Nocebo effects are stronger and more persistent than placebo effects in healthy individuals

Angelika Kunkel[1†], Katharina Schmidt[1†], Helena Hartmann[1], Torben Strietzel[1], Jens-Lennart Sperzel[1], Katja Wiech [1,2†] & Ulrike Bingel[1†*]

[1]Department of Neurology, Center for Translational Neuro- and Behavioral Sciences (C-TNBS), University Medicine Essen, University Duisburg-Essen, Hufelandstr. 55, 45147 Essen, Germany

[2] Wellcome Centre for Integrative Neuroimaging (WIN), Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK

† Both authors contributed equally

* Corresponding author: Ulrike Bingel (Ulrike.Bingel@uk-essen.de)

Keywords: placebo hypoalgesia, nocebo hyperalgesia, treatment expectation, pain modulation, conditioning

Abstract

Placebo and nocebo effects illustrate the profound influence of cognitive-affective processes on symptom perception and treatment outcomes, with the potential to significantly alter responses to medical interventions. Despite their clinical relevance, the question of how placebo and nocebo effects differ in strength and duration remains largely unexplored. In this preregistered study, we used a within-subject design in 104 healthy to investigate and directly compare the magnitude and persistence of placebo and nocebo effects on experimental pain. Effects were assessed immediately after their induction through verbal instructions and conditioning and at a one-week follow-up. Significant placebo and nocebo effects were detected on day 1 and day 8, but nocebo effects were stronger on both test days. Sustained effects after one week were primarily predicted by individuals' experienced effects on day 1. Our findings underscore the enduring nature of placebo and nocebo effects in pain, with nocebo responses demonstrating consistently greater strength, which is consistent with an evolutionarily advantageous 'better-safe-than-sorry' strategy. These insights emphasise the significant impact of nocebo effects and stress the need to prioritise efforts to mitigate them in clinical practice.

Introduction

Placebo and nocebo effects are intriguing phenomena that have generated considerable research interest in medicine, psychology, and neuroscience[1–4]. Belief in the effectiveness or ineffectiveness of a treatment can reduce or decrease symptoms, highlighting the powerful interaction between perception, physiology and cognitive-affective processes. Harnessing the power of positive expectations could complement standard medical treatments, and thereby enhance overall treatment outcome[5,6]. Conversely, awareness of nocebo effects is important to minimise negative expectations and side effects in clinical practice[3,7]. Moreover, it is relevant in placebo-controlled clinical trials where nocebo effects, manifesting as adverse events in the placebo group, can decrease treatment adherence and even lead to treatment discontinuation[8]. Recent insights into both phenomena have therefore led to a growing call to systematically utilize placebo effects and to learn to avoid nocebo effects in clinical care.

While extensive investigations have focused on the psychological and neurobiological mechanisms underlying positive expectations and their effect on symptom perception[2,6], our understanding of negative expectations and nocebo effects is comparably sparse despite evidence that nocebo effects can be moderate to large in size[9]. Even less is known about the longevity of the effect, a crucial factor for assessing its impact on treatment outcome in real life scenarios.

Importantly, there is evidence suggesting that an individual's susceptibility to nocebo information may not simply mirror their capacity for placebo analgesia. Early research by Colloca et al.[10] demonstrated that a single session using non-painful stimuli induced a nocebo effect but failed to elicit a placebo effect, indicating that negative expectations may be more readily triggered than positive ones. Moreover, nocebo effects seem to generalise more easily to other symptoms or treatments[11,12]. Given the evolutionary relevance of anticipating negative, threatening and potentially harmful events it seems reasonable to assume that negative expectation and its effect on health outcome is an integral aspect of promoting safety behaviours and are thus more persistent than positive expectation. To accurately gauge an individual's capacity to produce placebo and nocebo effects and compare their magnitude and duration, it is essential to investigate both effects within the same individual.

Here we investigated immediate and sustained effects of positive and negative treatment expectations on experimentally induced heat pain in N= 104 healthy volunteers. Our experimental approach allowed for the trial-by-trial modulation of expectations for pain relief and pain aggravation in a within-subject design. Verbal instructions were combined with conditioning along with a sham electrical stimulation, which was introduced to participants as a method to 'induce frequency-dependent changes in pain

sensitivity'. Treatment expectations and pain perception of physically identical medium-level heat stimuli were assessed immediately after expectancy induction (day 1), but also one week later (day 8) to investigate the longevity of both placebo analgesia and nocebo hyperalgesia. We also assessed psychological variables to explore whether they modulate or predict an individual's susceptibility, effects, and persistence of expectancy effects on pain. We hypothesized that negative expectations and nocebo effects would be stronger than placebo effects induced on day 1, and that negative expectations and their effects are more resistant to extinction and would therefore still be stronger on day 8.

Our data confirm that, although significant placebo and nocebo effects were found on days 1 and 8, the nocebo effect was consistently stronger. Both effects were primarily influenced by the most recent experience of pain reduction and pain increase but were also susceptible to psychological factors.

Results

The calibration procedure determined one temperature level for the placebo condition and one for the nocebo condition that were equidistant from the control condition. These temperatures were used in the conditioning procedure to induce the perception of pain reduction (placebo hyperalgesia) and pain aggravation (nocebo hyperalgesia), respectively (for details see Supplementary Results). In the test sessions on days 1 and 8, however, the same medium-level temperature of the control condition was applied in all three conditions. The analyses include comparisons between all three conditions (i.e., placebo, nocebo and control) and comparisons between placebo effects (i.e., control vs. placebo) and nocebo effects (i.e., nocebo vs. control). To identify variables associated with placebo or nocebo effects on day 1 or day 8, we conducted multiple regression analyses including expected and experienced effects as well as psychological variables as potential predictors.

Placebo and nocebo effects on day 1. The comparison of pain intensity ratings acquired after the conditioned expectancy manipulation in the first test session on day 1 confirmed differences between three conditions ($F(1.28, 131.96)= 96.32$, $p< .001$) with both a significant placebo effect (control vs. placebo condition: $t(103)= 3.92$; $p< .001$; 95% CI, 2.07 to 6.32; $d= 0.38$) and a significant nocebo effect (nocebo vs. control condition: $t(103)= 14.88$; $p< .001$; 95% CI, 9.78 to 12.79; $d= 1.46$; Fig. 1A). A direct comparison of both effects revealed a stronger nocebo effect than placebo effect (nocebo effect: M= 11.29, SD= 7.73; placebo effect: M= 4.19, SD= 10.92; $t(103)= 6.56$; $p < .001$; 95% CI, 4.95 to 9.24; $d= .64$; Fig. 1B).

Placebo and nocebo effects on day 8. In the second test session, seven days after the expectancy manipulation, pain intensity ratings remained to be different between conditions (F(1.58, 153.34)= 111.93, p< .001), despite the same stimulation intensity. Participants still showed a significant placebo effect (control vs. placebo condition: t(97)= 6.06; *p*< .001; 95% CI, 3.08 to 6.09; *d*= .61) as well as a nocebo effect (t(97)= 10.79, *p*< .001; 95% CI, 7.29 to 10.58; *d*= 1.09). As on day 1, a direct comparison between both effects using difference scores showed a stronger nocebo than placebo effect on day 8 (nocebo effect: M= 8.93, SD= 8.20; placebo effect: M= 4.58, SD= 7.50; t(97)= 3.90, *p* < .001, 95% CI, 2.14 to 6.56; *d* = .39) (Fig. 1B).

Comparison of day 1 and day 8. A direct comparison of placebo and nocebo effects on day 1 and day 8 pain intensity ratings showed a main effect of *Effect* with a stronger nocebo effect (F(1,97)= 53.93, *p*< .001, η^2= .36) but no main effect of *Day* (F(1,97)= 2.94, *p*= .089, η^2 = .029). The significant *Effect* x *Session* interaction indicated that the placebo effect and the nocebo effect developed differently over time (F(1,97)= 3.98, *p*= .049, η^2 = .039). While the nocebo effect decreased significantly from day 1 to day 8 (t(97)= 2.68, *p*= .018, 95% CI, 0.66 to 4.44; *d*= 0.27), the placebo effect did not change (t(97)= -0.517; *p*= .606; 95% CI, -2.47 to 1.45, *d*= -0.05), possibly due to the lower starting point on day 1. Of note, placebo and nocebo effects were significantly positively correlated at day 1 (r= 0.34; *p*< .001) but showed no significant relationship on day 8 (r= 0.01; *p*= .903).

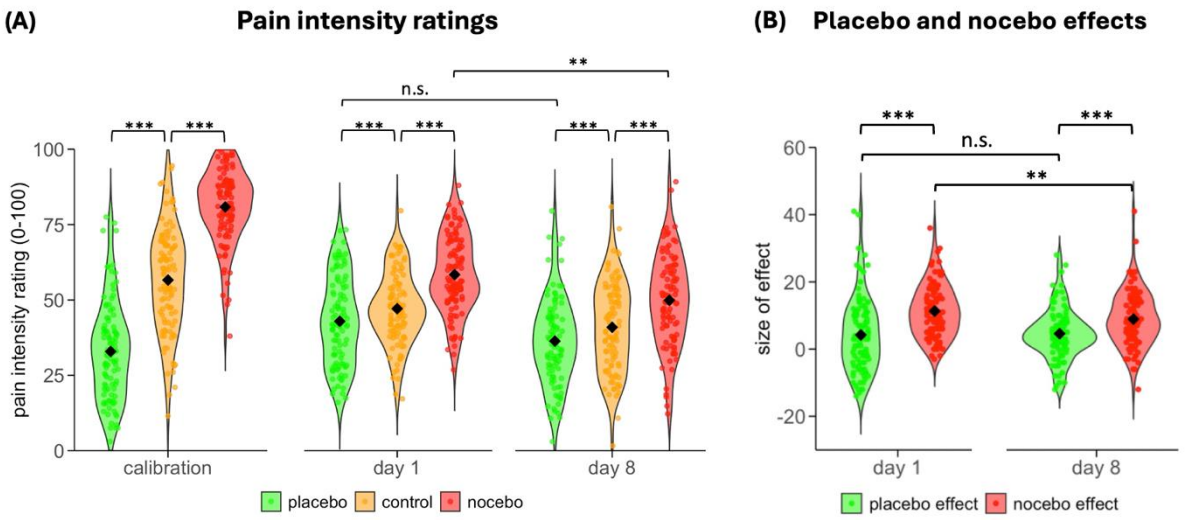


Figure 1. Pain intensity ratings and placebo and nocebo effects during calibration and test sessions. (A) Mean pain intensity ratings in the placebo, nocebo and control condition during calibration, and during the test sessions at day 1 and day 8. (B) Placebo effect (control condition vs placebo condition) and nocebo effect (nocebo condition vs control condition) on day 1 and day 8. Error bars indicate the standard error of the mean, circles indicate mean ratings of individual participants. ***: *p*< .001, **: *p*< .01, n.s.: non-significant.

Evolution of differences between placebo and nocebo effects. To test whether the difference between the placebo and the nocebo condition already evolved during conditioning, we first compared pain intensity ratings provided during conditioning where stimulus intensities were manipulated unbeknownst to the participant. As intended, heat stimuli applied during placebo conditioning were rated as less painful than stimuli applied in the control condition (control vs. placebo condition: t(103)= 20.56; $p< .001$; 95% CI, 20.98 to 25.45; $d= 2.02$). Similarly, stimuli applied during nocebo conditioning were rated as more intense than stimuli in the control condition: t(103)= 33.42; $p< .001$; 95% CI, 31.16 to 35.09; $d= 3.28$) (Fig. 2A). However, the pain ratings revealed a stronger conditioning effect for the nocebo condition than the placebo condition (nocebo effect: M= 33.12, SD= 10.11, placebo effect: 23.21, SD= 11.51; t(103)= 5.96, $p< .001$, 95% CI, 6.61 to 13.20; $d= .59$, Figure 2B).

To explore the formation of the placebo and nocebo effect during conditioning in more detail, we compared changes in trial-by-trial pain intensity ratings over the conditioning phase between the three conditions (Figure 2C). This analysis showed no significant main effect of *Trial* (F(4.37,341.01)= 1.25, $p= .289$, $\eta^2= 0.016$), indicating that there was no general change in ratings over time. However, as shown by a significant main effect of *Condition* (F(1.84,143.76) = 950.85, $p< .001$, $\eta^2= 0.924$) and more importantly a significant interaction between *Trial* and *Condition* (F(13.93,1086.45)= 4.93, $p< .001$, $\eta^2= 0.059$), changes in ratings over time differed between the three conditions. Separate regression analyses for each condition showed that although ratings decreased in the placebo condition (β= -0.22), the decrease was not significant ($p= .242$). Conversely, both the nocebo and the control condition showed an increase in ratings over time, but the increase only reached significance in the nocebo condition (β= 0.39, $p= .048$; control condition: β= 0.09, $p= .512$) indicating a stronger formation of nocebo hyperalgesia already during conditioning, despite rigorous calibration to intensities equidistant from the control condition.

To test whether the differences between placebo effects and nocebo effects on day 1 and day 8 could be explained by stronger nocebo conditioning, we repeated the previous comparisons between both effects, but this time included the difference in conditioning (nocebo condition minus placebo condition) as a covariate. While the difference in conditioning could indeed explain a significant part of the variance (F(1,102)= 5.85, $p= .017$, $\eta^2= 0.054$), the nocebo effect was still significantly stronger on day 1 (main effect *Effect*: F(1,102)= 20.79, $p< .001$, $\eta^2= 0.169$), indicating genuine differences in the underlying mechanisms and temporal dynamics. A similar (albeit weaker) result was found for day 8 with a significant difference between the placebo and the nocebo effect (main effect *Effect*: F(1,96)= 4.81, $p= .031$, $\eta^2= 0.048$) in addition to a significant effect of the difference in conditioning (F(1,96)= 4.38, $p= .039$, $\eta^2= 0.044$).

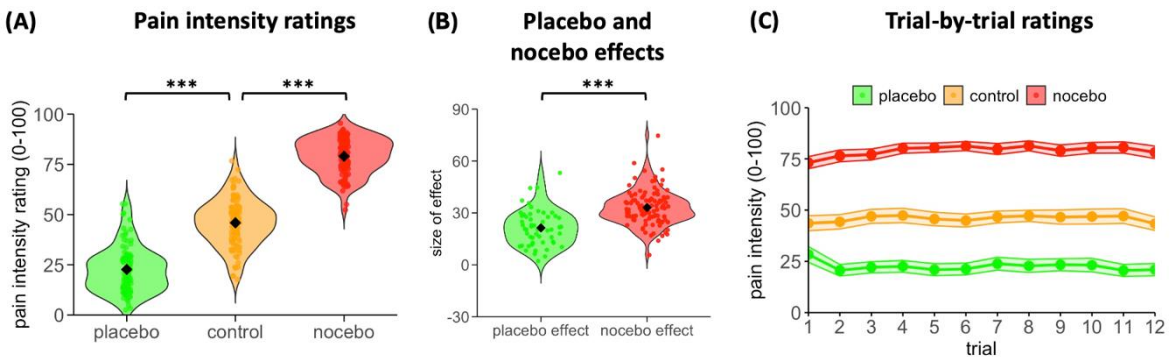


Figure 2. Mean and trial-by-trial pain intensity ratings, placebo and nocebo effects during conditioning. (A) Mean pain intensity ratings of the placebo, nocebo and control condition during conditioning. (B) Placebo effect (control condition vs. placebo condition) and nocebo effect (nocebo condition vs control condition) during conditioning. (C) Trial-by-trial pain intensity ratings (with confidence intervals) during conditioning. Error bars indicate the standard error of the mean, circles indicate mean ratings of individual participants. ***: $p < .001$.

Expectancy ratings. Given the proposed key role of expectations in placebo and nocebo effects, we also obtained expectancy ratings prior to each testing session. Because expectancy ratings were not normally distributed, we used a non-parametric analysis approach. Expectations that the pain would improve in the placebo condition and worsen in the nocebo condition did not differ significantly before conditioning, confirming that our verbal instruction had induced equally strong expectations ($Z(104) = -0.34$, $p = .737$; Fig. 3). The conditioning procedure on day 1 significantly increased the placebo improvement expectation ($Z(104) = -3.76$, $p < .001$) but not the nocebo worsening expectation ($Z(104) = -1.09$, $p = .556$) and a direct comparison showed significantly stronger placebo than nocebo expectations ($Z(104) = -2.71$, $p = .007$). Between day 1 and day 8, placebo expectations decreased significantly ($Z(98) = -3.09$, $p = .004$) and were no longer different from ratings before conditioning ($Z(104) = -0.96$, $p = 0.338$). Nocebo expectations also decreased ($Z(98) = -3.90$, $p < .001$) and were even significantly lower than before conditioning ($Z(98) = -2.30$, $p = .021$). As on day 1, the expected pain relief was significantly stronger than the expected worsening of pain ($Z(98) = -3.39$, $p = .001$).

Neither placebo nor nocebo expectations were significantly linked to the experienced effect on day 1 (placebo: Spearman's rho $(104) = 0.10$, $p = .335$; nocebo: Spearman's rho $(104) = 0.17$, $p = .093$) or day 8 (placebo: Spearman's rho $(98) = 0.13$, $p = .187$; nocebo: Spearman's rho $(98) = 0.88$, $p = .396$).

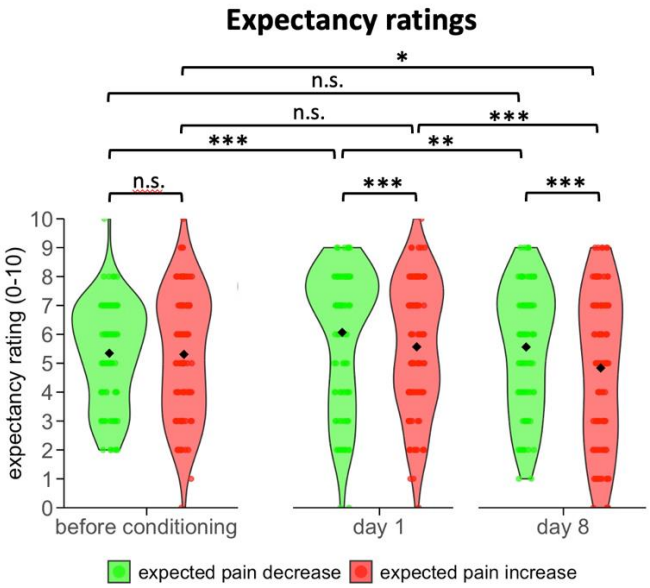


Figure 3: Expectancy ratings obtained before conditioning and before the test sessions on day 1 and day 8. Expectations were assessed using the *Generic Rating Scale for Previous Treatment Experiences, Treatment Expectations, and Treatment Effects* (GEEE, Rief et al., 2021). The expected pain relief was derived from the item asking how much improvement the participant expected from the treatment on a 10-point Likert-scale from 0 (= no improvement) to 10 (= greatest improvement imaginable). Analogously, the expected pain increase (nocebo effect) was taken from the item asking how much worsening of pain they expected from the treatment from 0 (= no worsening) to 10 (= greatest worsening imaginable). Black diamond shapes indicate the mean and circles the individual scores. ***: $p < .001$, **: $p < .01$, * $p < .05$, n.s.: non-significant

Multiple linear regression analyses (expected and experienced effects): Next, we employed multiple linear regression analyses to investigate the significance of expected (GEEE ratings) and experienced placebo and nocebo effects (VAS ratings) for subsequent effects on both test days. Overall, the regression model for the placebo effect on day 1 explained 9.7% of the variance (Supplementary Table S1). The only predictive variable for the placebo response on day 1 was the placebo effect during conditioning. In the equivalent model for the nocebo effect, none of the variables could significantly predict the nocebo response on day 1.

The regression model for the placebo effect on day 8 explained a total of 25.1% of the variance with two significant predictors: the placebo effect on day 1 and the placebo expectation on day 8 (Supplementary Table S1). For the nocebo response on day 8, the tested model explained 7.1% of the variance with the nocebo effect at day 1 as the only significant predictor (Supplementary Table S1). Together, these differences in the contribution of expectations and experienced effects between the

placebo and the nocebo condition further substantiate that both effects are driven by different mechanisms.

Multiple regression analyses (expected and experienced effects plus psychological variables)

In the final analysis step, we tested whether psychological variables that have been linked to placebo and nocebo effects in the past, such as trait anxiety[14] or practitioner characteristics[15] could increase the predictive power of the previously tested models. On day 1, in addition to the significant prediction from the experienced conditioning effect that had already been significant in the previous model, somatosensory amplification emerged as a negative predictor of the placebo effect, indicating that individuals with a higher tendency for somatosensory amplification were less likely to experience placebo analgesia. The total variance explained in this model was 14.5% (Supplementary Table S2). This influence of somatosensory amplification was no longer detectable on day 8 where only the experienced placebo effect on day 1 and placebo expectations on day 8 were significant predictors but none of the psychological variables (total amount of variance explained: 26.4%).

The equivalent analyses for the nocebo effect revealed that higher nocebo effects were found when participants had rated the experimenter competence as high (Supplementary Table S2), pointing towards a potential iatrogenic effect of experimenter when they implied that pain could become worse with the treatment. The total amount of variance explained by this model was 10.6%. As for the placebo effect, none of the psychological variables predicted the nocebo effect on day 8. The total variance explained by this model with only the perceived nocebo effect on day 1 as a significant predictor was 1.6%.

Discussion

In this pre-registered, experimental study in healthy individuals, we investigated placebo analgesic and nocebo hyperalgesic effects immediately after a conditioned expectancy manipulation and seven days later. Three key findings emerged from our investigation. First, medium-to-large scale placebo and nocebo effects were found not only on day 1 but also one week later. Second, nocebo effects were consistently stronger than placebo effects, including during the conditioning phase, despite analogous conditioning protocols in both conditions. Third, placebo and nocebo effects are primarily driven by the most recent experience of these effects but were also susceptible to some psychological factors.

Sustained placebo and nocebo effects. While placebo effects have been shown to persist for an extended period of time after they have been induced, there are only a few studies that have investigated the longevity of nocebo effects so far and these studies focused on sustained effects

within the same test session[10,16]. In our study, nocebo effects were not only sustained over the period of a week, but they were also significantly stronger than the placebo effect on both test days (Fig. 1B). This finding aligns with broader evidence from learning studies, which demonstrate a greater influence of negative information on sensory perception[17–19], as well as similar effects observed in placebo and nocebo trials. For example, nocebo hyperalgesia was more easily induced via instructions than placebo analgesia[20] and tended to extinguish more slowly[16,20]. Additionally, in an experimental study involving healthy individuals, Colloca and colleagues[10] found that one session of conditioning was sufficient to induce a nocebo effect but not a placebo effect.

Stronger and more sustained nocebo effects are likely to be the result of a combination of different factors. Evolutionary psychology suggests that humans may have evolved to be more attuned to potential threats for survival. Negative information or expectations about harm may have carried more evolutionary significance, making individuals more sensitive to nocebo suggestions, a tendency often referred to as 'better safe than sorry'. Confirmation for this assumption comes from brain imaging studies demonstrating a cognitive bias in which the brain tends to process negative information more readily than positive information. Moreover, negative expectations and fear tend to amplify sensory perception[17–19]. When individuals anticipate negative outcome, their attention is often heightened which makes them susceptive to perceiving symptoms, even in the absence of an actual stimulus. In line with this assumption, nocebo effects have been shown to lead to anticipatory anxiety and autonomic arousal which mediated the effect on extinction in an experimental learning model[16]. It may be argued that the dominant nocebo effect observed in our study is the result of the stronger conditioning in the nocebo condition (Fig. 2). This asymmetry is noteworthy in and of itself because it occurred despite the equidistant stimulus calibration relative to the control condition prior to conditioning. It may be the result of amplified learning in the nocebo condition, consistent with its heightened biological relevance, but it could also be a stronger effect of the verbal instructions in this condition. Importantly, the stronger nocebo effect observed on both test days remained significant even after accounting for the asymmetric conditioning effect, ruling out that conditioning differences alone explain the stronger nocebo effects. Instead, it suggests that the two effects may be induced and maintained by at least in part distinct mechanisms and temporal dynamics. This is supported by the observation that, similarly to a previous study[10], a significant correlation between placebo and nocebo effects was found on day 1 but was no longer detectable at the follow-up one week later. Interestingly, our expectancy manipulation increased placebo expectations, but had no significant effect on nocebo expectations (Fig. 3). Furthermore, expectations were not correlated with actual placebo or nocebo effects on either test day. While this may seem surprising, it has recently been suggested that these correlations depend on whether expectations are measured in the same format as the pain experience

or as a difference measure, as in our study[21]. Further research is therefore needed to investigate the effects of assessment methods on such associations.

Past effects predict future effects

To explore the relative influence of expectations and prior experience in more detail, we conducted separate regression analyses for placebo and nocebo effects on both test days, using expectations and perceived effects as predictors. The analyses revealed that experienced pain reduction and increase were significant predictors of subsequent effects, especially for the placebo effect on day 1 and day 8, and for the nocebo effect on day 8 (Supplementary Table S1). This highlights the strong impact of sensory experience on subsequent effects, in line with studies on learning[22], meta-analyses of behavioural placebo analgesia[23], and previous studies on carry-over effects between analgesic treatments[11]. Notably, the most recent experience was the most predictive in all three analyses; for instance, the placebo effect on day 8 was predicted by the placebo effect on day 1, not by the initial conditioning. This finding supports the Bayesian inference framework, where recent experiences are weighted more heavily in the process of model updating because they are more likely to reflect the current state of the environment, providing the most relevant and immediate information needed to guide future actions and predictions[24]. Interestingly, while a change in pain predicted subsequent nocebo effects, it seemed less influential than for placebo effects. This aligns with findings that longer conditioning enhanced placebo effects, while it did not affect nocebo responses[10] and the conclusion that nocebo instruction may be sufficient to trigger nocebo responses.

The role of psychological variables in immediate and sustained placebo and nocebo effects

Our extended regression models, incorporating psychological variables, highlight two interesting predictors: somatosensory amplification and perceived practitioner competence (Supplementary Table S2). Somatosensory amplification, described as a tendency to experience bodily symptoms as intense, noxious and disturbing[25], was associated with a weaker placebo effect on day 1. This may be due to higher-level evaluative processes[26], leading individuals to perceive symptoms as more threatening, which in turn diminishes the influence of cognitive processes that typically drive placebo effects. Additionally, our study suggests that nocebo effects can be linked to the perceived competence of the experimenter. While practitioner competence – alongside perceived warmth – usually enhances positive treatment expectations[27] and treatment outcome[15,28], it might also make negative suggestions more convincing and thereby amplify nocebo responses through increased anxiety or hypervigilance. This finding underscores the dual-edged nature of competence in patient-practitioner interactions, where heightened credibility could inadvertently strengthen nocebo effects.

Our findings have important implications for clinical research and practice. First, they underscore the necessity of prolonged observation periods in clinical trials to accurately capture the durability of these effects. Second, they emphasise the importance of not dismissing early signs of nocebo effects as they may persist and undermine otherwise treatments if left unaddressed. Third, our findings advocate for a stronger focus on the prevention of nocebo effects. While considerable effort has been made to leverage placebo effects, it is equally - if not more - crucial to focus on minimising nocebo effects, which seem to be triggered more easily. Fortunately, nocebo effects can often be avoided by adopting simple, effective strategies to improve patient-practitioner communication. For example, positive framing, avoiding unnecessary focus on potential side effects, or building a trusting relationship can reduce the likelihood of triggering nocebo effects. In a time where cost-effectiveness is paramount, and healthcare resources must be carefully allocated, prioritizing the prevention of nocebo effects should be a key strategy to enhance treatment outcome and reduce overall healthcare costs.

In summary, our findings indicate that nocebo effects are indeed more than the flipside of a placebo effect and that the two phenomena may be sustained by distinct mechanisms. These insights shed light on the factors that exacerbate nocebo effects and underscore the importance of carefully managing communication in clinical and experimental settings.

Materials and Methods

Participants

A total of N= 112 healthy volunteers were recruited through public adverts and received structured telephone interviews for screening purposes. Exclusion criteria comprised red-green colour blindness, drug use in the last four weeks, alcohol consumption in the last 24 hours, caffeine consumption on the test day, acute or chronic pain, a history of or acute psychiatric disorders (including major depression, schizophrenia and suicidality), hypersensitivity or other neurological diseases, acute infections, skin diseases, surgical procedure under anaesthesia in the last six months, use of analgesic or anticoagulant medications within the last 24 hours, use of any other medication in the last 7 days (except thyroid medication, hormonal contraceptives or allergy medication), pregnancy or breastfeeding. People were also ineligible if they had taken part in another study using electrical stimulation or experimental heat pain in the last six months before the study. Eight participants were excluded on the first testing day, two because of technical problems, two because they did not meet the inclusion criteria (due to caffeine consumption and yellow fever vaccination), and four showed a low or inconsistent pain sensitivity rendering the experimental manipulation ineffective (e.g., 80% of the pain stimuli were rated with a VAS score of 0). The final sample for the analyses of day 1 consisted of 104 participants (63 female and 41 male, mean ± SD age: 24.92 ± 3.47, range = 18 to 36 years). Six participants were unable to take part in the follow-up examination on day 8 for the following reasons: one due to personal illness, two because of the experimenter's illness, one failed to attend, another participated in a similar experiment between sessions, and one took pain medication on day 8. As a result, the final sample for day 8 consisted of 98 participants (59 female, 39 male, mean age ± SD: 24.86 ± 3.29 years, range: 18 to 36 years). The study was preregistered with the German Clinical Trials Register (https://www.drks.de; registration number: DRKS00029228). Ethics approval was granted by the University Hospital Essen (22-10597-BO). The experiment adhered to the principles outlined in the 2013 Declaration of Helsinki. Informed written consent was obtained from all participants, who received 120 Euros for their participation.

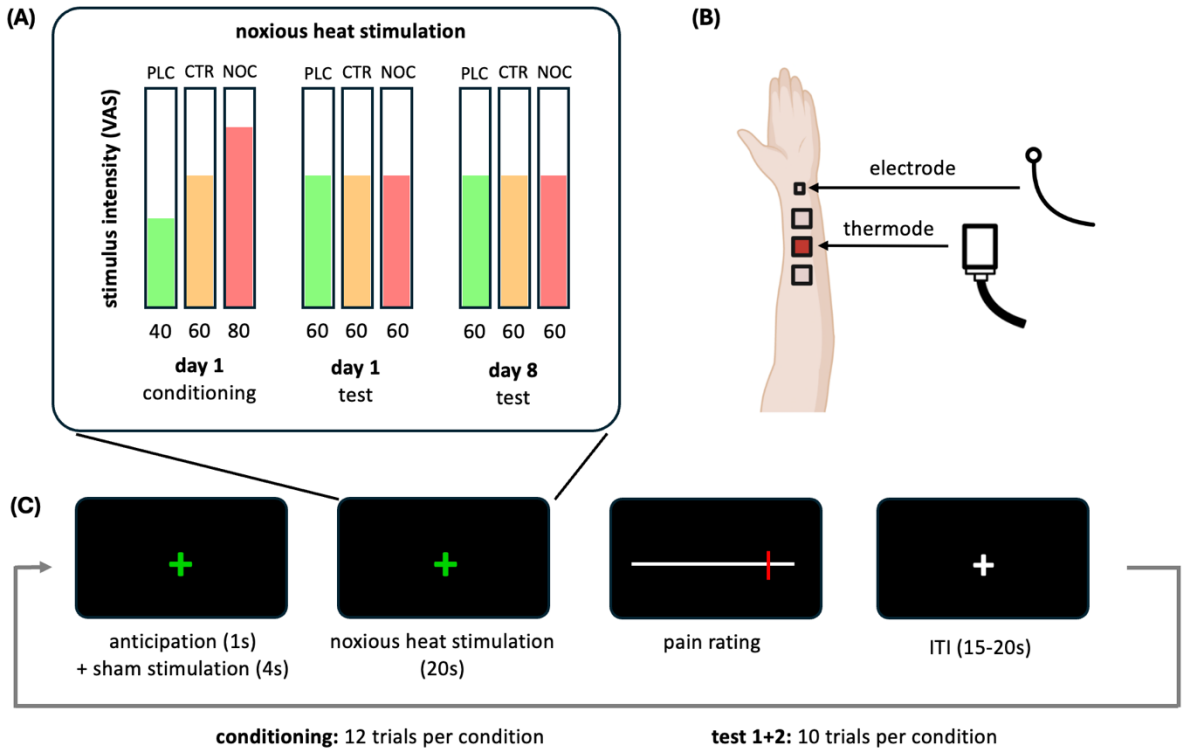
Study design and procedure

This study used a within-subjects design (Fig. 4) to investigate the immediate and sustained effects of three types of experimentally induced treatment expectations on heat pain perception: expectations of reduced pain (placebo condition), expectations of increased pain (nocebo condition) and expectations of no change in pain (control condition). The experiment was carried out on two days. On the first day (day 1), treatment expectations were induced using verbal instructions in combination with a conditioning procedure. During conditioning, participants learned to associate the presentation

of one of three visual, differently coloured cues with a reduction of heat-induced pain through a (sham) 'transcutaneous electrical nerve stimulation (TENS) device' that was introduced as an analgesic treatment in the placebo condition. A second cue signalled an increase in pain the nocebo condition and the third cue signalled no change in pain in the control condition. As in previous studies using conditioning to induce placebo and nocebo effects[10,29–31], unbeknownst to the participant the heat stimulation was reduced from VAS 60 to VAS 40 in the placebo condition, increased to VAS 80 in the nocebo condition and left unchanged at VAS 60 in the control condition. In the subsequent first test session, the same moderate stimulation intensity of VAS 60 was used in all three conditions. To explore the longevity of the induced conditioned effects, participants underwent the same testing procedure but no conditioning a week later (day 8) with all three visual stimuli again followed by the same moderate temperature stimulation (VAS 60). Participants' condition-specific treatment expectations and trial-by-trial pain intensity ratings were recorded as outcome measures. The study also comprised structural and functional magnetic resonance imaging (MRI) that took place on a separate day before day 1 (methods and data on this part will be reported elsewhere).

During the experiment, the participants were seated in a chair in front of a computer in a behavioural laboratory setting with a keyboard as response device. The left arm was positioned on a long cushion resting on the table while the right hand operated the keyboard. The experimenter faced the participant from the opposite side of the table with the computer screen between them.

Presentation of visual stimuli, delivery of thermal and electrical stimuli, and outcome recording were implemented using Presentation (Version 22.0, Neurobehavioral Systems, Inc., Berkeley, CA).



395

396 **Fig. 4: Study and trial design.** (A) Study design: On day 1, participants underwent a conditioning
397 procedure in which a noxious heat was applied directly after a (sham) TENS stimulation in three
398 conditions. In the placebo condition (PLC), the thermal stimulation was lowered to VAS 40, in the
399 nocebo condition (NOC), it was increased to VAS 80 and in the control condition (CTR) it remained
400 unchanged (VAS 60). During the two tests on day 1 and day 8, the same moderate stimulation intensity
401 of VAS 60 was applied in all three conditions. (B) Position of the electrode on the inner lower left arm
402 for (sham) TENS stimulation (approximately 2.5 cm above the wrist) and the thermode at three
403 possible locations (approximately, 3.5 cm above the electrode with a distance of 0.5 cm between each
404 of the three locations. (C) Trial design: Following the presentation of a visual cue to indicate the
405 condition (e.g., green cross for the placebo condition), first the sham TENS stimulation and then the
406 heat stimulus were applied before participants rated the pain intensity on a visual analogue scale.

407 **Calibration of the noxious thermal stimulation.** Heat stimuli were calibrated to each participant's
408 level of sensitivity. First, we used the Method of Limits[32] to determine the individual heat pain
409 threshold (HPT) in three consecutive trials. In the subsequent calibration procedure, participants rated
410 21 noxious heat stimuli with varying temperature levels around the HPT (-1°C – +3.5°C) on a visual
411 analogue scale (VAS) with endpoints 0 (= "not painful at all") and 100 (= "unbearably painful"). These
412 ratings were entered into a linear regression (lm(VAS rating ~ temperature)) in RStudio (except for the
413 first rating due to familiarisation effects) to determine the temperature levels rated as VAS 40, 60 and
414 80. These temperatures were applied twice in a short subsequent test to ensure that the calculated

heat levels induced the intended pain intensity. The 20-second contact heat pain stimuli were delivered using a Pathway advanced thermal stimulator with a 30 x 30 mm activation area (Pathway System, Medoc, Israel). The thermode was attached to one of three possible locations on the medial inner aspect of the left forearm using a tourniquet, maintaining a standardised distance of 3.5 cm from the electrode maintained via a template. To prevent sensitization or habituation, the three different stimulation sites were used. The thermode was moved to another of the three locations after calibration and conditioning, following a fixed, pseudorandomised order.

(Sham) transcutaneous electrical nerve stimulation. Participants were instructed that the applied non-painful electrical stimulation with different frequencies would either increase, decrease, or not influence pain perception, respectively**.** The electrical stimuli were applied to the left volar forearm approximately 2.5 cm proximal of the wrist using a Digitimer stimulator (Welwyn Garden City, England, model DS7A) that was connected to a surface electrode (Specialty Developments, Bexley, UK) with a diameter of approximately 5mm attached to the skin using medical tape. During calibration, the initial stimulation intensity for the four-second train of stimuli started at 0.9 mA and increased in increments of 0.1 mA until participants noticed a clear but non-painful sensation. This intensity was then tested by applying four four-second stimuli. If participants rated at least 2/4 of the stimuli between 25 and 35 on a VAS from 0-100, this final stimulation intensity was carried forward to be used throughout the test sessions. If the electrostimulation was not perceivable on day 8, the calibration was repeated once more before the start of the other experiments.

Conditioning procedure. During the conditioning session, participants' expectations of pain relief and pain increase were modulated using verbal instructions and electrical stimulation coupled with coloured visual cues. Specifically, participants were told that the electrical stimulation would either increase (nocebo instruction), decrease pain (placebo instruction), or have no influence on their pain perception (control instruction) depending on the frequency of the stimulation. The direction of change would be indicated by the colour of a cross that was shown in the centre of the computer screen. A green cross indicated a decrease in pain (placebo condition), a red cross indicated an increase of pain (nocebo condition) and a yellow indicated no change (control condition). In fact, unbeknownst to the participants, in placebo trials the green cross was followed by low-intensity heat stimulation calibrated at VAS 40 to induce a sense of pain reduction through the electrical stimulation, whereas the red cross was followed by a high-intensity heat pain calibrated at VAS 80 for a sense of pain increase (Fig. 4). In control trials the yellow cross was followed by a VAS 60 heat pain stimulus. The order of condition was pseudorandomised, and each trial type was repeated twelve times during the conditioning procedure. Due to a randomisation error, 25 participants received an unbalanced number of trials per condition (i.e., 10 x VAS 40, 14 x VAS 60, 12 x VAS 80). However, mean pain intensity ratings

during the conditioning phase did not differ significantly between these participants and the remaining sample in any of the three conditions (2-sample t-test (2-sided); placebo condition: t(102)= -0.806, *p*= .422, nocebo condition: (t(102)= 0.849, *p*= .398, control condition: (t(102)= 0.390, *p*= .697).

Test sessions*.* Placebo and nocebo responses were assessed during both test sessions on day 1 and day 8 following the same procedure as the conditioning session, but without temperature manipulation. Instead, the same target temperature corresponding to VAS 60 was maintained across all conditions (see Figure 4 for details of the design). On day 8, one stimulus per stimulation intensity (i.e., VAS 40, 60 and 80) was applied before the start of the test session to re-familiarise participants with the thermal stimulation.

Pain intensity ratings. During the conditioning and the test sessions, participants provided pain intensity ratings on a VAS with endpoints points 0 (= "not painful at all") and 100 (= "unbearably painful"). The cursor was positioned randomly on the scale at the beginning of the rating period. Participants could move the cursor by pressing the left or right arrow key and were asked to confirm their rating with the 'enter' key (no time limit).

Reaction time task. During the conditioning and test sessions, a reaction time task was included at the beginning of each trial in which participants had to respond as quickly as possible to a target stimulus (a blue cross) by pressing the left arrow key to ensure sustained attention. The blue cross appeared for 300 ms with a jittered onset at the beginning of each trial, i.e., 0 to 5 s after trial onset.

Psychological questionnaires. Before calibration on day 1, participants completed the German version of the following questionnaires using an online survey system (LimeSurvey, LimeSurvey GmbH, Hamburg, Germany): the Generic Rating for Treatment Pre-Experiences, Treatment Expectations, and Treatment Effects (GEEE[13]), the Somatosensory Amplification Scale (SSAS[25,33]), the Perceived Stress Scale (PSS-10[34,35]), anxiety and depression (STADI Trait[36]) and the Pain Catastrophizing Scale (PCS[37,38]). Warmth and competence of the experimenter were assessed as described in Seewald & Rief[27] at the end of day 1. In short, participants were asked the question how the experimenter seemed to them and provided ratings on a 5-point scale ranging from 1 (= not at all) to 5 (= extremely) for the following descriptors in German: "friendly", "well-intentioned", "trustworthy", "warm", "good-natured" and "sincere" to capture experimenter warmth and "competent", "confident", "capable", "efficient", "intelligent" and "skilful" for experimenter competence. The mean across items of each scale was used in further analyses.

Treatment expectation ratings using the GEEE and the emotional state using STADI State were also collected before conditioning, after conditioning, and before test session 2 on day 8. Treatment effects

were rated after conditioning and after test sessions 1 and 2. Note that participants also completed the following questionnaires as part of a larger project: Fear of Pain Questionnaire (FPQ-III), Behavioral Inhibition and Behavioral Activation (BIS/BAS) Scales, 10-item-Big-Five-Inventory (BFI-10) and the Positive and Negative Affect Schedule (PANAS). Responses to these questionnaires will be analysed elsewhere.

Statistical analyses

Data were analysed using R (version 4.4.1). For each of the three conditions, mean pain intensity ratings for the calibration phase, conditioning phase and tests on day 1 and day 8 were calculated across the trials of the respective phase. Nocebo effects were defined as the difference in pain intensity ratings between the nocebo and the control condition (nocebo - control), placebo effects as the difference between the control and the placebo condition (control - placebo). Comparisons of stimulation intensities and pain intensity ratings between conditions were carried out using repeated-measures ANOVAs with the within-subject factor *Condition* (placebo, nocebo, control) followed by post hoc Bonferroni-corrected pairwise comparisons. For a comparison of pain intensity ratings between time-points, an ANOVA with the within-subject factors *Condition* (placebo, nocebo, control) and *Session* (day 1, day 8) was carried out. For the comparison of placebo and nocebo effects between the two test days, an ANOVA with the with-subject factors *Effect* (placebo effect, nocebo effect) and *Session* (day 1, day 8) was used. The analysis of trial-by-trial ratings used an ANOVA with the within-subject factors *Condition* (placebo, nocebo, control) and *Trial* (trial 1 to 12). To account for interindividual differences in conditioning, the difference between the nocebo effect and the placebo effect during the conditioning phase was entered as a covariate in the comparison of pain intensity ratings at day 1 and day 8 (ANCOVA). Degrees of freedom were corrected using the Greenhouse-Geisser estimate of sphericity. To explore the relationship between placebo and nocebo effects on both test days, we calculated the Pearson correlation coefficient. Because expectancy ratings were not normally distributed, non-parametric Wilcoxon signed rank tests were used to compare these ratings between conditions and timepoints and Spearman's rho was calculated for correlations between pain intensity and expectancy ratings. All questionnaires were analysed according to their respective manuals.

Separate multiple linear regression analyses were performed to examine the influence of expectations (GEEE ratings) and experienced effects (VAS ratings) on subsequent placebo and nocebo effects. For day 1, the placebo effect was entered as the dependent variable and the following variables as potential predictors: (i) expected improvement with placebo before conditioning, (ii) placebo effect during conditioning and (iii) the expected improvement with placebo before the test session at day 1. The equivalent analysis was conducted for the nocebo effect but with (i) expected worsening with

nocebo before conditioning, (ii) nocebo effect during conditioning and (iii) the expected worsening with nocebo before the test session at day 1 as predictors.

To predict placebo responses a week later ($VAS_{control}$ – $VAS_{placebo}$ at day 8), the same independent variables were entered as for day 1 but with the following additional variables (i) the placebo effect at day 1 and (ii) the expected improvement with placebo before the test session at day 8. In the equivalent analysis for the nocebo effect on day 8 as dependent variable, we added (i) the nocebo effect at day 1 and (ii) the expected worsening with nocebo before the test session at day 8.

To explore whether psychological variables could explain additional variance in the regression analyses, we repeated all four analyses described above but included scores from these questionnaires as additional independent variables: Somatosensory Amplification Scale (SSAS), Perceived Stress Scale (PSS-10), trait anxiety and depression (STADI trait), Pain Catastrophizing Scale (PCS) and experimenter warmth and competence scores.

In all analyses, a significance level of $p < .05$ was used, and pairwise comparisons were conducted using two-tailed p-values. For all multiple regression analyses, the regression coefficient is reported.

Acknowledgments

The work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 422744262 - TRR 289 (gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 422744262 – TRR 289).

Author contributions

AK and KS contributed equally to this work as shared first authors. AK: Programming, Data acquisition, data analysis, writing, review & editing. KS: Conception, data analysis, interpretation, writing, review & editing. JS: data acquisition, writing. TS: data acquisition. HH: data curation, data analysis, interpretation, review & editing. KW: data analysis, interpretation, writing, review & editing. UB: funding acquisition, project administration, conceptualization, methodology, supervision, writing, review & editing

548 **References**

549 1. Wager, T. D. & Atlas, L. Y. The neuroscience of placebo effects: connecting context, learning and
550 health. *Nat. Rev. Neurosci.* **16**, 403–418 (2015).
551 2. Petrie, K. J. & Rief, W. Psychobiological mechanisms of placebo and nocebo effects: pathways to
552 improve treatments and reduce side effects. *Annu. Rev. Psychol.* **70**, 599–625 (2019).
553 3. Colloca, L. & Barsky, A. J. Placebo and nocebo effects. *N. Engl. J. Med.* **382**, 554–561 (2020).
554 4. Chen, C. *et al.* Neural circuit basis of placebo pain relief. *Nature* **632**, 1092–1100 (2024).
555 5. Enck, P., Bingel, U., Schedlowski, M. & Rief, W. The placebo response in medicine: minimize,
556 maximize or personalize? *Nat. Rev. Drug Discov.* **12**, 191–204 (2013).
557 6. Bingel, U. Placebo 2.0: the impact of expectations on analgesic treatment outcome. *Pain* **161**, S48–
558 S56 (2020).
559 7. Bingel, U. & for the Placebo Competence Team. Avoiding nocebo effects to optimize treatment
560 outcome. *JAMA* **312**, 693–694 (2014).
561 8. Colloca, L. The nocebo effect. *Ann. Rev. Pharmacol. Toxicol.* **64**, 171–190 (2024).
562 9. Petersen, G. L. *et al.* The magnitude of nocebo effects in pain: A meta-analysis. *Pain* **155**, 1426–
563 1434 (2014).
564 10. Colloca, L., Petrovic, P., Wager, T. D., Ingvar, M. & Benedetti, F. How the number of learning trials
565 affects placebo and nocebo responses. *Pain* **151**, 430–439 (2010).
566 11. Zunhammer, M. *et al.* The effects of treatment failure generalize across different routes of drug
567 administration. *Sci. Transl. Med.* **9**, 8 (2017).
568 12. Faasse, K. *et al.* The influence of side effect information framing on nocebo effects. *Ann. Beh. Med.*
569 **53**, 621–629 (2019).
570 13. Rief, W. *et al.* Generic rating scale for previous treatment experiences, treatment expectations,
571 and treatment effects (GEEE). (2021).
572 14. Kern, A., Kramm, C., Witt, C. M. & Barth, J. The influence of personality traits on the
573 placebo/nocebo response. *J. Psychosom. Res.* **128**, 109866 (2020).

15. Howe, L. C., Goyer, J. P. & Crum, A. J. Harnessing the placebo effect: exploring the influence of physician characteristics on placebo response. *Health Psychol.* **36**, 1074–1082 (2017).
16. Colagiuri, B., Quinn, V. F. & Colloca, L. Nocebo hyperalgesia, partial reinforcement, and extinction. *J. Pain* **16**, 995–1004 (2015).
17. van der Schaaf, M. E. *et al.* Acquisition learning is stronger for aversive than appetitive events. *Commun. Biol.* **5**, 302 (2022).
18. Forkmann, K., Wiech, K., Schmidt, K., Schmid-Köhler, J. & Bingel, U. Neural underpinnings of preferential pain learning and the modulatory role of fear. *Cereb. Cortex* **33**, 9664–9676 (2023).
19. Zika, O., Wiech, K., Reinecke, A., Browning, M. & Schuck, N. W. Trait anxiety is associated with hidden state inference during aversive reversal learning. *Nat. Commun.* **14**, 4203 (2023).
20. Colloca, L., Sigaudo, M. & Benedetti, F. The role of learning in nocebo and placebo effects. *Pain* **136**, 211–218 (2008).
21. Lunde, S. J., Rosenkjær, S., Matthiesen, S. T., Kirsch, I. & Vase, L. Conclusions regarding the role of expectations in placebo analgesia studies may depend on how we investigate it: a meta-analysis, systematic review, and proposal for methodological discussions. *Psychosom. Med.* **86**, 591–602 (2024).
22. Atlas, L. Y., Doll, B. B., Li, J., Daw, N. D. & Phelps, E. A. Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. *eLife* **5**, e15192 (2016).
23. Vase, L., Riley, J. L. I. & Price, D. D. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain* **99**, 443–452 (2002).
24. Büchel, C., Geuter, S., Sprenger, C. & Eippert, F. Placebo analgesia: a predictive coding perspective. *Neuron* **81**, 1223–1239 (2014).
25. Doering, B. K. *et al.* Is somatosensory amplification a risk factor for an increased report of side effects? Reference data from the German general population. *J. Psychosom. Res.* **79**, 492–497 (2015).

26. Nakao, M., Barsky, A. J., Nishikitani, M., Yano, E. & Murata, K. Somatosensory amplification and its relationship to somatosensory, auditory, and visual evoked and event-related potentials (P300). *Neurosci. Lett.* **415**, 185–189 (2007).
27. Seewald, A. & Rief, W. Therapist's warmth and competence increased positive outcome expectations and alliance in an analogue experiment. *Psychother. Res.* **34**, 663–678 (2024).
28. Ashton-James, C. E., Forouzanfar, T. & Costa, D. The contribution of patients' presurgery perceptions of surgeon attributes to the experience of trust and pain during third molar surgery. *PAIN Reports* **4**, e754 (2019).
29. Colloca, L. & Benedetti, F. How prior experience shapes placebo analgesia. *Pain* **124**, 126–133 (2006).
30. Montgomery, G. H. & Kirsch, I. Classical conditioning and the placebo effect. *Pain* **72**, 107–113 (1997).
31. Voudouris, N. J., Peck, C. L. & Coleman, G. The role of conditioning and verbal expectancy in the placebo response. *Pain* **43**, 121–128 (1990).
32. Fruhstorfer, H., Lindblom, U. & Schmidt, W. C. Method for quantitative estimation of thermal thresholds in patients. *J. Neurol. Neurosurg. Psychiatry* **39**, 1071–1075 (1976).
33. Barsky, A. J., Goodson, J. D., Lane, R. S. & Cleary, P. D. The amplification of somatic symptoms. *Psychosom. Med.* **50**, 510–519 (1988).
34. Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress. *J. Health Soc. Behav.* **24**, 385–396 (1983).
35. Klein, E. M. *et al.* The German version of the Perceived Stress Scale – psychometric characteristics in a representative German community sample. *BMC Psychiatry* **16**, 159 (2016).
36. Laux, L., Bergner-Köther, R., Hodapp, V. & Renner, K. H. Das State-Trait-Angst-Depressions-Inventar [The State-Trait Anxiety Depression Inventory]. (2013).
37. Sullivan, M. J. L., Bishop, S. & Pivik, J. The Pain Catastrophizing Scale: development and validation. *Psychol. Assess.* **7**, 524–532 (1995).

626 38. Meyer, K., Sprott, H. & Mannion, A. F. Cross-cultural adaptation, reliability, and validity of the
627 German version of the Pain Catastrophizing Scale. *J. Psychosom. Res.* **64**, 469–478 (2008).

628