Appendix S1

SUPPLEMENTARY MATERIALS

Normality and homogeneity of variance for test phase data

Itch intensity data from both the control and reinforced test trials were checked for normality by group using Shapiro–Wilks tests. Data from both evocation trials were normally distributed for classical conditioning (both $p>0.05$). Data from both trials for sham conditioning (control trial $p=0.01$, reinforced trial $p=0.031$) and observational learning (control trial $p<0.001$, reinforced trial $p=0.005$) were not normally distributed. Visual inspection of histograms of the data by group revealed this was due to a single participant in each group who rated her itch in both trials much higher than other participants, and an additional participant in the sham conditioning group who rated her itch on the reinforced trial much higher than other, but not for the control trial (Fig. 2). Homogeneity of error variance was assessed with Levene's test and was acceptable (all $p>0.05$). Results of the primary analysis comparing the nocebo effect between classical and sham conditioning did not change when the analysis was run with these participants excluded [$F(1,34)=4.34$, $p=0.045$, $\eta_g^2=0.01$]. The results of the secondary analysis comparing the nocebo effect between observational learning sham conditioning was also not changed when run with these participants excluded [$F(1,35)=2.31$, $p=0.137$, $\eta_g^2<0.01$]

Manipulation checks

At the end of the experiment, prior to debriefing, participants were asked 3 questions as manipulation checks: (*i*) "Please tell us what you think the purpose of the experiment was."; (*ii*) "Did you believe the information and instructions given to you by the experimenters?"; and (*iii*) "Did you notice an association between either of the gels and the intensity of the itch you experienced after receiving that gel?" Responses to the first item were coded as the participant (1) did not, (2) partially, or (3) accurately identified the aim of the experiment. Responses to the second and third items were coded as yes, partially/somewhat, or no. Results from these manipulation check items by group are shown in Tables SI and Table SII. None of these items correlated significantly with the magnitude of the nocebo effect in any group. Indicating a lack of complete belief in the instructions given throughout the experiment was negatively correlated with magnitude of the nocebo effect on itch at marginal significance in the classical ($r=-0.44$, $p=0.051$) indicating that participants who did not fully believe the instructions reported smaller nocebo effects than those who reported believing the instructions fully.

Participants in the observational learning group completed an additional manipulation check in the form of a post-video survey asking participants to identify the colour of the label on the bottle of gel used in each trial observed in the video, and to estimate how much itch the model participant experienced in each trial. Ten out of 22 participants (45%) misidentified the colour of the label on the bottle of the gel in at least one trial. Accuracy was better for the final 2 trials, in which 4 participants (18%) failed to indicate the correct label colour for 1 trial compared with the first 2 trials, in which 8 participants (36%) made errors. Accuracy of identifying the correct bottle used in each trial was measured on a 0–4 scale (number of correct answers), and was not found to correlate significantly with the magnitude of the nocebo effect ($r=-0.15$, $p=0.482$).

Learning phase results

In the learning phase, participants in the classical conditioning group were expected to report increased itch during the nocebo trials relative to the control trials (Fig. S1). A repeated measures ANOVA confirmed a main effect of trial type (control vs nocebo learning trial) [$F(1,18)=6.09$, $p=0.024$, $\eta_g^2=0.07$]. For the sham conditioning group, where all 4 learning trials were control trials, no differences between itch ratings on the learning trials was expected (Fig. S2). A repeated measures ANOVA detected no main effect of trial number (learning trials 1–4) on itch ratings [$F(1,17)=0.83$, $p=0.375$, $\eta_g^2=0.02$]. Participants in the observational learning group observed a video of the conditioning paradigm instead of experiencing it directly.

Psychological factors

This study tested the exploratory hypotheses that state anxiety, perceived stress, and negative affect would positively correlate with nocebo effects on itch induced through classical conditioning and observational learning, that positive affect would negatively correlate with these effects, and that trait empathy would positively correlate with nocebo effects induced through observational learning. For educational purposes (bachelor student theses) measures of sleep quality and social desirability were also included in the study, but are not reported here (data from these surveys is available upon request from the first author (Joseph S. Blythe, j.s.blythe@fsw.leidenuniv.nl)).

Reliability. The response data from the STAI had acceptable internal consistency ($\alpha=0.74$), as did the data from the PSS ($\alpha=0.85$), and the PANAS (positive affect $\alpha=0.86$, negative affect $\alpha=0.71$). Data from 2 of the B-IRI subscales indicated acceptable reliability (fantasy $\alpha=0.77$, perspective taking $\alpha=0.71$), while data from the other 2 scales was not found to be reliable (personal distress $\alpha=0.54$, empathic concern $\alpha=0.64$).

Correlations between psychological factors and the nocebo effect

No significant correlations were detected between psychological factors and the nocebo effect on itch in either the conditioning or observational learning group (Tables SII and SIII).