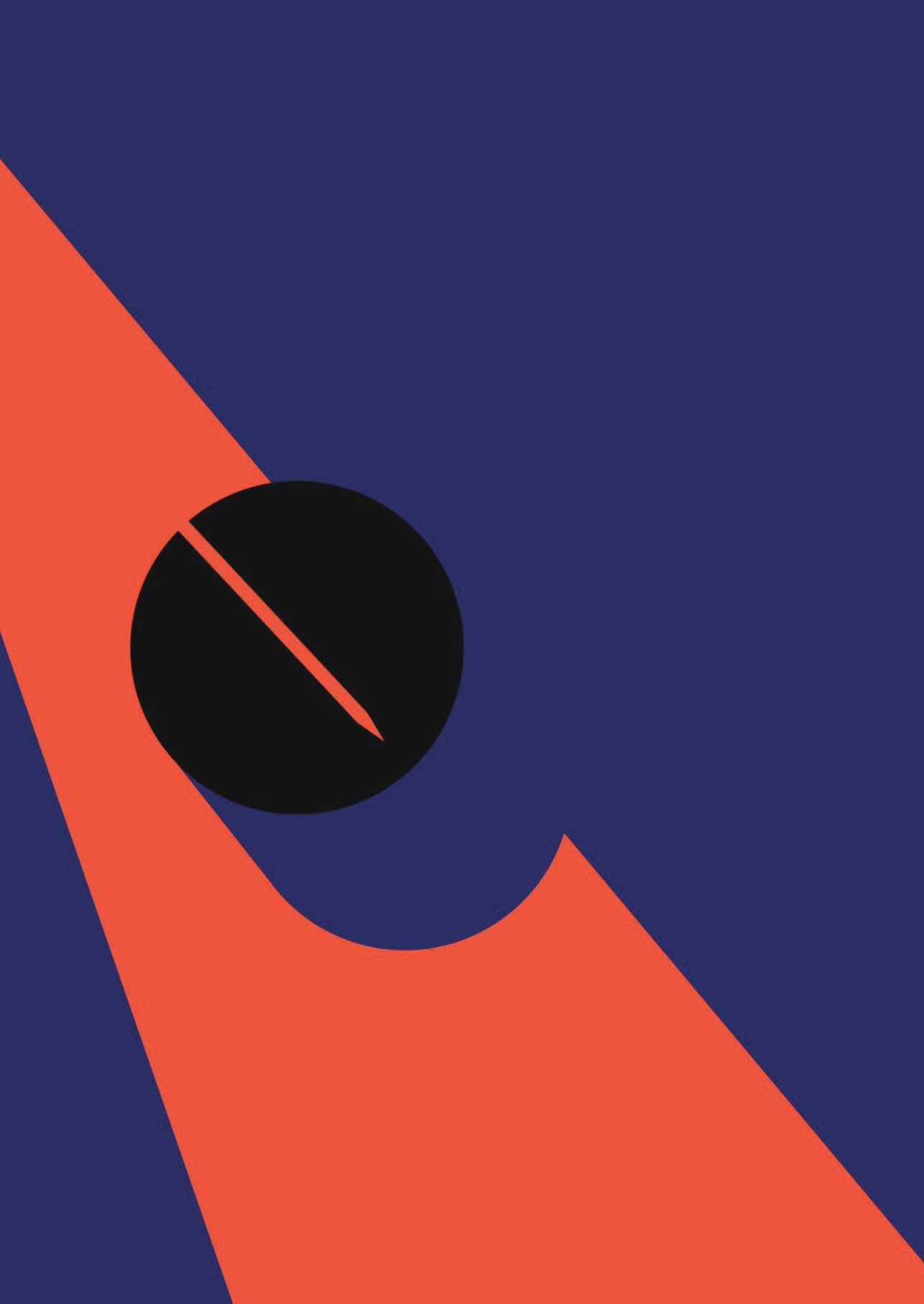
Nocebo hyperalgesia and pain progression: prediction, acquisition, and recovery

Karacaoglu, M.



CHAPTER 4

Nocebo hyperalgesia in patients with fibromyalgia and healthy controls: An experimental investigation of conditioning and extinction processes at baseline and one-month follow-up

Published as:

Karacaoglu, M., Peerdeman, K.J., Numans, M.E., Stolk, M.R., Meijer, S., Klinger, R., Veldhuijzen, D.S., van Middendorp, H., Evers, A.W.M. (2023). Nocebo Hyperalgesia in Patients With Fibromyalgia and Healthy Controls: An Experimental Investigation of Conditioning and Extinction Processes at Baseline and 1-Month Follow-up, *The Journal of Pain*, 24, 1696–1711. https://doi.org/10.1016/j.jpain.2023.05.003.

ABSTRACT

Nocebo effects are adverse treatment outcomes that are not ascribed to active treatment components. Potentially, their magnitude might be higher in patients with chronic pain compared to healthy controls since patients likely experience treatment failure more frequently. The current study investigated group differences in the induction and extinction of nocebo effects on pressure pain at baseline (*N*=69) and 1-month follow-up (*N*=56) in female patients with fibromyalgia and matched healthy controls. Nocebo effects were first experimentally induced via classical conditioning combined with instructions on the pain-increasing function of a sham TENS device, then decreased via extinction. One month later, the same procedures were repeated to explore their stability. Results suggest that nocebo effects were induced in the healthy control group during baseline and follow-up. In the patient group, nocebo effects were only induced during follow-up, without clear group differences. Extinction was only observed during baseline in the healthy control group. Further comparisons of nocebo effects and extinction indicated no significant changes across sessions; possibly suggesting their overall magnitudes were stable over time and across groups. In conclusion, contrary to our expectations, patients with fibromyalgia did not have stronger nocebo hyperalgesia; instead, they might be less responsive to nocebo manipulations than healthy controls.

Perspective: Current study is first to investigate group differences in experimentally manipulated nocebo hyperalgesia between chronic pain and healthy populations at baseline and 1-month follow-up. Since nocebo effects are common in clinical settings, their investigation in different populations is essential to explain and minimize their adverse effects during treatment.

Keywords: nocebo effect; classical conditioning; pressure pain; hyperalgesia; fibromyalgia

INTRODUCTION

Nocebo effects, which are adverse treatment outcomes unrelated to active treatment components, can occur in clinical or laboratory contexts after receiving an inert or active treatment[1]. They are presumably guided by negative expectations and can be induced and reduced by learning mechanisms[2–5]. An example of nocebo effects is the experiencing of side-effects after disclosing the potential side-effects of a medication, regardless of its pharmacological properties[6]. Various studies have investigated whether nocebo effects can be experimentally induced in healthy or in patient populations (e.g., with chronic back pain, post-operative pain, gastrointestinal disorders, or Parkinson's disease)[7–10]; however, to date, no study has directly compared the magnitude of nocebo effects between a patient and a healthy sample. Research with healthy participants indicates that nocebo hyperalgesia can be induced via classical conditioning and instructional learning, or their combination, with mixed findings on whether nocebo hyperalgesia could be extinguished by extinction[5,11–13]. One study in patients with chronic low back pain has combined conditioning of pain increase with a verbal suggestion that stated both positive and negative effects of a sham opioid treatment and found that placebo, instead of nocebo, effects were induced; possibly due to the ambiguity surrounding the verbal suggestions[10]. Conditioned nocebo effects need to be further investigated using pure verbal suggestions of pain increase, especially in chronic pain conditions such as fibromyalgia where the underlying etiopathogenesis is unclear[14].

Differences may exist in the extent in which patients with persistent physical symptoms, such as fibromyalgia, and healthy individuals are susceptible to nocebo effects. Firstly, patients have a higher cumulative exposure to treatments, which, given the existing challenges in diagnosing and treating fibromyalgia[14] and patients' possible dissatisfaction surrounding disease management[15,16], may have resulted in more negative treatment experiences surrounding treatment failure and patient-doctor exchanges[2,17,18]. Speculatively, along with biological dispositions, repetitive exposure to negative treatment experiences could establish nocebo effects that give rise to the emergence or aggravation of symptoms, and might even propagate symptom chronification over time[18–20]. Resultantly, patients may be more susceptible to acquiring stronger nocebo effects than healthy controls, which may be possibly harder to decrease via extinction[21]. Secondly, fear-conditioning studies have shown learning deficits during pain processing in fibromyalgia[22–25]. In particular, deficits related to contingency learning have been found, where a conditioned stimulus (CS+) paired with an unconditioned stimulus (US) could not be differentiated from another CS that is not paired with the US (CS-)[22]. This could eventually lead to problems with identifying safety cues in the environment that are not predictive of upcoming pain[22,23]. As such, these learning deficits may also

result in (stimulus) generalization of nocebo hyperalgesia, for instance making patients distinguish less clearly between safe and unsafe pain cues. However, the exact underlying mechanisms contributing to nocebo effects in fibromyalgia have not yet been unraveled.

With the goal of elucidating the role of nocebo hyperalgesia in fibromyalgia, the current study is the first to investigate group differences in inducing and decreasing nocebo effects on pressure pain in female patients with fibromyalgia compared to matched healthy controls. Since the majority of nocebo literature is based on findings from healthy participants, this allows us to examine whether patients have a larger magnitude of nocebo effects, which might be harder to decrease. Additionally, we explore whether inducing and decreasing nocebo effects after one month yields comparable findings with the baseline, to examine either the potential stability or progression of these effects over time. Previous literature is limited on the persistence of nocebo effects over time[10,26]. Nocebo effects on experimental pressure pain will be firstly induced by conditioning combined with verbal suggestions on the pain-increasing function of a sham TENS device, and afterwards decreased by extinction. Next, the stability of nocebo effects will be explored at one-month follow-up. Associations between psychological characteristics and the nocebo effect will also be explored for individual differences in the magnitude of nocebo effects.

METHODS

Study Design

This study is part of a larger prospective study on patients with fibromyalgia (ICTRP Identifier: NL8244) and has been approved by the Medical Ethical Committee Leiden-Den Haag-Delft (NL67541.058.18). The current study consists of two experimental sessions taking place at baseline and one-month follow-up (see Figure 1). During the baseline session, nocebo effects on pressure pain were experimentally induced in all participants via classical conditioning combined with verbal instructions about the pain-worsening function of a sham Transcutaneous Electrical Nerve Stimulation (TENS) device. With this procedure, the aim was to condition participants to expect more experimental pain in response to the sham activation of the TENS device. Next, an extinction procedure followed to examine the decrease of potentially induced nocebo effects on pain. All participants were invited to the lab for a second time after one month to take part in nearly the same experimental procedure to investigate the stability of these effects over time. The main difference was that at the 1-month follow-up, the nocebo conditioning and extinction procedures were preceded by a recall testing phase, where we aimed to assess the magnitude of recalled nocebo effects after the baseline session.

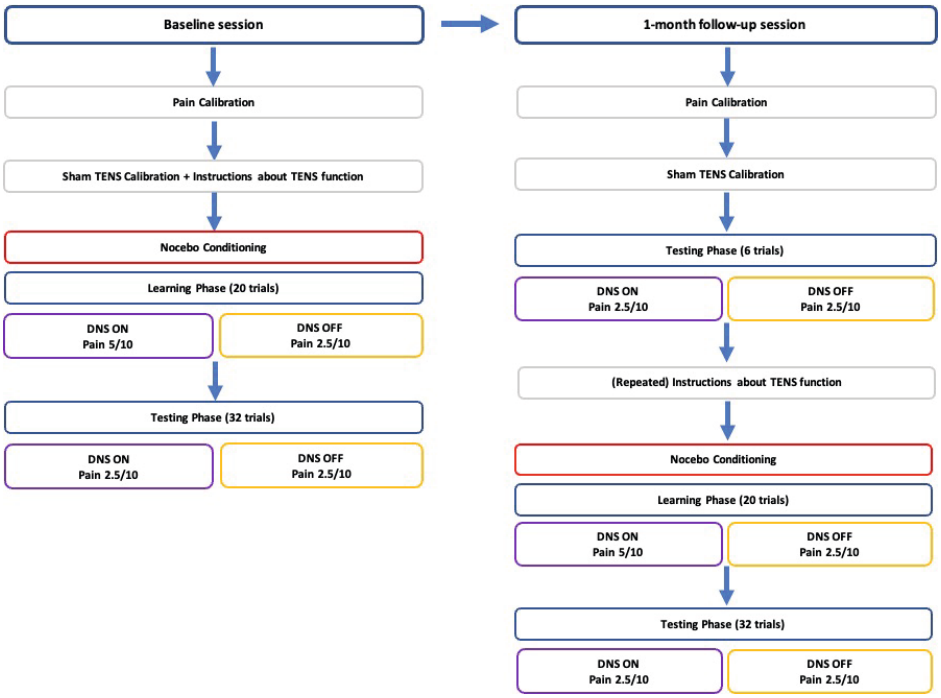


Figure 1. Illustration of the study design. Participants took part in a baseline session and a follow-up session after one month. Both lab sessions consisted of pain calibrations, sham TENS calibrations, instructions about the pain-worsening function of the TENS device, learning phase trials of nocebo conditioning and testing phase trials. The only difference between sessions was that the follow-up session began with a recall testing phase after which instructions about TENS function were repeated. During the learning phase of nocebo conditioning, participants received a moderate pressure pain stimulus when the sham TENS device (labeled as DNS device for participants), was supposedly activated, whereas they received a slight pressure pain stimulus when DNS was supposedly deactivated. In the testing phase, participants received a slight pressure pain stimulus regardless of supposed DNS (de)activation.

Participants

Sample size was calculated using G*Power 3.1[27]. Since to the best of our knowledge previous literature was not detected comparing nocebo effects in healthy and patient populations, it was decided to choose a minimal effect size that is considered clinically relevant[28], i.e., a medium effect size (Cohen's $d = 0.5$, $f = 0.25$) was selected for the planned primary analyses for the baseline and follow-up parts of the study. To conduct a mixed-design ANOVA with two groups and two repeated measurements with an alpha level of .05, a total sample size of $N = 54$ (27 per group) was needed, per session, to demonstrate a power of .95.

All participants were required to be between 18 and 65 years, fluent in the Dutch language, and able to sign an informed consent form. Since fibromyalgia is more prevalent in women than men[29], the current sample consisted of only females to increase the comparability of current findings with existing literature. Healthy controls were matched to patients based on sex, age, and education level. Education level was assessed using the Verhage scale[30], where primary education up to higher general secondary education was categorized as lower education, and higher vocational education up to university education was categorized as higher education. Patients were required to have a fibromyalgia diagnosis by a rheumatologist, which was verified during the telephone screening by patients' self-report of the year, location, and the provider of their diagnosis. Additionally, all participants, including healthy controls, filled in the Fibromyalgia Survey Questionnaire[31] to verify the presence or absence of key fibromyalgia symptoms in each group. Patients were excluded if they received a medical diagnosis other than fibromyalgia explaining their chronic pain symptoms (e.g., rheumatoid arthritis, polyarthritis) or had severe physical or mental co-comorbidities that were not related to fibromyalgia (e.g., cancer, schizophrenia). Patients were allowed to continue treatment as usual and were specifically asked not to make any changes to their usual dose of analgesic medication 24 h prior to the measurements. Healthy controls were excluded if they had chronic pain complaints (≥3 months) in the past or present, a fibromyalgia diagnosis, severe physical or mental co-morbidities that could interfere with the study protocol, current pain on the measurement days (common types of pain such as localized muscle soreness after work-out rated ≤ 3/10 on the Numeric Rating Scale were included), or used analgesic medication within 24 h prior to the measurements. The common exclusion criteria for both groups were: pregnancy or breastfeeding, color blindness, injuries or wounds on the non-dominant hand or arm, refusal to remove possible artificial nails or nail polish covering the thumbnail of the non-dominant hand, an unsuccessful pressure pain calibration procedure, i.e., not being able to stably distinguish between pressure intensities during pressure pain calibration, and as an additional safety measure due to the brief TENS activation: carrying a pacemaker or implanted pumps, or having implanted metals in the non-dominant hand or arm.

Participants were recruited via advertisements, such as flyers shared at various fibromyalgia patient organizations, pain rehabilitation centers, or Facebook. A portion of the healthy control sample was recruited via the Dutch online registry for neuroscience Hersenonderzoek.nl (www.hersenonderzoek.nl). Study participation involved taking part in the telephone screening, filling out baseline questionnaires at home, and attending two lab sessions, one at baseline and one at one-month follow-up. Participants received an ascending share of the total reimbursement in each lab session, in order to provide an extra motivation to complete the study. All participants received €50 compensation for completing all study parts with additional reimbursement of travel costs to the

lab. If a participant dropped out or was excluded during the calibration procedure, the compensation amount was adjusted based on the amount of time spent in the study. Participants gave verbal informed consent for the information collected during the telephone screening, digital informed consent for the online questionnaire, and signed informed consent for the experimental data collection in the lab.

Pressure Pain Application

Pressure pain is an ecologically valid stimulus type for disorders involving musculoskeletal pain[32], such as fibromyalgia[11]. Pressure pain was induced on the thumbnail of the non-dominant hand using a custom-built automatic pressure administrator called Pneumatic Electronic Pressure Pain Administrator (PEPPA) (see Figure 2), engineered by the Support for Research, Laboratory and Education (SOLO) team of Leiden University based on a prototype design from Karolinska Institute in Sweden[33]. To apply pressure pain, the thumb of the non-dominant hand was inserted in a transparent cylinder hand piece built by the Development and Instrumental Affairs department of Leiden University Medical Center (LUMC). Pressure was applied to the middle of the thumbnail via a piston with a 1 cm^2 probe, which automatically moved up and down by the pressured air supplied by an air compressor. Each pressure stimulus lasted 2.5 s, with a 30 s inter-stimulus interval. As a safety measure, the extension air of the cylinder was limited to 80 kPa, corresponding to a thumb force of 15 kgf/cm^2, which is the maximum pain tolerance in healthy participants that is known to be safe for pain administration[34]. Nevertheless, the current study took extra precautions by not exceeding the maximum thumb force of 13 kgf/cm^2 for both patients and healthy participants. Additionally, an emergency stop button was provided for participants to stop the pressure administration if they could no longer endure the pain. E-prime version 3.0 (Psychology Software Tools, Pittsburg, PA) was used for presenting the pressure pain stimuli and for entering participants' pain ratings after each trial.

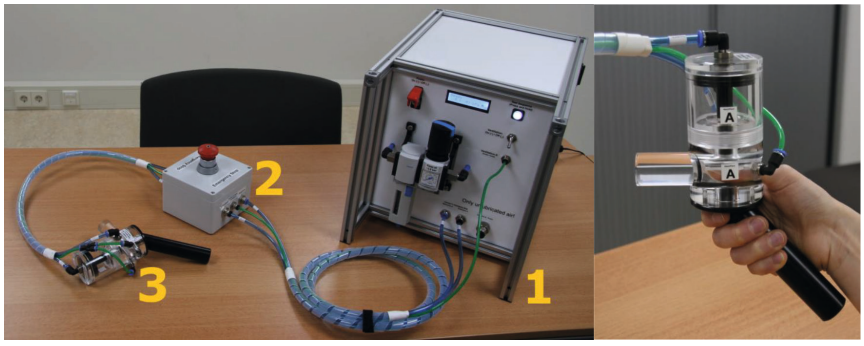


Figure 2. Picture on the left depicts the components of PEPPA. The first is the main device containing the electronics and pneumatics, the second is the emergency stop button, and the third is the hand piece for applying pressure to the thumbnail. Picture on the right demonstrates the thumb insertion into the hand piece.

Pain Measure

Following each experimental pressure stimulus, participants rated their pain intensity on a Numeric Rating Scale (NRS), with the end points 0 representing no pain and 10 the worst pain imaginable. Participants rated their pain by positioning a pointer on a digital horizontal line with anchors, each line representing a decimal on the 0-10 NRS. Participants were instructed to rate above zero (thus 0.1 upwards) when they started to feel pain next to a pressure sensation.

Pressure Pain Calibration

The calibration procedure consisted of three parts, with 5-minute breaks in-between, to minimize sensitization or habituation from repeated stimulus administration. Breaks were extended by one minute, up to five minutes, if the participant indicated still having pain ≥1 on the NRS. No participant has asked for a break exceeding the initial five minutes. Pressure intensities were administered starting from 1 kgf/cm^2 with 0.5 kgf/cm^2 increments until participants rated ≥ 5.5 on the NRS or until 13 kgf/cm^2 was reached. By choosing the highest intensity scored as zero on the NRS and the highest scored pressure intensity, 3 intermittent pressure intensities were calculated that were equidistant from each other in magnitude. Together, these five intensities were then randomly administered three times to determine the pressure intensities rated 0 (ranges 0-1), 2.5 (ranges 2-3), 5 (ranges 4.5-5.5) on the NRS to determine the non-painful, slight, and moderate pain intensities, respectively. Next, a calibration check followed where the pressure intensities for no pain, slight pain, and moderate pain were randomly administered with slight pain presented thrice and the rest presented twice. The experimenter controlled whether the pain ratings were within the targeted ranges; if not, adjusted pressure intensities were based on E-prime's calculations using standard formulas (see Supplemental File I). If manual adjustments were not possible due to the requirement of less pressure than the minimum or more pressure than the maximum amount that PEPPA could safely administer, participants were excluded.

Experimental manipulation of nocebo effects

Sham TENS device

A sham TENS device (Bentrotens T37, Bentronic Gesellschaft fuer Medizintechnik GmbH, Wolnzach, Germany) was used as conditioned stimulus in the conditioning paradigm, wherein a chip was inserted to cease the device from sending any electrical signals after 1 minute. The device was renamed as "Dermal Nerve Stimulation" (DNS) device to prevent possible preconceptions about TENS from interfering with the experimental manipulations. Participants were given a fake device leaflet that read: *"DNS is a device that stimulates nerves via electrical signals. This stimulation helps increase the communication*

between the nerve cells. Nerve cells in the skin communicate with other nerve cells in the spine via electrical signals. The DNS device can influence these signals, for example, by increasing the intensity of the signals coming from a painful stimulus. When these signals are sent from the spine to the brain, you become aware of the sensation of pain. The DNS device applies electrical signals via electrodes attached to your skin. An advantage of DNS is that a light and an (almost) unnoticeable signal is sufficient to influence the communication between the nerve cells; and therefore, to increase your pain sensation."
After participants read the leaflet, the experimenter further explained that the clinical use of DNS is to increase sensations, e.g., to treat numbness that might occur after a surgery or an accident, and that the purpose of the current study is to investigate whether there is a difference in pain sensitivity between patients with fibromyalgia and healthy participants. The real purpose of the experiment, i.e., the investigation of nocebo effects, was not disclosed until the end of the study to not bias any pain-related expectations. A sham calibration procedure followed, where the intention was not to actually calibrate the DNS device but to demonstrate how electrical signals feel on the skin to increase the believability of DNS device function. After cleaning the skin with alcohol, two electrodes were attached vertical to each other on the radial side of the forearm of the non-dominant hand. While the experimenter slowly increased the electrical intensity, participants were asked to indicate the moment that they just noticed a sensation, which was told to be the intensity they would eventually receive throughout the experiment. In reality, all electrical activity stopped after 1 minute. A flashing light allowed the DNS device to appear as if it was still working.

Nocebo Conditioning with Verbal Suggestions

Nocebo effects on pressure pain were induced through conditioning and verbal suggestions using the DNS device. Participants were instructed that when the DNS device was activated, the text "DNS on" would appear on the computer screen, signaling that the device would increase their pain sensitivity, and that the "DNS off" message would appear when the device was deactivated and would not have any influence on their pain sensitivity. DNS on/off messages were presented for 2.5 s using E-prime version 3.0 (Psychology Software Tools, Pittsburg, PA) and were color-coded in either orange or purple, counterbalanced across participants. After the message disappeared, participants received a painful pressure stimulus on their thumbnail for 2.5 s, which was rated on the NRS after each trial with an inter-trial interval of 30 s. The learning phase consisted of 20 trials, where DNS was supposedly activated in half of them. During the experimental trials of the learning phase, i.e., when DNS was supposedly activated, participants received a moderately painful pressure intensity on their thumbnail; during the control trials, i.e., when DNS was supposedly deactivated, they received slight pain. All trials were semi-randomized and not presented more than twice in a row.

Testing phase: Nocebo Effects and Extinction

Directly after nocebo conditioning, a testing phase including extinction took place. The testing phase consisted of 16 experimental (DNS on) and 16 control (DNS off) trials, which were all paired with only slight pain on the thumbnail regardless of the supposed DNS (de) activation, to no longer reinforce the conditioned nocebo effects. After the first 6 testing phase trials, which were used to determine the magnitude of the nocebo effect after nocebo conditioning[11,35,36], a 10-minute break took place. Following this short break, participants were told that the next part of the experiment would be similar to before and that the DNS on/off text would appear signaling DNS (de)activation. No additional verbal suggestions were provided about extinction. Then, the remaining 26 trials ensued. All trials were semi-randomized and not presented more than twice in a row.

The remaining magnitude of nocebo effects after extinction was determined based on the final 6 testing phase trials (3 experimental and 3 control) [11,35,36].

Stability of nocebo effects and extinction across sessions

The same nocebo conditioning and extinction procedures were repeated at one month follow-up. The main difference was that nocebo conditioning and testing phases were preceded by a recall testing phase, to identify the magnitude of nocebo effects recalled after the extinction procedure in the baseline session. The recall testing phase consisted of 6 trials paired with only slight pain, half of which were experimental (DNS on) trials and the other half control trials (DNS off). All trials were semi-randomized and not presented more than twice in a row.

Self-report measures

The Dutch versions of various questionnaires were used to assess participants' clinical and psychological characteristics, which were filled in once before arriving to the first lab session. The Fibromyalgia Survey Questionnaire (FSQ)[31], which is based on the American College of Rheumatology 2010/2011 diagnostic criteria, was filled in by both groups to assess the presence or absence of key symptoms of fibromyalgia. A Fibromyalgia Severity (FS) score was calculated by summing the symptom severity score, ranging between 0-12, and the widespread pain index, ranging between 0-19; a cut-off score of FS ≥12 was considered reliable to satisfy the diagnostic criteria[37]. Those with FS <12 who had already received a fibromyalgia diagnosis were considered to be improving[37].

The Fibromyalgia Impact Questionnaire (FIQ)[38] was filled in by patients to assess their functional disability related to fibromyalgia (Cronbach's α = 0.85). The first item consists of 11 questions on physical functioning, which is scored by taking the mean of all ratings ranging between 0 (always) and 3 (never). The second item assesses how many days

they felt good in the past week, scored inversely between 0 and 7, and the third item assesses how many days of work they missed in the past week, scored between 0 and 7. Items 4-10 assess the severity of various symptoms, ranging between 0 (no impairment) and 10 (maximum impairment). The first 3 scores are subjected to a normalization procedure, after which all scores are averaged and if a patient didn't answer all questions an equalization calculation was employed. The scores range between 0-100, where an average patient scores 50 and higher scores indicate a larger functional disability[38].

The short version of the Depression Anxiety and Stress Scale (DASS-21)[39] was filled in by all participants to assess the negative emotional states of depression, anxiety, and stress subscales (depression subscale Cronbach's α = 0.83; anxiety subscale Cronbach's α = 0.73; stress subscale Cronbach's α = 0.87). The scale consists of 21 statements that are rated between 0 (did not apply to me at all) and 3 (applied to me very much or most of the time). Scores from each subscale are summed and then adjusted to range between 0-42 per subscale for comparability with DASS-42, with higher scores indicating greater symptom severity.

The Life Orientation Test-Revised (LOT-R)[40] was used for assessing dispositional optimism in all participants (Cronbach's α = 0.73). The LOT-R is a 10-item measure consisting of positive, negative, and filler items rated on a 5-point scale between 0 (strongly disagree) and 4 (strongly agree). To calculate optimism, the negative items were reverse coded and then summed with the positive items, resulting in a total score ranging between 0-24, with higher scores indicating higher optimism.

The Pain Catastrophizing Scale (PCS)[41] was used for assessing pain catastrophizing thoughts in all participants (Cronbach's α = 0.91). PCS is a 13-item measure consisting of rumination, magnification, and helplessness subscales, which is rated on a 5-point scale between 0 (not at all) and 4 (all the time). To calculate a PCS score, a sum score of all items was calculated, ranging between 0-52, with higher scores indicating more pain-catastrophizing thoughts.

The Body Vigilance Scale (BVS)[42] was used for assessing participants' attention to bodily sensations (Cronbach's α = 0.93). The first three items in the BVS are directly rated on an 11-point scale between 0 (never) and 10 (always), whereas the fourth item consists of 15 sub-items that are rated separately. To calculate the BVS score, ratings in the fourth item were averaged and afterwards summed with the first three items, ranging between 0-40, with higher scores indicating greater focus on bodily sensations.

The Pearlin Mastery Scale (PMS)[43] was used for assessing the psychological coping resources of all participants based on self-mastery (Cronbach's α = 0.74). The PMS consists of 7 items rated between 1 (strongly disagree) and 5 (strongly agree). Items are summed up, ranging between 7-35, with higher scores indicating greater levels of mastery.

The state scale of State-Trait Anxiety Inventory short-form (STAI-S-6)[44] was used for assessing state anxiety on the day of experimentation in all participants (session 1: Cronbach's α = 0.77; session 2: Cronbach's α = 0.81). The scale consists of 6-items that are rated on a 4-point scale between 1 (not at all) and 4 (very much so). Positive items were reverse coded and then the sum of all items was calculated. Scores were adjusted to range between 20-80 for comparability with STAI-S.

Patients rated their clinical pain and fatigue levels on the day of experimentation using 11-point scales between 0 (no pain/fatigue) and 10 (worst pain/fatigue imaginable), with higher ratings indicating greater symptom severity. Lastly, exit questionnaires were filled in at the end of the study on the perceived aim of the study, perceived effect of DNS on pain sensitivity, trust in the experimenter, perceived competence of the experimenter, and perceived experiment length. The first item required an open-ended answer, whereas the rest of the items were rated on a 0-10 NRS, with higher scores indicating higher intensity. The perceived experiment length was anchored "exactly long enough" around 5/10 on the NRS.

Procedure

Interested individuals were screened for eligibility via a telephone call, which took approximately 10-20 min. A verbal informed consent was obtained prior to screening. If eligible, participants were invited to the lab sessions, 2-2.5 h each, at the Leiden University Treatment and Expertise Center (LUBEC; Leiden, the Netherlands). Before the first lab appointment, participants were asked to fill in an online battery of questionnaires (Qualtrics, Provo, UT) at home taking about 20-30 min, before which they digitally provided an informed consent. After arriving at the lab, explanations were provided about the upcoming experimental procedures and that the study participation was voluntary. After all questions were answered, the experimenter controlled if the participant fulfilled the eligibility criteria for the day of testing, and then the informed consent form was signed. All participants filled in an online questionnaire to assess their current state anxiety levels, where only patients were asked to additionally indicate their current pain and fatigue levels. A brief demonstration of the PEPPA followed, involving practicing the thumb insertion and pain ratings, and then the pressure pain calibration ensued. Next, written and verbal instructions were provided about the DNS device, after which electrodes were attached on participants' arm and the sham calibration of the DNS device

took place. Directly afterwards, after a non-painful practice trial, the nocebo conditioning and testing phases began. When the experiment finished, the experimenter left the room and participants did a 4-minute relaxation task in the form of a guided breathing exercise instructed via headphones to help recover from the potential stress arising from pain administration. At the end of the session, patients were assisted in downloading an app on their phone for rating their daily pain intensity in the coming 3 weeks, which was a procedure pertaining to the larger patient study and will not be addressed in the current paper.

The follow-up lab session took place one month later at LUBEC. The procedure was the same as during the baseline session, with two exceptions. First, the pressure pain calibration was shorter. The pain ratings from the baseline session were used here to replace the first calibration step, i.e., ascending series, since pain thresholds were not expected to change over one-month. However, the remaining calibration steps, i.e., random series and calibration check, still took place to check whether the pressure intensities from the ascending series were successfully rated again within the targeted pain ranges, and if necessary, adjustments were made using the same formulas. Second, the experimental manipulations now began with 6 additional (recall) testing phase trials to measure the magnitude of recalled nocebo effects remaining from the baseline session. After a 5-minute break, participants were orally reminded again about the function of the DNS device, and then the nocebo conditioning and extinction procedures ensued as before with a 10-minute break halfway into the experiment. At the end of the session, participants completed the relaxation task, filled out exit questionnaires, and were reimbursed for their participation.

Statistical Analyses

Data analyses were conducted using the R software environment, version 4.1.0[45]. ANOVA assumptions of normality, homogeneity of variances, and sphericity were checked with QQ plots, Levene's test, and Mauchly's test of sphericity, respectively. When sphericity was violated, either the Greenhouse-Geisser correction (epsilon < 0.75) or the Huynh-Feldt correction (epsilon > 0.75) was considered[46]. Statistical outliers were detected based on z-scores ($z < -3$ or $z > 3$) of the dependent variable. A *p*-value below 0.05 was considered statistically significant unless indicated otherwise. Partial eta-squared (η_p^2) was calculated as the effect size of ANOVA. A partial eta-squared effect size around 0.01 is considered small, 0.06 considered medium, and 0.14 considered large[47]. Cohen's *d* was calculated as the effect size of pairwise t-tests, where 0.2 is considered small, 0.5 considered medium, and 0.8 considered a large effect size[47]. To check whether groups were successfully matched on age and education level, an independent-samples t-test was conducted on the mean age between groups, and a chi-square test was conducted

on the education level (lower vs higher) between groups, respectively. Independent-samples t-tests were used for analyzing between-group differences in calibration intensities, perceived effect of DNS on pain sensitivity, trust in experimenter, perceived competence of the experimenter, and perceived experiment length. Because of multiple comparisons, a Bonferroni correction was applied such that a *p*-value below 0.01 was considered statistically significant.

As a manipulation check, it was examined whether learning has occurred during nocebo conditioning in both sessions. Four paired-sample t-tests were conducted on the mean pain ratings between experimental and control trials during the learning phase of nocebo conditioning in each session to identify whether the associations of "DNS on" with moderate pain and "DNS off" with slight pain were correctly made by each group. Moreover, open-ended answers describing the perceived aim of the study were checked to see whether any participants identified the DNS as a sham device.

To investigate whether nocebo effects were successfully induced during nocebo conditioning in both sessions and whether this induction of nocebo effects differed between groups, a 2 x 2 mixed-design ANOVA was conducted per session, with group (patient vs healthy control) as between-subjects variable and trial type (experimental vs control) as within-subjects variable on the average pain ratings from the first 3 experimental and first 3 control trials of the testing phase. When a significant interaction effect of group by trial type was detected, Bonferroni-corrected pairwise comparisons were applied to more closely examine the manipulation effects between experimental and control trials in each group.

To examine the change in nocebo effects after extinction in both sessions and whether this extinction in nocebo effects differed between groups, a different analysis plan was chosen including difference scores to facilitate the interpretation of findings. A 2 x 2 mixed-design ANOVA was conducted per session, with group (patient vs healthy control) as between-subjects variable and time (nocebo conditioning vs extinction) as within-subjects variable on the difference scores. The difference score after nocebo conditioning was calculated by subtracting the average pain ratings given to the first 3 control trials from the first 3 experimental trials of the testing phase. The difference score after extinction was calculated by subtracting the average pain ratings given to the last 3 control trials from the last 3 experimental trials of the testing phase. The difference score after nocebo conditioning determined the magnitude of nocebo effects, whereas after extinction, it determined the magnitude of nocebo effects remaining after extinction. By comparing the difference scores after nocebo conditioning and after extinction, we investigated whether the magnitude of nocebo effects was significantly lower after

extinction. When a significant interaction effect between group and time was detected, Bonferroni-corrected pairwise comparisons were applied to determine the manipulation effects between nocebo conditioning and extinction on nocebo effects in each group.

To explore the stability of the induction and extinction of nocebo effects across sessions and whether this differed between groups, a 2 x 5 mixed-design ANOVA was conducted with group (patient vs healthy control) as between-subjects variable and time (nocebo conditioning and extinction from sessions 1 and 2, and the recall testing phase from session 2) as within-subjects variable on the difference scores. The difference score after the recall testing phase was calculated by subtracting the average pain ratings given to the 3 control trials from the 3 experimental trials. If a significant interaction effect was detected between group and time, Bonferroni-corrected multiple pairwise comparisons were computed to determine the time level differences in each group. To examine the stability of induction of nocebo effects across sessions, Bonferroni-corrected pairwise comparisons were applied between the time levels 1) nocebo conditioning in sessions 1 and 2, 2) nocebo conditioning in session 1 and the recall testing phase in session 2. To examine the stability of extinction across sessions, Bonferroni-corrected pairwise comparisons were applied between the time levels 1) extinction in sessions 1 and 2, 2) extinction in session 1 and the recall testing phase in session 2. A *p*-value below .025 was considered to indicate a statistically significant lack of stability in the induction or extinction of nocebo effects across sessions.

To allow for the assessment of extinction efficacy for a subgroup of participants who were observed to be susceptible to learning nocebo effects, sensitivity analyses were conducted for the extinction of nocebo effects after removing nocebo non-responders, i.e., participants with difference scores equal to or below zero, from the analyses. The same analyses were subsequently conducted in the subgroup of nocebo responders for the extinction of nocebo effects and for the stability of extinction across sessions. To allow for the assessment of nocebo and extinction efficacy for participants who could be clearly differentiated in their fibromyalgia symptomatology, another set of sensitivity analyses were conducted after removing patients scoring FS <12 or healthy controls scoring FS ≥12 on the FSQ, using the same analyses for the induction and extinction of nocebo effects in both sessions. Additionally, we checked whether the induced nocebo magnitudes were associated across sessions. This was explored with a repeated measures correlation analysis conducted for the magnitude of nocebo effects between two sessions firstly per group and then after pooling both samples. All sensitivity analyses were reported under Supplemental File II.

Lastly, we conducted Pearson's correlation analyses to examine the relation between the magnitude of nocebo effects in session 1 and depression (DASS-21), trait anxiety (DASS-21), stress (DASS-21), optimism (LOT-R), pain-catastrophizing thoughts (PCS), body vigilance (BVS), and mastery (PMS) in both participant groups, as well as fibromyalgia disability (FIQ) in the patient group. Moreover, for each session, we examined the relationship between the magnitude of nocebo effect induction and the state anxiety (STAI-S-6) and pain and fatigue levels (NRS) on the experiment day.

RESULTS

A total of 81 participants were eligible for participating in the experiment (patients $N = 46$; healthy controls $N = 35$). Of these, 8 participants (6 patients, 2 healthy controls) were excluded during the first session due to problems with pressure pain calibration (i.e., pain ratings were lower than the required pain ranges for moderate and slight pain) and 1 patient dropped out due to misunderstanding the instructions for rating pain intensity. During the second session, 4 participants (1 patient, 3 healthy controls) were excluded due to problems with pressure pain calibration and 8 participants (7 patients, 1 healthy control) dropped out for personal reasons (e.g., scheduling issues, testing positive for COVID-19). Moreover, due to technical and software-related problems, data could not be retrieved from 3 participants (2 patients, 1 healthy control) in session 1 and from another 3 participants (2 patients, 1 healthy control) in session 2. Considering that 28% of patients and 9% of healthy controls dropped out after the baseline session, a total of 69 participants (37 patients and 32 healthy controls) were included in session 1 to also reach a minimum sample size in the follow-up session, which resulted in a total inclusion of 56 participants in session 2 (patients $N = 29$; healthy controls $N = 27$). All included participants per session were considered for statistical analyses.

Descriptive statistics

Table 1 displays the demographic and health-related characteristics of the sample, and Table 2 displays the group means and SDs from psychological characteristics and exit questionnaires. The fibromyalgia severity score in the patient group was between 6 and 26, where 3 patients had scores <12, indicating that they might be in a recovery period; all patients were considered for the main analyses. In the healthy control group, scores ranged between 0-9, where no healthy participant reached the cut-off score. There were no significant group differences in the mean age ($t(67) = 1.64$, $p = .11$) or the education level ($X^2(1) = .31$, $p = .58$) of participants, suggesting a successful group matching. Table 3 displays the means and SDs of calibration values (kgf/cm^2). No significant group differences were observed for the calibration values of slight and moderate pressure

pain intensities in session 1 (slight pain: $t(67) = 1.98$, $p = .053$; moderate pain: $t(67) = 1.98$, $p = .051$) and session 2 (slight pain: $t(54) = .70$, $p = .51$; moderate pain: $t(54) = .92$, $p = .36$). Neither were there any group differences in the perceived effect of DNS on pain sensitivity ($t(59) = 1.38$, $p = .17$), trust in the experimenter ($t(59) = .76$, $p = .45$), perceived competence of the experimenter ($t(59) = .17$, $p = .87$), or the perceived experiment length ($t(59) = 1.35$, $p = .18$). Moreover, Table 3 displays the overall mean pain intensity ratings and Figure 3A and Figure 3B display the trial-by-trial change in mean pain intensity ratings across sessions. Note that in Figures 3A and 3B, an upward trend can be observed in the horizontal lines, which is also reflected in Table 3 by an increase in pain ratings during the extinction phase, both of which potentially illustrating a pain sensitization across trials.

Table 1

Demographic and health-related characteristics of female participants in the study.

	Session 1		Session 2	
Characteristics	Patient (*N* = 37)	Healthy control (*N* = 32)	Patient (*N* = 29)	Healthy control (*N* = 27)
Age (years) [mean (SD)]	37.81(10.47)	33.56(10.97)	34.21(9.96)	33.78(11.31)
Higher education Level [n (%)]	28(76)	26(81)	21(72)	21(78)
Partner [n (%)]	32(87)	20(63)	25(86)	16(59)
Work status [n (%)]				
Student	13(35)	18(56)	13(45)	15(56)
Employed	34(92)	32(100)	27(93)	27(100)
Work (hours per week) [mean (SD)]	24.96(9.99)	26.27(11.67)	26.32(10.11)	27.63(10.74)
Unemployed	1(3)	1(3)	1(3)	1(4)
Volunteer work	9(24)	13(41)	10(34)	11(41)
Run household	16(43)	3(9)	13(48)	3(11)
Disability pension	7(19)	1(3)	4(15)	1(4)
Retired	0	1(3)	0	1(4)
Fibromyalgia Severity (FSQ) [median (IQR)]	18(8)	4(3)	17(9)	4(3)
Fibromyalgia Disability (FIQ) [mean (SD)]	40.95(13.48)	-	41.06(14.04)	-
Fibromyalgia complaints (years) [mean (SD)]	14.51(9.81)	-	14.32(8.59)	-
Fibromyalgia diagnosis (years) [mean (SD)]	6.59(6.16)	-	5.58(3.49)	-

Table 2

Group means and SDs for psychological characteristics and exit questionnaires.

	Patient	Healthy control
Characteristics	*Mean (SD)*	
Depression	7.73(6.83)	1.63(2.98)
Anxiety	5.46(5.07)	1.31(2.25)
Stress	14.11(7.53)	4.81(4.28)
Dispositional Optimism	15.73(3.25)	16.78(2.69)
Pain Catastrophizing	12.27(7.66)	7.28(7.63)
Body Vigilance	13.65(7.16)	10.97(6.18)
Self-Mastery	26.16(3.88)	27.91(3.14)
State Anxiety prior to testing during session 1	34.78(7.19)	29.53(7.67)
State Anxiety prior to testing during session 2	35.64(9.57)	28.35(7.49)
Pain prior to testing during session 1	4.32(1.87)	-
Pain prior to testing during session 2	4.11(2.03)	-
Fatigue prior to testing during session 1	4.59(2.05)	-
Fatigue prior to testing during session 2	4.61(1.91)	-
Perceived DNS effect on pain sensitivity	4.74(2.33)	3.74(2.80)
Trust in experimenter	9.00(1.00)	9.30(0.87)
Perceived competence of experimenter	9.04(0.90)	9.11(1.22)
Perceived length of study	5.67(1.04)	6.11(1.28)

Note. Total sample size for trait and state characteristics in session 1 was 69 (patient *N*=37; healthy control *N*=32) whereas for state characteristics in session 2 and exit questionnaires was 56 (patient *N*=29; healthy control *N*=27).

Table 3

Group means and SDs for pressure intensity levels (calibration) and pain intensity ratings (recall testing phase, nocebo conditioning, extinction) across sessions.

	Session 1		Session 2	
	Patient (*N*=37)	Healthy Control (*N*=32)	Patient (*N*=29)	Healthy Control (*N*=27)
Calibration (kgf/cm2)		*Mean (SD)*		
Slight Pain	4.66(1.90)	5.59(2.00)	5.24(2.28)	5.64(2.25)
Moderate Pain	6.76(2.72)	8.05(2.67)	7.14(2.65)	7.81(2.86)
Recall Testing Phase (0-10 NRS)				
Experimental Trials	-	-	2.73(0.97)	2.77(0.94)
Control Trials	-	-	2.59(0.99)	2.67(1.12)
Difference Score	-	-	0.14(0.67)	0.10(0.55)
Nocebo Conditioning (0-10 NRS)*				
Learning Phase (20 Trials)				
Experimental Trials	5.71(1.37)	5.70(0.90)	5.70(1.22)	5.49(1.22)
Control Trials	3.69(1.41)	3.52(1.05)	3.29(1.35)	3.23(1.20)
Difference Score	2.02(0.50)	2.17(1.03)	2.41(1.11)	2.25(1.13)

Table 3

Continued.

	Session 1		Session 2	
	Patient (*N*=37)	Healthy Control (*N*=32)	Patient (*N*=29)	Healthy Control (*N*=27)
Testing Phase (Trials 1-6)				
Experimental Trials	4.17(1.86)	4.30(1.48)	4.06(1.77)	3.77(1.55)
Control Trials	3.98(1.82)	3.69(1.53)	3.72(1.77)	3.29(1.32)
Difference score	0.19(0.74)	0.61(0.62)	0.34(0.91)	0.48(0.72)
*Extinction** (0-10 NRS)*				
Testing Phase (Trials 26-32)				
Experimental Trials	5.09(1.96)	4.65(1.83)	4.97(1.88)	4.43(1.90)
Control Trials	4.80(1.96)	4.44(1.79)	4.67(1.79)	3.94(1.60)
Difference score	0.29(0.60)	0.21(0.55)	0.29(0.76)	0.49(0.66)

Note for session 1: *patient sample excluding 1 outlier (*N*= 36); **patient sample excluding 2 outliers (*N* = 35)

Assumption checks

The ANOVA assumptions of normality and homogeneity of variances were not violated. In cases where Mauchly's test of sphericity was violated, corrections were made on the degrees of freedom. Notably, two patients were detected as statistical outliers based on the difference scores after nocebo conditioning (z = 4.04) or after extinction (z = -3.72) in session 1. Given the extremity of these statistical outliers and since they had a significant impact on the study findings, they were considered not representative of the sample and were therefore excluded from the corresponding analyses relating to session 1. For more detailed results including these outliers see Supplemental File II. No statistical outliers were detected based on data from session 2.

Manipulation check

Results from the paired-samples t-tests showed that learning had successfully occurred during the learning phase of nocebo conditioning in both sessions for patients (session 1 $t(36) = 14.43$, $p < .001$; session 2 $t(28) = 11.71$, $p < .001$) and healthy controls (session 1 $t(31) = 11.92$, $p < .001$; session 2 $t(26) = 10.35$, $p < .001$). Factors such as having prior knowledge of, or experience with, a TENS device, and in case of experience finding it effective, did not have any significant impact on the magnitude of nocebo effects in either session (for more details see Supplemental File III). Moreover, open-ended answers given to the perceived aim of the study was aligned with the information provided about the study, where no participants suspected that the DNS device was never activated.

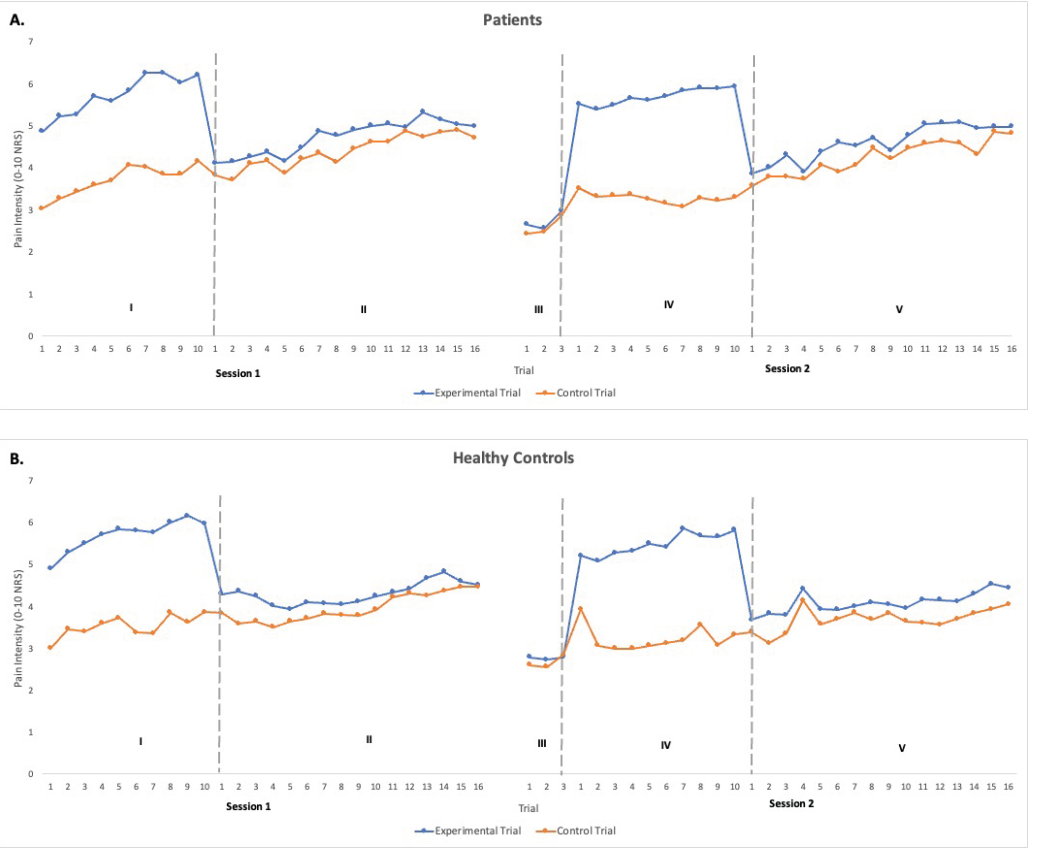


Figure 3. Mean pain intensity ratings across all trials in sessions 1 and 2 in the patient group excluding outliers **(A)** and the healthy control group **(B)**. Experimental and control trials are represented in separate lines. Section I: Trials in the learning phase of nocebo conditioning; Section II: Trials in the testing phase; Section III: Trials in the recall testing phase; Section IV: Trials in the learning phase of nocebo conditioning; Section V: Trials in the testing phase.

Induction of nocebo effects in session 1

A 2 x 2 mixed-design ANOVA showed a significant interaction effect between group and trial type in session 1 ($F(1,66) = 6.36$, $p = .01$, $\eta_p^2 = .08$) and a main effect of trial type ($F(1,66) = 23.43$, $p < .001$, $\eta_p^2 = .27$), but no main effect of group ($F(1,66) = .04$, $p = .84$, $\eta_p^2 < .01$). Bonferroni-corrected pairwise comparisons between the trial type levels at each group showed that the mean pain ratings in experimental trials were significantly higher than control trials in the healthy control group ($p < .001$, $d = .41$). In the patient group, however, the mean pain ratings were not significantly higher in the experimental trials compared to control trials ($p = .13$, $d = .10$). Figure 4 displays the magnitude of induced nocebo effects, across sessions and groups.

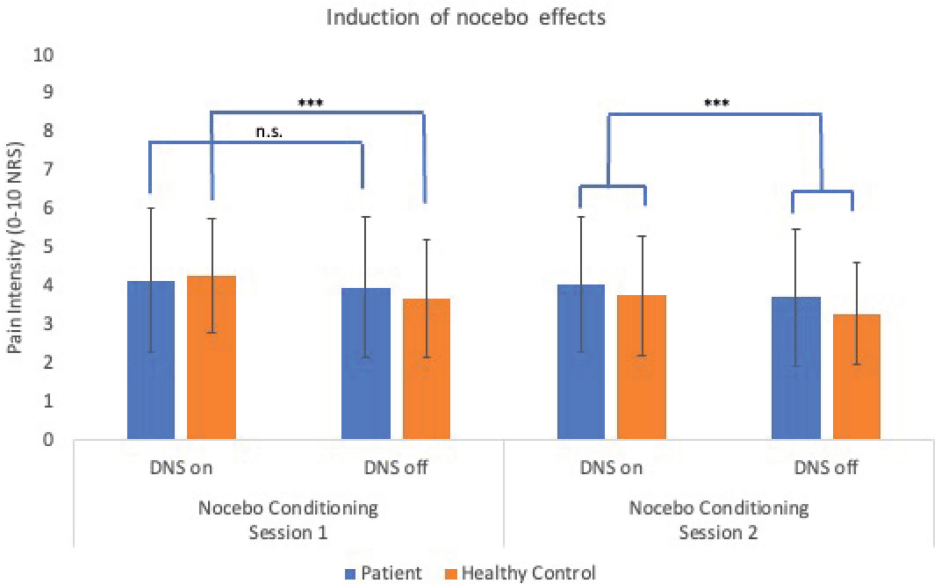


Figure 4. Mean pain intensities from the first 3 experimental (DNS on) and first 3 control (DNS off) trials of the testing phase across groups and sessions. Sample sizes per group are excluding the outliers. If a group x trial type interaction was found, significance levels were presented between groups. If only a main effect of trial type was found, significance levels were presented across groups. Error bars indicate ± SE. ***: $p < .001$; n.s.: not significant.

Extinction of nocebo effects in session 1

A 2 x 2 mixed-design ANOVA showed a significant interaction effect between group and time in session 1 ($F(1,65) = 10.35$, $p = .02$, $\eta_p^2 = .14$), but no main effect of time ($F(1,65) = 2.72$, $p = .10$, $\eta_p^2 = .04$) nor a main effect of group ($F(1,65) = 2.07$, $p = .15$, $\eta_p^2 = .031$). Bonferroni-corrected pairwise comparisons between time levels at each group showed that the mean difference score was significantly lower after extinction compared to nocebo conditioning in the healthy control group, indicating a significant

decrease in nocebo effects ($p < .001$, $d = .68$). In the patient group, the mean difference score was higher after extinction compared to nocebo conditioning; however, this was not significant ($p = .34$, $d = .20$). Figure 5 displays the magnitude of nocebo decrease after extinction, across sessions and groups.

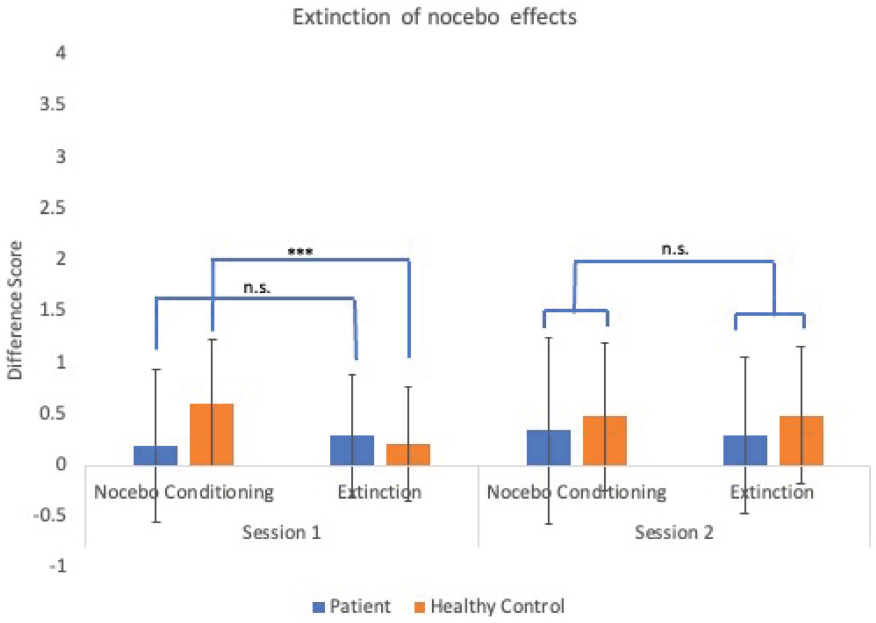


Figure 5. Difference scores based on the first 6 trials (nocebo conditioning) and last 6 trials (extinction) of the testing phase across groups and sessions. Sample size per experimental manipulation consists of all participants in a given session excluding the outliers. If a group x time interaction was found, significance levels were presented between groups. If only a main effect of time was found, significance levels were presented across groups. Error bars indicate ± SE. ***: $p < .001$; n.s.: not significant.

Induction and extinction of nocebo effects in session 2

For the induction of nocebo effects in session 2, the 2 x 2 mixed-design ANOVA showed that there was no interaction effect ($F(1,54) = .41$, $p = .52$, $\eta_p^2 = .01$) nor a main effect of group ($F(1,54) = .75$, $p = .39$, $\eta_p^2 = .01$), but there was a significant main effect of trial type ($F(1,54) = 13.85$, $p < .001$, $\eta_p^2 = .20$), where experimental trials ($M = 3.91$, $SE = .22$) were rated significantly higher than the control trials ($M = 3.50$, $SE = .21$), indicating that nocebo effects were induced across groups. Since this overall finding did not align with the nocebo results from session 1, post-hoc analyses were conducted to get a better insight into the potential group differences in nocebo induction in session 2. Pairwise comparisons of trial type levels at each group showed that the mean pain ratings were significantly higher in experimental trials compared to control trials in the healthy control group ($p = .002$, $d = .33$); however, not significantly higher in the patient group ($p = .054$, $d = .19$).

Moreover, for the extinction of nocebo effects in session 2, the 2 x 2 mixed-design ANOVA showed that there was no interaction effect ($F(1,54) = .06$, $p = .81$, $\eta_p^2 = .001$), nor a main effect of group ($F(1,54) = .92$, $p = .34$, $\eta_p^2 = .02$), or time $F(1,54) = .03$, $p = .87$, $\eta_p^2 = .001$), giving no indication for extinction of nocebo effects across groups.

The stability of nocebo effects and of extinction across sessions 1 and 2

Figure 6 displays the fluctuations in difference scores across all experimental manipulations, with patients showing a relatively more stable trend and lower nocebo effects compared to the healthy control group. The 5 x 2 mixed-design ANOVA showed there was no significant interaction effect ($F(3.59,179.64) = 1.95$, $p = .11$, $\eta_p^2 = .04$) nor a main effect of group ($F(1,50) = 2.25$, $p = .14$, $\eta_p^2 = .04$), but there was a significant main effect of time ($F(3.59,179.64) = 2.54$, $p = .048$, $\eta_p^2 = .05$). Pairwise comparison of the time levels showed that the magnitude of nocebo effects was significantly higher ($M = .33$, $SE = .13$) after nocebo conditioning in session 1 compared to the recall testing phase in session 2 ($p = .01$, $d = .44$). Compared to session 2, the magnitude of nocebo effects after nocebo conditioning in session 1 was not statistically different ($p = .98$, $d = .03$). This indicates that the magnitude of nocebo effects induced during the baseline session were significantly decreased at one-month follow-up, and that the efficacy of the nocebo conditioning paradigm did not significantly differ between sessions. Moreover, pairwise comparisons showed that the difference score after extinction in session 1 did not significantly differ from the difference score after the recall testing phase in session 2 ($p = .05$, $d = .26$). Also, the difference scores after extinction in session 1 and 2 were not significantly different ($p = .65$, $d = .22$). This indicates that the magnitude of nocebo effects observed after extinction at baseline was not different from the magnitude of nocebo effects recalled after one-month, and that the efficacy of the extinction paradigm did not statistically differ between sessions.

Questionnaires

Pearson's correlation analyses indicated that there was no significant relation between the magnitude of nocebo effects during session 1 and each of the nine questionnaire scores (FIQ: $r = -.05$, $p = .79$; DASS depression: $r = -.08$, $p = .49$; DASS anxiety: $r = -.09$, $p = .42$; DASS stress: $r = -.03$, $p = .81$; BVS: $r = .03$, $p = .81$; PCS: $r = -.04$, $p = .78$; PMS: $r = .13$, $p = .28$, LOT-R: $r = .06$, $p = .64$; state anxiety session 1: $r = -.09$, $p = .45$; state anxiety session 2: $r = .06$, $p = .65$). Moreover, patients' pain and fatigue levels on the experiment day were not significantly related to the magnitude of nocebo effects (pain session 1: $r = .20$, $p = .24$; pain session 2: $r = -.08$, $p = .69$ fatigue session 1: $r = .02$, $p = .93$; fatigue session 2: $r = .09$, $p = .66$).

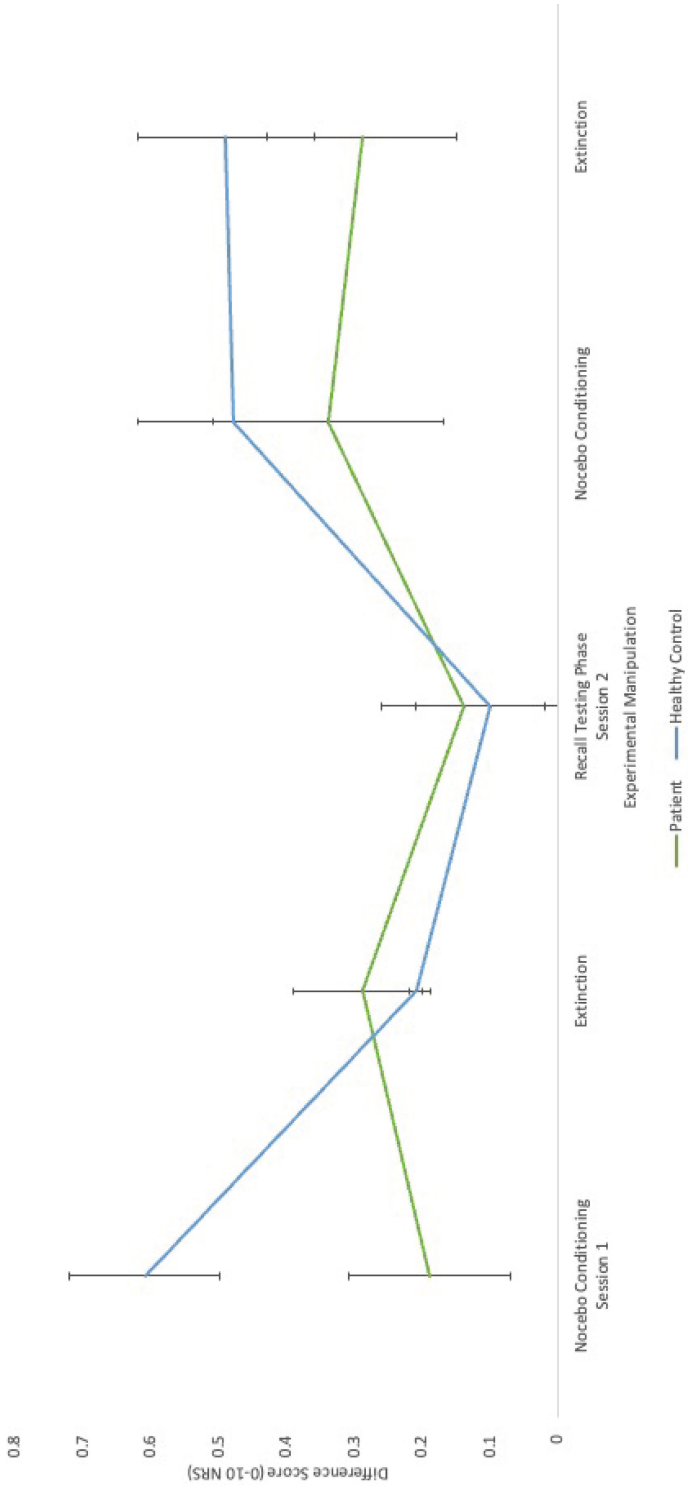


Figure 6. Difference scores per experimental manipulation across groups and sessions. Sample size per experimental manipulation consist of all participants in a given session excluding the outliers. Participant groups are represented in separate lines and the error bars indicate ± SE.

DISCUSSION

The current study investigated potential group differences in inducing and decreasing nocebo effects on experimental pressure pain in female patients with fibromyalgia and matched healthy controls. Additionally, the stability of nocebo effects at a one-month follow-up was examined. Nocebo effects on pressure pain were experimentally induced through classical conditioning with verbal suggestions, and were decreased via extinction. Our results suggest that nocebo effects were induced in the healthy control group, but not in the patient group during the baseline session, although this group difference was not robust. Nocebo effects decreased in the healthy control group after extinction. During the follow-up session, nocebo effects were induced across both groups; however, insights from post-hoc investigations suggest that this effect was primarily observable in the healthy control group, generally aligning with our results from the baseline session. However, unlike the baseline session, extinction was not observed in either group. Moreover, across all participants, the magnitude of nocebo induction and decrease appeared stable over 1-month, although note that only less than half of participants qualified as nocebo responders in both sessions. Contrary to our hypotheses, we did not find stronger nocebo effects, or more resistance to extinction, in the patient group compared to healthy controls. Instead, patients with fibromyalgia might be less responsive towards the experimental manipulation of nocebo effects than healthy controls.

Current literature on the experimental investigation of nocebo effects is largely based on findings from healthy samples[13], with only a number of studies focusing on patients with acute post-operative pain[7] or with chronic pain complaints from irritable bowel syndrome[8]. In these studies, nocebo effects were induced by providing verbal suggestions about the pain-increasing function of a placebo agent[7,8]. The role of classical conditioning in inducing nocebo effects in chronic pain conditions is far less researched[10]. In healthy participants on the other hand, the nocebo conditioning paradigm has been found to successfully induce nocebo effects on a variety of pain modalities, such as heat, electrical, and pressure pain[5,11,12]. In line with previous research, we found that nocebo effects were induced on pressure pain in the healthy control group in both sessions; however, our findings in the patient group were somewhat elusive. Nocebo effects were observed in the patient group only during the follow-up session. However, when including one patient who had an unlikely large nocebo score (i.e., an outlier), significant nocebo effects were observed during baseline, and group difference at 1-month follow-up were not clear. Thus, the group differences found in the current study were not robust. Additionally, it was observed that a lower percentage of patients than healthy controls were nocebo responders in each session. Thus, the current data suggests that patients with fibromyalgia either could be equally or less

responsive to the experimental manipulation of nocebo effects compared to healthy controls. Future studies might consider including equivalence testing or Bayesian statistics in their methodology to better establish whether group differences were not observable.

But how do these findings align with previous literature which suggests that patients could be at risk of developing nocebo effects[1,6,21,48,49]? One methodological explanation could be that the experimental pain intensities administered in the current study may not have been high enough to induce fear in patients, as higher pain intensities have been found to be associated with larger nocebo hyperalgesia, mediated through fear[50]. Patients' ongoing pain experiences in real-life might have been more intense than our administered pressure pain intensities, which might have led patients to experience less fear of pain during nocebo manipulations compared to healthy controls.

Another potential explanation could be related to group differences in pain-reporting variability. A recent study in patients with osteoarthritis of the knee has shown that accuracy in experimental pain-reporting correlates negatively with responsiveness to a placebo[51]. The implication of this finding is that the ability to direct one's attention inwardly, rather than externally, could be related to being able to resist external cues that contribute to placebo responses, and thereby lead to more accurate reporting of pain experiences due to a higher awareness of bodily sensations[51,52]. We did not assess this in the current study, but we speculate that patients' attention towards pain might have been more inwardly-directed compared to healthy controls, potentially due to their ongoing pain experiences in daily-life which might affect the salience networks in the brain[53]. If so, patients might have been less influenced by the sham activation of the TENS device, i.e., the external (placebo/nocebo) cue. However, preliminary findings e.g., on a heartbeat perception task, have shown a reduced awareness in fibromyalgia patients compared to healthy controls[54,55]. Thus, further research is warranted on the interoceptive awareness of pain and attention to placebo/nocebo cues in fibromyalgia.

Moreover, patients with fibromyalgia have been previously found to suffer from contingency learning deficits where safety cues in the environment could not be distinctly identified[22,24]. Potentially, the inability to identify safe pain cues from unsafe ones may have implications for the strength of nocebo hyperalgesia induction, although the current data is insufficient to support this argument. To get a better insight into whether the US-CS contingency awareness plays a role in nocebo learning, future research could consider including additional measurements of contingency awareness between the experimental and control (i.e., safety) cues during the testing phase. This could be useful in identifying whether the ability to learn the predictive cues in the environment (contingency learning) intersects with expectations of adverse treatment outcomes (nocebo effects).

The same experimental procedures were repeated at follow-up. The overall magnitudes of nocebo effects and their extinction did not statistically differ across sessions. However, group differences observed during baseline were no longer clearly present during follow-up, which could be potentially explained by two things. Firstly, nocebo learning might have been more strongly reinforced in patients than healthy controls after repeating the experimental procedure for a second time. Secondly, due to drop-outs, a smaller sample was included in the follow-up analyses than in the baseline analyses, which might have influenced the group effects in the follow-up session. A closer look into the recall testing phase tells us that the magnitude of nocebo effects recalled after one month was comparable to the magnitude of effects remaining after the extinction procedure during the baseline session. The passing of one month probably had no additional influence on the further extinction of nocebo effects. Also, no spontaneous recovery[56], i.e., return of nocebo effects, was detected during recall testing phase. Although the inclusion of the recall-testing phase was necessary in the study design, its potential interference on the subsequent nocebo conditioning procedure cannot be ruled out; nevertheless, our manipulation check indicates that participants did not detect any discrepancy in the DNS device function throughout the experiment and regular breaks were included to reduce any contrast between procedures. A study limitation was that our conclusions on the stability of nocebo induction could not be based on a pure comparison between the nocebo induction procedures in both sessions, as the potential influence of additional procedures which took place in between, i.e., extinction procedure during baseline and recall testing phase, cannot be overlooked. Future studies might consider including a control group without these additional manipulations to purely examine the role of follow-up period length on nocebo stability. Also, longer follow-up periods might present different outcomes in stability, especially if disease progression also occurs on the side.

As a study limitation, the potential influence of floor effects due to small nocebo scores cannot be ruled out entirely. The generalizability of our findings using the nocebo conditioning paradigm on pressure pain requires further replication in healthy and chronic pain populations. Moreover, the pain sensitization observed in the current study was unique, and this issue has not been raised previously in nocebo studies using pressure pain or other pain modalities[5,11,57]. During extinction, an overall increase in pain ratings was observed as a result of pain sensitization; our sensitivity analyses suggest that extinction took place once the nocebo effects were induced in either group. Considering that conditioned nocebo responses are common in clinical practice[6], future research is recommended to take these points into consideration when designing nocebo studies in chronic pain conditions.

To conclude, the current study is first to investigate group differences in conditioned nocebo effects in patients with a chronic pain condition and healthy controls. Contrary to our expectations, we did not find stronger nocebo effects on pressure pain in patients with fibromyalgia compared to healthy controls, if anything, patients might be less, or potentially equally, responsive to the experimental manipulation of nocebo effects as compared to healthy controls. This finding could be related to the current methodological limitations as well as the potential learning differences in patients. Moreover, the overall magnitudes of nocebo effects and their extinction were stable over 1-month. Considering that conditioned nocebo responses are common in clinical settings, further investigation of nocebo effects is essential to minimize their detrimental role during treatment.

Acknowledgements

The authors would like to thank Elio Sjak-Shie, Evert Dekker, and Iris Spruit from the technical support team of Leiden University (SOLO) and Michael Boonekamp from the Design & Prototyping team of Leiden University Medical Center for their incredible support with designing, manufacturing, and programming the pressure pain device components. The authors would like to also thank Ruben Sars, Suzanne Derksen, Suzie Wesselman, Joy Franke, Shannen Pieper, Sofie Schutte, Sarah Vreijling, and Fabienne van den Ende for their exceptional support during the procedural design and/or the data collection parts of this experiment. Lastly, the authors thank our patient partners for their valuable insights during the study design.

REFERENCES CHAPTER 4

[1] A.W.M. Evers, L. Colloca, C. Blease, M. Annoni, L.Y. Atlas, F. Benedetti, U. Bingel, C. Büchel, C. Carvalho, B. Colagiuri, A.J. Crum, P. Enck, J. Gaab, A.L. Geers, J. Howick, K.B. Jensen, I. Kirsch, K. Meissner, V. Napadow, K.J. Peerdeman, A. Raz, W. Rief, L. Vase, T.D. Wager, B.E. Wampold, K. Weimer, K. Wiech, T.J. Kaptchuk, R. Klinger, J.M. Kelley, Implications of placebo and nocebo effects for clinical practice: Expert consensus, Psychother. Psychosom. 87 (2018) 204–210. https://doi.org/10.1159/000490354.

[2] A.J. Barsky, R. Saintfort, M.P. Rogers, J.F. Borus, Nonspecific medication side effects and the nocebo phenomenon, Jama. 287 (2002) 622–627.

[3] F. Benedetti, M. Amanzio, S. Vighetti, G. Asteggiano, F.G. Miller, L. Colloca, R.A. Crouch, T.J. Kaptchuk, The biochemical and neuroendocrine bases of the hyperalgesic nocebo effect, Placebo A Reader. (2013) 176–185. http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-38719-028&site=ehost-live.

[4] V.F. Quinn, B. Colagiuri, Using Learning Strategies to Inhibit the Nocebo Effect, first ed., Elsevier Inc., 2018. https://doi.org/10.1016/bs.irn.2018.01.011.

[5] M.A. Thomaidou, D.S. Veldhuijzen, K.J. Peerdeman, N.Z.S. Wiebing, J.S. Blythe, A.W.M. Evers, Learning mechanisms in nocebo hyperalgesia: the role of conditioning and extinction processes, Pain. 161 (2020) 1597–1608. https://doi.org/10.1097/j.pain.0000000000001861.

[6] L. Colloca, F.G. Miller, The Nocebo Effect and Its Relevance for Clinical Practice, Psychosom. Med. 73 (2011) 598–603. https://doi.org/10.1097/PSY.0b013e3182294a50.

[7] F. Benedetti, M. Amanzio, C. Casadio, A. Oliaro, G. Maggi, Blockade of nocebo hyperalgesia by the cholecystokinin antagonist proglumide, Pain. 71 (1997) 135–140.

[8] L. Vase, M.E. Robinson, G.N. Verne, D.D. Price, The contributions of suggestion, desire, and expectation to placebo effects in irritable bowel syndrome patients. An empirical investigation, Pain. 105 (2003) 17–25.

[9] F. Benedetti, A. Pollo, L. Lopiano, M. Lanotte, S. Vighetti, I. Rainero, Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/nocebo responses, J Neurosci. 23 (2003) 4315–4323. https://doi.org/10.1523/jneurosci.23-10-04315.2003.

[10] J. Schmitz, M. Muller, J. Stork, I. Eichler, C. Zollner, H. Flor, R. Klinger, Positive Treatment Expectancies Reduce Clinical Pain and Perceived Limitations in Movement Ability Despite Increased Experimental Pain: A Randomized Controlled Trial on Sham Opioid Infusion in Patients with Chronic Back Pain, Psychother Psychosom. 88 (2019) 203–214.

[11] S. Meijer, M. Karacaoglu, H. van Middendorp, D.S. Veldhuijzen, K.J. Peerdeman, A.W.M.M. Evers, K.B. Jensen, K.J. Peerdeman, A.W.M.M. Evers, Efficacy of open-label counterconditioning for reducing nocebo effects on pressure pain, Eur. J. Pain. 27 (2023) 831-847. https://doi.org/https://doi.org/10.1002/ejp.2112.

[12] B. Colagiuri, V.F. Quinn, L. Colloca, Nocebo Hyperalgesia, Partial Reinforcement, and Extinction, J Pain. 16 (2015) 995–1004. https://doi.org/10.1016/j.jpain.2015.06.012.

[13] G.L. Petersen, N.B. Finnerup, L. Colloca, M. Amanzio, D.D. Price, T.S. Jensen, L. Vase, The magnitude of nocebo effects in pain: A meta-analysis, Pain. 155 (2014) 1426–1434. https://doi.org/10.1016/j.pain.2014.04.016.

[14] P. Sarzi-Puttini, V. Giorgi, D. Marotto, F. Atzeni, Fibromyalgia: an update on clinical characteristics, aetiopathogenesis and treatment, Nat. Rev. Rheumatol. 16 (2020) 645–660. https://doi.org/10.1038/s41584-020-00506-w.

[15] E. Briones-Vozmediano, C. Vives-Cases, E. Ronda-Pérez, D. Gil-González, Patients' and professionals' views on managing fibromyalgia, Pain Res.

Manag. 18 (2013) 19–24. https://doi.org/10.1155/2013/742510.
[16] M.M. Cunningham, C. Jillings, Individuals' descriptions of living with fibromyalgia, Clin. Nurs. Res. 15 (2006) 258–273. https://doi.org/10.1177/1054773806291853.
[17] K.J. Peerdeman, A.I.M. van Laarhoven, M.L. Peters, A.W.M. Evers, An integrative review of the influence of expectancies on pain, Front. Psychol. 7 (2016) 1–7. https://doi.org/10.3389/fpsyg.2016.01270.
[18] M. Greville-Harris, P. Dieppe, Bad is more powerful than good: The nocebo response in medical consultations, Am. J. Med. 128 (2015) 126–129. https://doi.org/10.1016/j.amjmed.2014.08.031.
[19] M. Ingvar, Learning mechanisms in pain chronification-teachings from placebo research, Pain. 156 (2015) S18–S23. https://doi.org/10.1097/j.pain.0000000000000093.
[20] F. Benedetti, E. Frisaldi, D. Barbiani, E. Camerone, A. Shaibani, Nocebo and the contribution of psychosocial factors to the generation of pain, J. Neural Transm. 127 (2020) 687–696. https://doi.org/10.1007/s00702-019-02104-x.
[21] M. Manai, H. van Middendorp, D.S. Veldhuijzen, T.W.J. Huizinga, A.W.M. Evers, How to prevent, minimize, or extinguish nocebo effects in pain: a narrative review on mechanisms, predictors, and interventions, Pain Rep. 4 (2019) e699. https://doi.org/10.1097/PR9.0000000000000699.
[22] J. Jenewein, H. Moergeli, H. Sprott, D. Honegger, L. Brunner, D. Ettlin, C. Grillon, K. Bloch, M. Brügger, K. Schwegler, S. Schumacher, G. Hasler, Fear-learning deficits in subjects with fibromyalgia syndrome?, Eur. J. Pain. 17 (2013) 1374–1384. https://doi.org/10.1002/j.1532-2149.2013.00300.x.
[23] A. Meulders, P.A. Karsdorp, N. Claes, J.W.S. Vlaeyen, Comparing Counterconditioning and Extinction as Methods to Reduce Fear of Movement-Related Pain, J. Pain. 16 (2015) 1353–1365. https://doi.org/10.1016/j.jpain.2015.09.007.
[24] A. Meulders, Y. Boddez, F. Blanco, M. Van Den Houte, J.W.S. Vlaeyen, Reduced selective learning in patients with fibromyalgia vs healthy controls, Pain. 159 (2018) 1268–1276. https://doi.org/10.1097/j.pain.0000000000001207.
[25] A. Meulders, M. Meulders, I. Stouten, J. De Bie, J.W.S. Vlaeyen, Extinction of Fear Generalization: A Comparison Between Fibromyalgia Patients and Healthy Control Participants, J. Pain. 18 (2017) 79–95. https://doi.org/10.1016/j.jpain.2016.10.004.
[26] K. Faasse, A. Huynh, S. Pearson, A.L. Geers, S.G. Helfer, B. Colagiuri, The Influence of Side Effect Information Framing on Nocebo Effects, Ann Behav Med. 53 (2019) 621–629.
[27] F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behav. Res. Methods. 39 (2007) 175–191. https://doi.org/10.3758/BF03193146.
[28] P. Cuijpers, E.H. Turner, S.L. Koole, A. Van Dijke, F. Smit, What is the threshold for a clinically relevant effect? The case of major depressive disorders, Depress. Anxiety. 31 (2014) 374–378.
[29] M.B. Yunus, The role of gender in fibromyalgia syndrome, Curr. Rheumatol. Rep. 3 (2001) 128–134.
[30] F. Verhage, Intelligence and age in a Dutch sample, Hum. Dev. 8 (1965) 238–245.
[31] F. Wolfe, D.J. Clauw, M.A. Fitzcharles, D.L. Goldenberg, W. Häuser, R.S. Katz, P. Mease, A.S. Russell, I.J. Russell, J.B. Winfield, Fibromyalgia criteria and severity scales for clinical and epidemiological studies: A modification of the ACR preliminary diagnostic criteria for fibromyalgia, J. Rheumatol. 38 (2011) 1113–1122. https://doi.org/10.3899/jrheum.100594.
[32] F. Petzke, D.J. Clauw, K. Ambrose, A. Khine, R.H. Gracely, Increased pain sensitivity in fibromyalgia: Effects of stimulus type and mode of presentation, Pain. 105 (2003) 403–413. https://doi.org/10.1016/S0304-3959(03)00204-5.
[33] K.B. Jensen, E. Kosek, F. Petzke, S. Carville, P. Fransson, H. Marcus, S.C.R. Williams, E. Choy, T. Giesecke, Y. Mainguy, R. Gracely, M. Ingvar, Evidence of dysfunctional

pain inhibition in Fibromyalgia reflected in rACC during provoked pain, Pain. 144 (2009) 95–100. https://doi.org/10.1016/j.pain.2009.03.018.

[34] T.E. Lacourt, J.H. Houtveen, L.J.P. van Doornen, Experimental pressure-pain assessments: Test-retest reliability, convergence and dimensionality, Scand. J. Pain. 3 (2012) 31–37. https://doi.org/10.1016/j.sjpain.2011.10.003.

[35] J.S. Blythe, M.A. Thomaidou, K.J. Peerdeman, A.I.M. Van Laarhoven, Placebo effects on cutaneous pain and itch : a systematic review and meta-analysis of experimental results and methodology, Pain. 164 (2023) 1181-1199.

[36] D.J.P. Bartels, A.I.M. van Laarhoven, M. Stroo, K. Hijne, K.J. Peerdeman, A.R.T. Donders, P.C.M. van de Kerkhof, A.W.M. Evers, Minimizing nocebo effects by conditioning with verbal suggestion: A randomized clinical trial in healthy humans, PLoS One. 12 (2017) 1–19. https://doi.org/10.1371/journal.pone.0182959.

[37] F. Wolfe, D.J. Clauw, M.A. Fitzcharles, D.L. Goldenberg, W. Häuser, R.L. Katz, P.J. Mease, A.S. Russell, I.J. Russell, B. Walitt, 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria, Semin. Arthritis Rheum. 46 (2016) 319–329. https://doi.org/10.1016/j.semarthrit.2016.08.012.

[38] C.S. Burckhardt, S.R. Clark, R.M. Bennett, The fibromyalgia impact questionnaire: development and validation, J. Rheumatol. 18 (1991) 728–733.

[39] S.H. Lovibond, P.F. Lovibond, Manual for the depression anxiety stress scales, Psychol. Found. Aust. (1995).

[40] M.F. Scheier, C. Carver, Distinguishing Optimism From Neuroticism (and Trait Anxiety, Self-Mastery, and Self-Esteem): A Reevaluation of the Life Orientation Test, Artic. J. Personal. Soc. Psychol. 67 (1994) 1063-1078. https://doi.org/10.1037/0022-3514.67.6.1063.

[41] M.J.L. Sullivan, S.R. Bishop, J. Pivik, The Pain Catastrophizing Scale: Development and Validation, Psychol. Assess. 7 (1995) 524–532. https://doi.org/10.1037/1040-3590.7.4.524.

[42] N.B. Schmidt, D.R. Lerew, J.H. Trakowski, Body vigilance in panic disorder: Evaluating attention to bodily perturbations, J. Consult. Clin. Psychol. 65 (1997) 214–220. https://doi.org/10.1037/0022-006X.65.2.214.

[43] L.I. Pearlin, C. Schooler, The structure of coping, J. Health Soc. Behav. 19 (1978) 2–21.

[44] T.M. Marteau, H. Bekker, The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI), Br. J. Clin. Psychol. 31 (1992) 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x.

[45] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. https://www.r-project.org/.

[46] E.R. Girden, ANOVA: Repeated measures, Sage, Newbury Park, CA, 1992.

[47] A. Field, Discovering statistics using IBM SPSS statistics, fourth ed., Sage, London, 2013.

[48] L. Colloca, D. Finniss, Effects, Nocebo Communication, Patient-clinician Outcomes, Therapeutic, Jama. 307 (2019) 567–568. https://doi.org/10.1001/jama.2012.115.Nocebo.

[49] L. Colloca, R. Panaccione, T.K. Murphy, The Clinical Implications of Nocebo Effects for Biosimilar Therapy, Front. Pharmacol. 10 (2019) 1372. https://doi.org/10.3389/fphar.2019.01372.

[50] M.A. Thomaidou, D.S. Veldhuijzen, A. Meulders, A.W.M. Evers, An experimental investigation into the mediating role of pain-related fear in boosting nocebo hyperalgesia, Pain. 162 (2021) 287-299. https://doi.org/10.1097/j.pain.0000000000002017.

[51] R. Treister, L. Honigman, O.D. Lawal, R.K. Lanier, N.P. Katz, A deeper look at pain variability and its relationship with the placebo response: Results from a randomized, double-blind, placebo-controlled clinical trial of naproxen in osteoarthritis of the knee, Pain. 160 (2019) 1522–1528. https://doi.org/10.1097/j.pain.0000000000001538.

[52] A.I.M. van Laarhoven, F.W. Kraaimaat, O.H. Wilder-Smith, A.W.M. Evers, Role of attentional focus on bodily sensations in sensitivity to itch and pain, Acta Derm. Venereol. 90 (2010) 46–51. https://doi.org/10.2340/00015555-0743.

[53] D. Borsook, R. Edwards, I. Elman, L. Becerra, J. Levine, Pain and analgesia: The value of salience circuits, Prog. Neurobiol. 104 (2013) 93–105. https://doi.org/10.1016/j.pneurobio.2013.02.003.

[54] S. Duschek, C.I. Montoro, G.A. Reyes del Paso, Diminished Interoceptive Awareness in Fibromyalgia Syndrome, Behav. Med. 43 (2017) 100–107. https://doi.org/10.1080/08964289.2015.1094442.

[55] C. Valenzuela-Moguillansky, A. Reyes-Reyes, M.I. Gaete, Exteroceptive and interoceptive body-self awareness in fibromyalgia patients, Front. Hum. Neurosci. 11 (2017) 1–14. https://doi.org/10.3389/fnhum.2017.00117.

[56] R.A. Rescorla, Spontaneous recovery, Learn. Mem. 11 (2004) 501–509. https://doi.org/10.1101/lm.77504.

[57] L. Colloca, M. Sigaudo, F. Benedetti, The role of learning in nocebo and placebo effects, Pain. 136 (2008) 211–218. https://doi.org/10.1016/j.pain.2008.02.006.

CHAPTER 4 SUPPLEMENTARY MATERIALS

Supplemental File I

Formulas used during pressure pain calibration

Step 1: Ascending Series

During this step, participants receive an increasing amount of pressure stimuli (kgf/cm^2) and rate their subjective pain intensity after each stimulus (0-10 Numeric Rating Scale (NRS)). When a certain pain threshold has been reached (> 5.5 NRS), the Ascending Series stops. Based on the participant ratings, 5 pressure stimulus intensities are calculated that are to be used during Randomized Series. This is done by taking the stimulation intensities on which the lowest (highest stimulus intensity rated between 0-1) and the highest (highest stimulus intensity rated between 4.5-5.5) pain rating was given, and calculate the 3 values that lie in between (with equal distances between each two subsequent values), ending up with 5 intensities in total. If participants did not rate their pain between 0-1 or the 4.5-5.5 ranges, the adjustments were made by E-prime using the following formulas:

- *Lowest pressure intensity:*
  If there are no pressure intensities that are scored 0-1, the lowest possible pressure intensity (1 kgf/cm^2) will be chosen.
- *Highest pressure intensity:*
  Interpolation: When there are no pressure intensities rated between 4.5-5.5, the highest pressure intensity is calculated using the first pain rating higher than the higher bound (5.5) and the first pain rating that is lower than the lower bound (4.5). These ratings are then used to interpolate and calculate the pressure intensity that corresponds to the middle value of the highest pressure intensity (5.0).
  Extrapolation: When there are no pressure intensities rated between 4.5-5.5, and no pressure intensities rated lower than the lower bound (4.5) or no pressure intensities rated higher than the higher bound (5.5) , the two pressure intensities corresponding to the first two ratings higher than the higher bound, or the first two ratings lower than the lower bound, respectively, are used to extrapolate the pressure intensity that corresponds to the middle value of the highest pressure intensity (5.0).

Step 2: Randomized Series

During this step, participants receive each of the 5 pressure intensities 3 times. This results in a total of 15 pressure stimuli. Using ratings of the pressure intensities entered in randomized series, 3 pressure intensities that are directly used in the experiment are calculated. These pressure intensities correspond to: no pain (0-1), slight pain (2-3), and moderate pain (4.5-5.5).

- *No Pain:*
  Extrapolation: When there are no pressure intensities rated between 0-1, the two pressure intensities corresponding to the first two ratings higher than the higher bound (1) are used to extrapolate the pressure intensity that corresponds to the middle value of no pain range (0.5).
- *Slight Pain:*
  Interpolation: When there are no pressure intensities rated between 2-3, slight pain is calculated using the first pain rating higher than the higher bound (3) and the first pain rating that is lower than the lower bound (2). These ratings are then used to interpolate and calculate the pressure intensity that corresponds to the middle value of the slight pain range (2.5).
  Extrapolation: When there are no pressure intensities rated between 2-3, and no pressure intensities rated lower than the lower bound (2) or no pressure intensities rated higher than the higher bound (3), the two pressure intensities corresponding to the first two ratings higher than the higher bound, or the first two ratings lower than the lower bound, respectively, are used to extrapolate the pressure intensity that corresponds to the middle value of the slight pain range (2.5).
- *Moderate Pain:*
  Interpolation: When there are no pressure intensities rated between 4.5-5.5, moderate pain is calculated using the first pain rating higher than the higher bound (5.5) and the first pain rating that is lower than the lower bound (4.5). These ratings are then used to interpolate and calculate the pressure intensity that corresponds to the middle value of the moderate pain range (5.0).
  Extrapolation: When there are no pressure intensities rated between 4.5-5.5, and no pressure intensities rated lower than the lower bound (4.5) or no pressure intensities rated higher than the higher bound (5.5), the two pressure intensities corresponding to the first two ratings higher than the higher bound, or the first two ratings lower than the lower bound, respectively, are used to extrapolate the pressure intensity that corresponds to the middle value of the moderate pain range (5.0).

Step 3: Calibration Check

During this step, participants receive the 3 pressure intensities calculated during the Randomized Series for a final check. No pain intensity (0-1) is presented twice, slight pain intensity (2-3) is presented thrice, and moderate pain intensity (4.5-5.5) is presented twice in a randomized order, where the no pain and moderate pain intensities are never presented subsequently (or other way around). No adjustment is necessary when at least 1/2 no pain intensities, 2/3 slight pain intensities, and 1/2 moderate pain intensities are rated within the correct ranges.

- *No Pain:*
  When none of the no pain intensities were rated 0-1, extrapolation to a pain rating of 0.5 is used to adjust the no pain intensity.
- *Slight Pain:*
  When none of the slight pain intensities were rated 2-3, interpolation or extrapolation to a pain rating of 2.5 is used to adjust the slight pain intensity. When only one of the slight pain pressure intensities were rated in the slight pain range, interpolation or extrapolation is still used, using this one rating within the slight pain range.
- *Moderate Pain:*
  When none of the moderate pain intensities were rated 4.5-5.5, interpolation or extrapolation to a pain rating of 5.0 is used to adjust the moderate pain intensity.

Note on the interpolation/extrapolation: When the first two (or three, however much) **pressure intensities** higher than the higher bound or lower than the lower bound have the same rating, the median of these pressure intensities and the pressure intensity corresponding to the next rating is used for extrapolation. However, when the first two (or three, however much) **pain ratings** higher than the higher bound and lower than the lower bound have the same pressure intensities, nothing is done. When the pressure intensities used for interpolation are the same, the interpolation will result in the same pressure intensity. Same for extrapolation, when the pressure intensities used for extrapolation are the same, the extrapolation will result in the same pressure intensity.

Supplemental File II

Results from the sensitivity analyses

Sensitivity Analyses

It was identified that nocebo effects were induced in 65% of all participants (54% of patients and 78% of healthy controls) in session 1, i.e., the difference scores after nocebo conditioning were above zero. In session 2, this was the case in 66% of all participants (62% of patients and 70% of healthy controls). 43% of all participants (37% of patients and 52% of healthy controls) participating in both sessions were nocebo responders in each session. As part of sensitivity analyses, the same analyses for the extinction of nocebo effects were conducted for only the nocebo responders. To assess the extinction efficacy in session 1, a 2 x 2 mixed-design ANOVA analysis was conducted with 43 nocebo responders (patient N = 18; healthy control N = 25). The results showed no significant interaction effect ($F(1,41) = 1.45$, $p = .23$, $\eta_p^2 = .03$) nor a main effect of group ($F(1,41) = .24$, $p = .63$, $\eta_p^2 = .01$), but there was a significant main effect of time ($F(1,41) = 38.78$, $p < .001$, $\eta_p^2 = .49$). After extinction in session 1, the mean difference score was significantly reduced by 0.43 (SE = .07) NRS points across groups. Moreover, to investigate the extinction efficacy in session 2, the same 2 x 2 mixed-design ANOVA was conducted with 37 nocebo responders from session 2 (patient N = 18; healthy control N = 19). The results showed no significant interaction effect ($F(1,35) = .35$, $p = .56$, $\eta_p^2 = .01$) nor a main effect of group ($F(1,35) = .45$, $p = .51$, $\eta_p^2 = .01$), but there was a significant main effect of time ($F(1,35) = 4.35$, $p = .04$, $\eta_p^2 = .11$). After extinction in session 2, the mean difference score was significantly reduced by 0.28 (SE = .13) NRS points across groups.

Moreover, to investigate the stability of nocebo reduction between groups and across sessions, the same 5 x 2 mixed-design ANOVA analysis was conducted with 22 participants (patient N = 8; healthy control N = 14) who were nocebo responders in both sessions. The results showed no interaction effect ($F(4, 80) = .16$, $p = .96$, $\eta_p^2 = .01$) nor a main effect of group ($F(1,20) = .009$, $p = .92$, $\eta_p^2 = .00$), but there was a significant main effect of time ($F(4, 80) = 8.65$, $p < .001$, $\eta_p^2 = .30$). Similar to earlier findings, Bonferroni-corrected pairwise comparisons between extinction in session 1 and the recall testing phase in session 2 were not statically different ($p = .10$) and neither was the comparison between extinction in session 1 and 2 ($p = .07$).

Moreover, removing 3 patients with FS <12 did not significantly influence the findings on the induction and extinction of nocebo effects in both sessions.

Furthermore, there were no statistically significant associations in the magnitude of nocebo effects between sessions in the patient group ($r(25)= .22$, $p = .27$) nor in the healthy control group ($r(25)= .07$, $p = .74$). After pooling the samples, the results were the same ($r(52)= .18$, $p = .18$).

Results without the exclusion of outliers

Induction of nocebo effects in session 1

Without excluding one patient outlier, the results no longer showed an interaction effect ($F(1,67) = 2.81$, $p = .09$, $\eta_p^2 = .04$), nor a main effect of group ($F(1,67) = .01$, $p = .93$, $\eta_p^2 = .00$), but there was a significant main effect of trial type ($F(1,67) = 21.67$, $p < .001$, $\eta_p^2 = .24$), where mean pain ratings were significantly higher in experiments trials ($M = 4.24$, $SE = 0.20$) compared to control trials ($M = 3.79$, $SE = 0.21$), indicating that nocebo effects were induced across groups.

Extinction of nocebo effects in session 1

Without excluding two patient outliers, the results no longer showed an interaction effect ($F(1,67)= 2.84$, $p = .09$, $\eta_p^2 = .04$), nor a main effect of group ($F(1,67)=1.09$, $p = .29$, $\eta_p^2 = .02$), but there was a significant main effect of time ($F(1,67) = 4.63$, $p = .04$, $\eta_p^2 = .06$), with a significant reduction of nocebo effects of 0.23 NRS points ($SE = .10$) after extinction across groups ($p = .04$).

Supplemental File III

Assessment of prior experience with a TENS device

During baseline questionnaire assessment, all participants were asked to categorically indicate (yes/no) to their prior knowledge of and experience with a TENS device, and in case of experience whether they found it was effective in reducing pain.

Statistical Analyses

Frequency of responses were calculated. As a manipulation check, 6 One-Way ANOVAs were conducted to investigate whether:

1. Prior TENS knowledge influenced nocebo response (in sessions 1 and 2)
2. Prior TENS experience influenced nocebo response (in sessions 1 and 2)
3. Prior TENS efficacy influenced nocebo response (in sessions 1 and 2)

Bonferroni corrections were applied, where a p-value below .008 was considered statistically significant

Results

Only a small number of participants had prior knowledge of ($N_{patient}$ = 17, $N_{healthy}$ = 3) and experience with ($N_{patient}$ = 8) a TENS device. Amongst those who have experience, 4 patients found that it was effective in reducing pain.

Regarding TENS, neither having knowledge (session 1 $F(1,68) = .22$, $p = .64$; session 2 $F(1,55) = .52$, $p = .47$), experience (session 1 $F(1,68) = .05$, $p = .82$; session 2 $F(1,55) = .03$, $p = .86$), nor perceived efficacy (session 1 $F(1,8) = 4.38$, $p = .08$; session 2 $F(1,4) = 1.86$, $p = .27$) influenced nocebo responses.