Pre-Stressor), after the stressor (Post-Stressor), and after the stressor recovery (Post-Recovery). Each set of three VAS was combined to yield a single negative mood score ranging from 0 (positive mood) to 100 (negative mood) (Cronbach's alphas on each measurement $> .78$).

2.4. Interpretation bias modification and assessment: scenario completion training

The scenario completion training consisted of 80 scenarios. To match the content of the scenarios to the nature of the socio-evaluative performance stressor, ambiguous scenarios covering test-related and social-evaluative situations were drawn from Mackintosh, Mathews, Eckien, and Hoppitt (2013) and Mathews and Mackintosh (2000). Fig. 1 illustrates the full trial sequence. Each scenario consisted of three lines, and participants pressed the space bar to make the next line appear. A crucial disambiguating word was missing from the last line and was presented as a word fragment. Participants were asked to imagine themselves being in the situation and to press the space bar when they knew how to complete the word fragment. RTs were measured from the onset of the word fragment. Next, participants typed in the first missing letter, and they completed a yes/no comprehension question to verify that they had processed the scenario content and did not simply focus on completing the word fragment.

Prior to the training phase, participants completed 3 neutral practice scenarios. Participants were then randomly assigned to either positive or negative training, either with or without cognitive load. In the positive training groups, the correct word fragment solutions of 64 training scenarios always disambiguated the scenario positively, while in the negative training groups, the correct word fragments solutions of these 64 training scenarios always disambiguated the scenario negatively.

We also included 16 assessment scenarios, with 8 word fragment completions yielding a positive outcome and 8 word fragment completions yielding a negative outcome. These scenarios were the same for all participants. RT differences between positive and negative assessment scenarios were used as an index of IB. Each 10-trial sequence



The final part of your interview involves taking a written intelligence test.

Ten minutes before the end, you glance around the room to see how the others are doing.

Compared with them, you think that your answers will make you seem …

inf-r-or

Please type in the first missing letter of the word fragment below:

inf-r-or

Do you do poorly on the intelligence test?

YES NO

Fig. 1. Schematic overview of the trial sequence in the scenario completion training task. In this negative example, the correct solution is "inferior", while in the positive training groups, the word-fragment would be "cl-v-r", with the correct solution "clever".

comprised 8 training scenarios, 1 positive assessment scenario, and 1 negative assessment scenario. Scenarios were presented in 4 blocks of 20 trials each and participants took self-paced breaks between blocks.

Participants in the no-load groups completed the scenarios as described above, while participants in the load groups were shown a randomly generated 6-digit number for 4 s before every sequence of 4 trials. They were asked to hold these numbers in memory while completing the scenarios, and after every 4 scenarios, they indicated whether a single digit was part of the number that they had memorized. This procedure has been used in previous studies (e.g. Booth et al., 2017; Bowler et al., 2012; Clarke et al., 2017) and has been argued to deprive participants of processing resources (Gilbert & Hixon, 1991). Next, a new 6-digit number was presented, and the following sequence of 4 scenarios started.

2.5. *Interpretation bias assessment: recognition task*

The recognition task consisted of a block with and a block without cognitive load, presented in counterbalanced order. Each block consisted of two phases. First, 8 novel ambiguous test-related scenarios and comprehension questions similar to the ones used in the training were presented in random order, but completing the word fragment did not resolve the scenario's ambiguity. The scenarios also had a title and they were printed in yellow on a black background. Next, each of the titles of the preceding scenarios was shown, together with one positive and one negative interpretation of the scenario, each on a separate page. The positive and negative interpretations were presented in a fixed randomized order. Participants rated on 4-point Likert scales how (dis-)similar to the original scenario each interpretation was (1 = very different; 4 = very similar). Higher similarity ratings for negative compared to positive interpretations indicate a more negative IB. In the block without cognitive load, the task was presented as described above. In the block with cognitive load, the same load induction was used as the one described in the scenario completion training. This cognitive load manipulation was maintained during the similarity rating phase, with participants memorizing and responding to new 6-digit numbers before and after every 4 scenario interpretations.

2.6. *Anagram stress task*

Prior to the stress task, which was based on the task used by MacLeod, Rutherford, Campbell, Ebsworthy, and Holker (2002), the experimenter attached the HR devices. Inter-beat intervals (IBIs: larger IBIs reflect slower HR) were measured using a Suunto HR belt connected to a Windows 7 PC with a Suunto PC-POD and custom made software. The stress task was preceded by a 5-min rest phase, during which baseline HR was measured, followed by a VAS mood assessment. Next, participants were asked to solve 15 anagrams. They were told that success on this task predicted future success in a variety of domains, that students are typically good at the task, and that they were being video recorded while solving the anagrams. Anagrams were presented in a fixed order and 10 s after the onset of each anagram, a 10 s countdown clock was shown. The anagrams varied in difficulty so that participants would be able to solve some but not all of them. Participants were required to type in the complete solution, and the next anagram was automatically presented if participants did not respond within 20 s. HR was measured for as long as participants were solving the anagrams (max. 5 min). After the anagram phase, participants again completed the VAS, followed by a 5-min recovery period during which they relaxed and read a magazine while we measured HR. After the recovery period, they completed the VAS for a last time.

2.7. *Procedure*

The experiment was performed individually in a soundproof cubicle. Participants were informed of the general nature of the tasks and that they could terminate their participation at any time prior to providing written informed consent. Participants started by completing the questionnaires, followed by the scenario completion training, the recognition task, and the anagram stress task. After completing the experiment, participants were debriefed and rewarded with course credits or money. The entire experiment lasted for about 1 h and was approved by the ethical committee of the University of Amsterdam (ref. number 2015-DP-4093).

2.8. *Statistical approach*

We used mixed-measures ANOVAs to address our main hypotheses, always with the between-subjects factors Training Valence (positive/negative) and Training Load (with/without load). The ANOVAs had different within-subjects factors depending on the outcome measures:

The analysis of the scenario completion RTs included Scenario Valence (positive/negative), the analysis of the recognition task similarity ratings included Assessment Load (with/without load) and Target Sentence Valence (positive/negative), and the stress reactivity analyses included the factor Time (pre-stressor/post-stressor/post-recovery). We used the conventional p-value of .05 as a threshold for statistical significance.

3. Results

3.1. Data reduction and outlier analysis

For the scenario completion task, we removed trials with errors on word completions (1.25%), trials with RTs of 0 (0.39%), and trials with RTs deviating more than 3*SD*s from the group mean (2.02%) or each individual's mean (1.81%). Next, we calculated mean RTs on positive and negative assessment trials. For the recognition task, we calculated mean similarity ratings for positive and negative interpretations for the load and no-load blocks separately. With respect to the number memorization, all participants in the load groups performed well above chance during the scenario completion training (M = 87.03% correct, SD = 9.46, range = 65–100%), while 11 participants made errors on at least half of the number memorization trials in the load block of the recognition task. In the anagram task, there were no significant group differences in either the percentage of correctly solved anagrams (overall M = 38.12%, SD = 14.31) or the mean RT for solving the anagrams (overall M = 13690 ms, SD = 2361), all Fs $<$ 1, all ps $>$ .48.

For each phase of the stress test, we calculated the total summed IBIs to check the continuity of the HR signal. We retained the HR data of only those participants for whom a maximum of 30 s (i.e., 10% of the total duration of the baseline or recovery phase) in each phase was missing. The most likely cause for these missing data was a loose HR belt. The HR data of 20 participants were considered (partly) compromised and were not further analysed. For the remaining 51 participants, we used ARTiiFACT (Kaufmann, Sütterlin, Schulz, & Vögele, 2011) to automatically detect artefacts (i.e., excessively long or short IBIs) following the recommendations of Berntson, Quigley, Jang, and Boysen (1990). Artefacts were visually inspected and overruled if deemed necessary (while being blind to participants' group allocations). Remaining artefacts were corrected using cubic spline interpolation, after which we calculated mean IBIs and HRV (RMSSD: root mean square of successive differences in IBIs) for each phase.

3.2. Baseline group characteristics

Our final sample consisted of 71 participants (50 women, *M*age = 21.73, SD = 6.63).[1] Groups did not differ significantly on any of the baseline anxiety measures, all Fs $<$ 1 (Table 1).

3.3. Effects of interpretation bias modification on interpretation bias

Assessing the effects of IBM on IB during the training, the 2 (Training Valence) x 2 (Training Load) x 2 (Scenario Valence) mixed-design ANOVA on the assessment scenario RTs yielded no significant 3-way interaction, $F(1, 67) = 1.17$, $p = .283$. There was a significant Scenario Valence by Training Valence interaction, $F(1, 67) = 9.83$, $p = .003$, $f = 0.38$ (Fig. 2), qualifying the main effects of Scenario Valence, $F(1, 67) = 6.32$, $p = .014$, $f = 0.31$, and Training Valence, $F(1, 67) = 8.17$, $p = .006$, $d = 0.68$. Follow-up within-group comparisons showed that participants in the positive training groups were significantly faster to complete positive than negative word fragments, $F(1, 34) = 23.07$, $p < .001$, $f = 0.82$, with no such difference in the negative training groups, $F < 1$. No other effects in the ANOVA were significant, all Fs $<$ 1. The positive training groups thus had a positive IB during the training, while the negative training groups had no bias. Cognitive load did not affect the acquisition of IB.

Post-training differences in IB were addressed in a 2 (Training Valence) x 2 (Training Load) x 2 (Assessment Load) x 2 (Target Sentence Valence) mixed-design ANOVA on the recognition task similarity ratings. The 4-way interaction was not significant, $F < 1$, nor were any of the 3-way interactions, all Fs $<$ 3.55, all ps $>$ .06. The interaction between Target Sentence Valence and Assessment Load was significant, $F(1, 67) = 5.03$, $p = .028$, $f = 0.27$. This interaction is not of theoretical relevance, reflecting the absence of a main effect of Target Sentence Valence in the block with load, $F(1, 70) = 2.41$, $p = .13$, but in the block without cognitive load participants rated positive target sentences ($M = 2.49$, $SD = 0.55$) as more similar to the original scenario than negative target sentences ($M = 2.09$, $SD = 0.56$), $F(1, 70) = 13.06$, $p = .001$, $f = 0.43$. More importantly, the Target Sentence Valence by Training Valence interaction was highly significant, $F(1, 67) = 133.45$, $p < .001$, $f = 1.41$ (Fig. 3), qualifying the main effects of Target Sentence Valence, $F(1, 67) = 25.18$, $p < .001$, $f = 0.61$, and Training Valence, $F(1, 67) = 5.78$, $p = .019$, $d = 0.58$. Following-up on this interaction, participants in the negative training groups rated negative target sentences as more similar to the scenarios than participants in the positive training groups did, $t(69) = 9.85$, $p < .001$, $d = 2.34$. Inversely, participants in the positive training groups rated positive target sentences as more similar to the scenarios than participants in the negative training groups did, $t(69) = 6.44$, $p < .001$, $d = 1.53$. Within-group comparisons further illustrated that participants in the negative training groups rated negative target sentences as more similar to the scenarios than positive target sentences, $F(1, 35) = 18.90$, $p < .001$, $f = 0.74$, while participants in the positive training groups rated positive target sentences as more similar to the scenarios than the negative target sentences, $F(1, 34) = 141.28$, $p < .001$, $f = 2.04$. We thus found strong training-congruent differences in IB, but the addition of cognitive load during either the training or the assessment did not affect these differences.

3.4. Effects of interpretation bias modification on self-reported mood and stress reactivity

Addressing the effects of different training manipulations on mood, a 2 (Time: baseline versus post-training) x 2 (Training Load) x 2 (Training Valence) mixed-design ANOVA (Table 2) revealed no significant 3-way interaction, $F < 1$, nor any significant 2-way interactions, all Fs $<$ 3.64, all ps $>$ .06, indicating that the training manipulations did not differentially affect mood. The main effect of Time was significant, $F(1, 67) = 5.07$, $p = .028$, $f = 0.27$, indicating that negative mood increased from baseline ($M = 28.39$, $SD = 14.92$) to post-training ($M = 31.42$, $SD = 13.00$). Both other main effects were not significant, both Fs $<$ 3.14, both ps $>$ .08.

Determining the effects of training on self-reported stress reactivity, the 3 (Time: pre-stressor, post-stressor, post-recovery) x 2 (Training Load) x 2 (Training Valence) mixed-design ANOVA (Table 2) revealed neither a significant 3-way interaction, $F(2, 66) = 1.82$, $p = .170$, nor any significant 2-way interactions, all Fs $<$ 1.02, all ps $>$ .31. Only the main effect of Time was significant, $F(2, 66) = 52.76$, $p < .001$, $f = 1.26$, reflecting the intended elevation of negative mood from pre-stressor ($M = 30.28$, $SD = 13.21$) to post-stressor ($M = 44.92$, $SD = 17.65$), $F(1, 70) = 74.08$, $p < .001$, $f = 1.03$, and a subsequent decline in negative mood from post-stressor to post-recovery ($M = 31.07$, $SD = 13.23$), $F(1, 70) = 98.27$, $p < .001$, $f = 1.18$. The other main effects were not significant, both Fs $<$ 2.37, both ps $>$ .12. So although the stressor and the recovery period had the expected effects on negative mood, this emotional reactivity was unaffected by both the IBM training and the load manipulation.

---

[1] Due to an experimenter error, the age data of 7 participants were missing.

Table 1
Baseline group characteristics.

	Negative No Load		Negative Load		Positive No Load		Positive Load	
N (Female/Male)	18	(13/5)	18	(10/8)	16	(11/5)	19	(16/3)
Age (*SD*)	20.73	(2.60)	20.25	(2.27)	26.13	(12.38)	20.22	(1.99)
STAI-T (*SD*)	36.67	(7.81)	37.44	(8.20)	39.50	(9.84)	37.47	(8.49)
STAI-S (*SD*)	34.78	(8.95)	32.39	(4.47)	34.44	(9.97)	32.00	(5.59)
Baseline negative mood (*SD*)	32.13	(19.00)	27.52	(11.88)	17.52	(13.55)	26.40	(14.77)

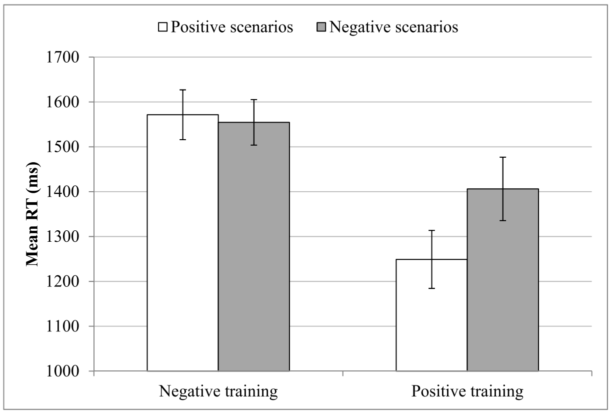


Fig. 2. Mean positive and negative assessment scenario reaction times in the scenario completion training as a function of training valence. Error bars reflect standard errors of the mean.

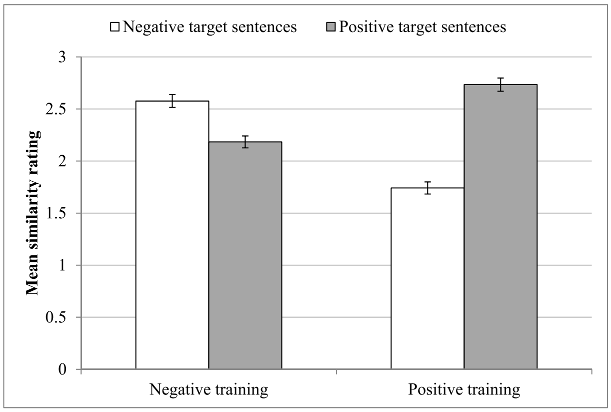


Fig. 3. Mean similarity ratings of positive and negative target sentences in the recognition task as a function of training valence. Error bars reflect standard errors of the mean.

Table 2
Average negative mood scores throughout the experiment.

	Baseline		Post-training		Pre-stressor		Post-stressor		Post-recovery	
Training Group	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)
Negative – No load	32.13	(19.00)	34.78	(15.34)	29.89	(13.24)	42.20	(18.28)	32.52	(13.89)
Negative – Load	27.52	(11.88)	35.81	(14.48)	35.59	(16.11)	51.65	(18.94)	34.67	(14.89)
Positive – No load	27.52	(13.55)	27.17	(10.89)	27.94	(10.50)	43.90	(17.99)	27.92	(11.05)
Positive – Load	26.40	(14.77)	27.67	(8.73)	27.60	(11.64)	42.00	(15.00)	28.93	(12.60)

3.5. *Effects of interpretation bias modification on physiological stress reactivity*

The effects of IBM on HR and HRV exhibited during and after the anagram stressor, shown in Table 3, were determined by two separate 3 (Time: baseline, stressor, recovery) x 2 (Training Load) x 2 (Training Valence) mixed-design ANOVAs. The HR analysis yielded no significant 3-way interaction, $F < 1$. Of all other effects, only the Time by Training Valence interaction was significant, $F(2, 46) = 4.11$, $p = .023$, $f = 0.42$, both other interactions *Fs* < 1.34, *ps* > .25, all main effects *Fs* < 3.01, all *ps* > .05. Following-up on the interaction, comparing the positive and negative training group during only the baseline and stress phases revealed a significant main effect of Time, $F(1, 49) = 6.10$, $p = .017$, $f = 0.35$, reflecting increased HR to the stressor that was unaffected by Training Valence, with the interaction $F < 1$. Contrasting the stress and recovery phases, the Time by Training Valence interaction was significant, $F(1, 49) = 5.81$, $p = .020$, $f = 0.34$. While participants in the negative training group showed no decreasing HR during the recovery phase, $F < 1$, participants in the positive training group had a significantly lower HR in the recovery phase than in the stress phase, $F(1, 24) = 8.08$, $p = .009$, $f = 0.58$, suggesting faster recovery from stress.

The HRV analysis yielded no significant 3-way interaction, $F(2, 46) = 1.08$, $p = .348$, nor any other crucial interactions involving the factor Time, both *Fs* < 1, indicating that neither IBM nor the addition of cognitive load affected stress reactivity as indexed by HRV. There was a significant Training Valence by Training Load interaction, $F(1, 47) = 5.72$, $p = .021$, indicating that average HRV (across all three phases) was smaller in the positive training group without load ($M = 42.92$, $SD = 15.84$) than the positive training group with load ($M = 70.16$, $SD = 33.00$), $F(1, 23) = 6.31$, $p = .019$, while there was no such difference between the negative training groups with ($M = 50.76$, $SD = 17.56$) versus without load ($M = 57.02$, $SD = 26.35$), $F < 1$. Finally, the main effect of Time was significant, $F(2, 46) = 3.76$, $p = .031$, $f = 0.41$, indicating that, as expected, HRV decreased from the baseline to the stress phase, $F(1, 50) = 6.98$, $p = .011$, and increased from the stress phase to the recovery phase, $F(1, 50) = 5.93$, $p = .019$.

4. Discussion

Our results are relatively easily summarized. IBM resulted in training-congruent differences in IB. Adding cognitive load during IBM did not affect the magnitude of the training-induced IB change, either with or without cognitive load during the measurement, nor did it alter the impact of IBM on mood or self-reported or physiological stress reactivity and recovery. IBM did not affect participants' self-reported emotional reactivity to a stressor. However, positive IBM improved the physiological recovery from stress as indicated by changes in HR (but

Table 3
Heart rate indices of stress reactivity.

	Baseline		Stress		Recovery	
Training Group	Mean IBI	(*SD*)	Mean IBI	(*SD*)	Mean IBI	(*SD*)
Negative – No load	873.75	(78.36)	867.87	(70.70)	859.83	(89.00)
Negative – Load	868.32	(120.69)	837.77	(106.91)	836.84	(109.68)
Negative – Total	871.24	(98.03)	853.98	(88.67)	849.22	(97.72)
Positive – No Load	855.80	(106.07)	844.28	(107.07)	862.83	(100.12)
Positive – Load	907.69	(141.57)	884.54	(128.10)	922.59	(145.71)
Positive – Total	877.92	(112.45)	866.83	(118.67)	896.30	(128.82)
	Mean HRV	(*SD*)	Mean HRV	(*SD*)	Mean HRV	(*SD*)
Negative – No load	58.65	(27.30)	53.69	(22.90)	58.74	(32.01)
Negative – Load	55.86	(21.64)	47.53	(16.71)	48.91	(17.89)
Negative – Total	57.36	(24.40)	50.84	(20.13)	54.20	(26.43)
Positive – No Load	42.62	(17.59)	42.06	(16.05)	44.07	(18.41)
Positive – Load	74.14	(35.62)	62.69	(32.48)	73.66	(38.52)
Positive – Total	60.27	(32.73)	53.61	(28.07)	60.64	(34.20)

Note: IBI = Inter-Beat Interval, HRV = Heart Rate Variability.

not HRV).

We found no support for the idea that adding cognitive load during IBM would strengthen the effects of IBM on both IB and emotional reactivity. The nature of our load manipulation could partly accounts for this. In dual task paradigms, stimulus encoding can be compromised by the presence of other stimuli belonging to the same domain (e.g. verbal, visual; see Cowan & Morey, 2007). Keeping a (verbal) number sequence active in working memory may have interfered with the processing of the semantic content of the scenarios, thus hampering the transfer of training under load. However, if our load manipulation hampered the semantic processing of the scenarios, this should have compromised the effects of training on IB, with reduced training effects in the load groups. As the addition of cognitive load did not compromise the effects of IBM on IB, the idea that our load manipulation interfered with semantic processing (thus hampering far transfer) seems unlikely. Nevertheless, this idea may be tested experimentally using a non-verbal secondary cognitive load task, like the random interval repetition task (Vandierendonck, De Vooght, & Van der Goten, 1998).

IBM (or the addition of cognitive load) did not affect self-reported stress reactivity, despite producing the intended changes in bias. The successful modification of bias is considered an important factor distinguishing between studies in which IBM did versus did not affect emotional vulnerability (Grafton et al., 2017): Only when IB is successfully modified should emotional vulnerability also be affected. Our self-report findings are not consistent with this pattern (see e.g. also Van Bockstaele, Notebaert et al., 2019), suggesting that while successfully changing bias may be necessary to successfully change emotional vulnerability, it is not always sufficient. However, this does not negate the findings of other studies in which experimentally induced changes in IB did lead to changes in self-reported stress reactivity (e.g. Tran et al., 2011; Wilson, MacLeod, Mathews, & Rutherford, 2006). Future research may identify potentially crucial moderating factors of the relation between changes in IB and changes in self-reported stress reactivity, like for instance strong imagery instructions during IBM (Holmes, Lang, & Shah, 2009).

Interestingly, positive IBM did improve recovery from stress as indexed by changes in HR. To our knowledge, only two previous studies addressed the effects of IBM on HR indices of stress reactivity. Nowakowski, Antony, and Koerner (2015) found that positive IBM reduced HR during recovery from stress, but they found a similar recovery-effect in a control group who had completed a 50/50 training. Investigating IBM in the context of depression, Joormann, Waugh, and Gotlib (2015) found that participants in a positive IBM group showed a smaller increase in HR in response a stressor than participants in a negative training group, but no group differences during recovery. In line with our current results, neither of the previous studies found any significant effects of IBM on self-reported stress reactivity. So although there are inconsistencies as to whether IBM affects HR during the stressor or during recovery, the finding that IBM affects HR rather than self-reported indices of stress reactivity is consistent.

The divergence between self-reported and physiological outcomes may seem surprising. While self-reports can provide valid and accurate indications of subjective feelings, most emotion theories postulate that subjective feelings are only one component of the emotional response, next to cognitions, physiological changes, and behaviours (Moors, 2009). These different emotion indices are typically only loosely related, and correlations between different measures are often small and inconsistent (Mauss & Robinson, 2009; Van Bockstaele et al., 2011). The poor convergence between self-reported and physiological outcomes may be due to participants being unwilling or incapable to accurately report their emotions. Alternatively, we measured HR continuously during each phase, while self-report ratings were provided after each phase. Our HR measures may thus reflect increasing anxiety during stress induction and decreasing anxiety during recovery, while the self-reports may reflect the end result of the stress induction and recovery phases. Finally, our HR and HRV results also diverged, despite the fact that changes over phases between both outcomes were strongly correlated (*rs* = .71). One tentative explanation for this divergence could be that HRV as indexed by RMSSD primarily reflects parasympathetic nervous system activity (Thayer, Åhs, Fredrikson, Sollers III, & Wager, 2012), while HR reflects a combination of sympathetic and parasympathetic activity (Mauss and Robinson, 2009). Effects of IBM on physiological stress reactivity may thus be limited to the sympathetic but not the parasympathetic nervous system. Future research including additional physiological measures that are primarily driven by the sympathetic nervous system, like skin conductance levels, may add empirical weight to this idea.

Our study also has limitations. First, our sample was relatively small, with between 16 and 19 participants per group, and even less for the physiological outcomes. A post-hoc power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) – given our sample size and a minimal statistical power of .80 – showed that our sample was still large enough to detect medium and for some tests even small within-between interactions, depending on the observed correlations between repeated measures. Second, our single-session training may have been too short for IBM to affect self-reported stress reactivity. However, Menne-Lothmann et al.'s (2014) meta-analysis does not support this idea, as they found that increasing the number of training sessions did not strengthen the effects of IBM on stress reactivity. Finally, it should be stressed that a substantial amount of HR data was excluded due to measurement errors, we did not find the improved physiological recovery in the HRV index, we found no group differences in HR from the baseline to the stress phase, and we did not correct our alpha-levels for multiple comparisons. Our results should therefore be interpreted with caution and further replication of the effects of IBM on physiological stress reactivity and recovery is warranted.

In sum, despite training-congruent effects on IB, adding cognitive load to IBM did not strengthen these effects. We found no effects of IBM on self-reported stress reactivity, but we did find evidence for improved physiological recovery from stress after positive IBM. Once again, the addition of cognitive load to IBM did not serve to enhance the emotional impact of the training.

Author contributions

ES, PC, and CM developed the broad study concept. All authors were involved in the development of the study design and hypotheses. The data were collected under supervision of ES. BVB transformed the raw data and analysed the data, under supervision of ES. All authors contributed to the interpretation of the results. BVB drafted the manuscript, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

Funding and acknowledgements

This study was supported by a University of Western Australia Research Collaboration Award awarded to Lies Notebaert, Patrick Clarke, Bram Van Bockstaele, Reinout Wiers, and Elske Salemink. Colin MacLeod is supported by Australian Research Council Laureate Fellowship FL170100167. The authors thank Sanne Buter and Vera Balk for their aid in the data collection.

Declaration of competing interest

All authors acknowledge that they have exercised due care in ensuring the integrity of the work. Further, none of the original material contained in the manuscript has been submitted for consideration nor will any of it be published elsewhere except in abstract form in connection with scientific meetings. We have no conflicts of interest to disclose.

Open practice and data availability statement

The raw data, outlier analysis description, transformed data, and the analysis output are available on the following OSF-page: https://osf.io/cjdsg/ (doi:10.17605/OSF.IO/CJDSG).

References

Berntson, G. G., Quigley, K. S., Jang, J. F., & Boysen, S. T. (1990). An approach to artifact identification: Application to heart period data. *Psychophysiology, 27*, 586–598. https://doi.org/10.1111/j.1469-8986.1990.tb01982.x.

Booth, R. W., Mackintosh, B., & Sharma, D. (2017). Working memory regulates trait anxiety-related threat processing biases. *Emotion, 17*, 616–627. https://doi.org/10.1037/emo0000264.

Bowler, J. O., Mackintosh, B., Dunn, B. D., Mathews, A., Dalgleish, T., & Hoppitt, L. (2012). A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: Effects on anxiety, depression, attentional control, and interpretive bias. *Journal of Consulting and Clinical Psychology, 80*, 1021–1033. https://doi.org/10.1037/a0029932.

Clarke, P. J. F., Branson, S., Chen, N. T. M., Van Bockstaele, B., Salemink, E., MacLeod, C., et al. (2017). Attention bias modification training under working memory load increases the magnitude of change in attentional bias. *Journal of Behavior Therapy and Experimental Psychiatry, 57*, 25–31. https://doi.org/10.1016/j.jbtep.2017.02.003.

Cowan, N., & Morey, C. C. (2007). How can dual-task working memory retention limits be investigated? *Psychological Science, 18*, 686–688. https://doi.org/10.1111/j.1467-9280.2007.01960.x.

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*, 175–191. https://doi.org/10.3758/BF03193146.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517. https://doi.org/10.1037/0022-3514.60.4.509.

Grafton, B., MacLeod, B., Rudaizky, D., Holmes, E. A., Salemink, E., Fox, E., et al. (2017). Confusing procedures with process when appraising the impact of cognitive bias modification on emotional vulnerability. *The British Journal of Psychiatry, 211*, 266–271. https://doi.org/10.1192/bjp.bp.115.176123.

Hertel, P. T., & Mathews, A. (2011). Cognitive bias modification: Past perspectives, current findings, and future applications. *Perspectives on Psychological Science, 6*, 521–536. https://doi.org/10.1177/1745691611421205.

Hirsch, C. R., Meeten, F., Krahé, C., & Reeder, C. (2016). Resolving ambiguity in emotional disorders: The nature and role of interpretation biases. *Annual Review of Clinical Psychology, 12*, 281–305. https://doi.org/10.1146/annurev-clinpsy-021815-093436.

Holmes, E. A., Lang, T. J., & Shah, D. M. (2009). Developing interpretation bias modification as a "cognitive vaccine" for depressed mood: Imagining positive events makes you feel better than thinking about them verbally. *Journal of Abnormal Psychology, 118*, 76–88. https://doi.org/10.1037/a0012590.

Joormann, J., Waugh, C. E., & Gotlib, I. H. (2015). Cognitive bias modification for interpretation in major depression: Effects on memory and stress reactivity. *Clinical Psychological Science, 3*, 126–139. https://doi.org/10.1177/2167702614560748.

Kaufmann, T., Sütterlin, S., Schulz, S., & Vögele, C. (2011). ARTiiFACT: A tool for heart rate artifact processing and heart rate variability analysis. *Behavior Research Methods, 43*, 1161–1170. https://doi.org/10.3758/s13428-011-0107-7.

Krebs, G., Pile, V., Grant, S., Degli Esposti, M., Montgomery, P., & Lau, J. Y. F. (2018). Research review: Cognitive bias modification of interpretations in youth and its effect on anxiety: A meta-analysis. *Journal of Child Psychology and Psychiatry, 59*, 831–844. https://doi.org/10.1111/jcpp.12809.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*, 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010.

Mackintosh, B., Mathews, A., Eckien, D., & Hoppitt, L. (2013). Specificity effects in the modification of interpretation bias and stress reactivity. *Journal of Experimental Psychopathology, 4*, 133–147. https://doi.org/10.5127/jep.025711.

MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology, 111*, 107–123. https://doi.org/10.1037/0021-843X.111.1.107.

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology, 109*, 602–615. https://doi.org/10.1037//0021-843X.109.4.602.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion, 23*, 209–237. https://doi.org/10.1080/02699930802204677.

McNally, R. J. (1995). Automaticity and the anxiety disorders. *Behaviour Research and Therapy, 33*, 747–754. https://doi.org/10.1016/0005-7967(95)00015-p.

Menne-Lothmann, C., Viechtbauer, W., Höhn, P., Kasanova, Z., Haller, S. P., Drukker, M., ... Lau, J. Y. F. (2014). How to boost positive interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *PloS One, 9*, e100925. https://doi.org/10.1371/journal.pone.0100925.

Moors, A. (2009). Theories of emotion causation: A review. *Cognition & Emotion, 23*, 625–662. https://doi.org/10.1080/02699930802645739.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9.

Nowakowski, M. E., Antony, M. M., & Koerner, N. (2015). Modifying interpretation biases: Effects on symptomatology, behavior, and physiological reactivity in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 49*, 44–52. https://doi.org/10.1016/j.jbtep.2015.04.004.

Ouimet, A. J., Gawronski, B., & Dozois, D. J. (2009). Cognitive vulnerability to anxiety: A review and an integrative model. *Clinical Psychology Review, 29*, 459–470. https://doi.org/10.1016/j.cpr.2009.05.004.

van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (1980). *Handleiding bij de Zelfbeoordelings vragenlijst. [Manual for the Dutch adaptation of the STAI-Y]*. Lisse: Swets & Zeitlinger.

Salemink, E., van den Hout, M., & Kindt, M. (2007). Trained interpretive bias and anxiety. *Behaviour Research and Therapy, 45*, 329–340. https://doi.org/10.1016/j.brat.2006.03.011.

Salemink, E., van den Hout, M. A., & Kindt, M. (2009). Effects of positive interpretive bias modification in highly anxious individuals. *Journal of Anxiety Disorders, 23*, 676–683. https://doi.org/10.1016/j.janxdis.2009.02.006.

Teachman, B. A., Joormann, J., Steinman, S. A., & Gotlib, I. H. (2012). Automaticity in anxiety disorders and major depressive disorder. *Clinical Psychology Review, 32*, 575–603. https://doi.org/10.1016/j.cpr.2012.06.004.

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J., III, & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews, 36*, 747–756. https://doi.org/10.1016/j.neubiorev.2011.11.009.

Tran, T. B., Siemer, M., & Joormann, J. (2011). Implicit interpretation biases affect emotional vulnerability: A training study. *Cognition & Emotion, 25*, 546–558. https://doi.org/10.1080/02699931.2010.532393.

Van Bockstaele, B., Notebaert, L., Salemink, E., Clarke, P. J. F., MacLeod, C., Wiers, R. W., et al. (2019). Effects of interpretation bias modification on unregulated and regulated emotional reactivity. *Journal of Behavior Therapy and Experimental Psychiatry, 64*, 123–132. https://doi.org/10.1016/j.jbtep.2019.03.009.

Van Bockstaele, B., Verschuere, B., Koster, E. H. W., Tibboel, H., De Houwer, J., & Crombez, G. (2011). Differential predictive power of self report and implicit measures on behavioural and physiological fear responses to spiders. *International Journal of Psychophysiology, 79*, 166–174. https://doi.org/10.1016/j.ijpsycho.2010.10.003.

Vandierendonck, A., De Vooght, G., & Van der Goten, K. (1998). Interfering with the central executive by means of the random interval repetition task. *The Quarterly Journal of Experimental Psychology Section A, 51*, 197–218. https://doi.org/10.1080/713755748.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1997). *Cognitive psychology and the emotional disorders* (2nd ed.). New York, NY: Wiley.

Wilson, E. J., MacLeod, C., Mathews, A., & Rutherford, E. M. (2006). The causal role of interpretative bias in anxiety reactivity. *Journal of Abnormal Psychology, 115*, 103–111. https://doi.org/10.1037/0021-843X.115.1.103.