Was it so bad?

The role of retrospective memory in symptom reporting

Marta Walentynowicz°

Katleen Bogaerts°

Ilse Van Diest°

Filip Raes*

Omer Van den Bergh°

°Health Psychology

*Learning and Experimental Psychopathology

University of Leuven, Leuven, Belgium

Corresponding author:

Omer Van den Bergh, PhD

Health Psychology

University of Leuven

Tiensestraat 102 – box 3726, 3000 Leuven, Belgium.

Tel.: +32-(0)16-32.60.58, Fax: +32-(0)16-32.61.44.

Email: omer.vandenbergh@ppw.kuleuven.be

Abstract

Objective: Retrospective symptom reports are an important source of information in both laboratory and clinical settings. The present study investigated memory for experimentally induced pain and dyspnea in high and low habitual symptom reporters (HSR).

Methods: Healthy women (N=48; 24 high/24 low HSR) participated in two laboratory studies. One study included two pain episodes (cold pressor task), the other study included two dyspnea episodes (rebreathing task). Pain and dyspnea ratings were collected (1) continuously during symptom inductions, (2) after each trial, (3) immediately after the experiment, and (4) at 2-week follow up. Symptom ratings, negative affect (NA) and anxiety measures were also completed following each trial.

Results: While the retrospective pain ratings were higher in the high compared with the low HSR group ($p = .01$), both groups rated recalled dyspnea higher relative to concurrent dyspnea ($p < .001$). A further increase in bias over time was only found for dyspnea in high HSR ($p = .02$). Moreover, dyspnea induction was associated with higher state NA ($p = .03$) and anxiety ($p = .007$) than pain induction.

Conclusions: Our findings show that even though memory for pain and dyspnea is overall distorted, the extent of bias in symptom recall clearly differs between symptoms and groups. The observed increase of dyspnea reporting over time may have important implications for diagnostic assessment based on symptom reporting.

Keywords: pain, dyspnea, symptom memory, biased symptom recall, habitual symptom reporting

Retrospective descriptions of somatic experiences are important sources of information for health care professionals and can influence clinical diagnosis and treatment choice. Interestingly, a considerable number of patients in both primary (Barsky, Orav, & Bates, 2005; Khan, Khan, Harezlak, Tu, & Kroenke, 2003) and secondary health care (Carson et al., 2000; Nimnuan, Rabe-Hesketh, Wessely, & Hotopf, 2001) tend to report symptoms in the absence of underlying physical dysfunction (often called Medically Unexplained Symptoms; MUS). In the recently proposed DSM-5 (American Psychiatric Association, 2013) most of these patients would meet criteria for somatic symptom disorder (SSD), which emphasizes the presence of persistent distressing somatic symptoms, as well as excessive thoughts, feelings and behaviors linked to those symptoms. Various studies have explored the perceptual-cognitive processes underlying symptom overreporting in this group (Rief & Broadbent, 2007; Van den Bergh, Bogaerts, & Van Diest, in press), but few studies have focused on the role of memory processes herein.

Although research on memory for symptoms often results in contradictory findings, one consistent conclusion is that memory for symptoms is relatively inaccurate and mostly results in retrospective overestimation of experienced symptoms (Broderick et al., 2008; Giske, Sandvik, & Røe, 2010; Linton & Melin, 1982). Several sources of bias have been identified: (1) variability of real-time symptom levels (Lefebvre & Keefe, 2002; Sohl & Friedberg, 2008; Stone, Schwartz, Broderick, & Shiffman, 2005), (2) symptom intensity (Feine, Lavigne, Thuan Dao, Morin, & Lund, 1998; Giske et al., 2010; Hunter, Philips, & Rachman, 1979; Sohl & Friedberg, 2008), (3) emotional state during symptom experience (Everts et al., 1999; Gedney & Logan, 2004), (4) symptom intensity during recall (Eich, Reeves, Jaeger, & Graff-Radford, 1985; Lefebvre & Keefe, 2002; Meek, Lareau, & Anderson, 2001; Smith & Safer, 1993), (5) time since actual symptom episode (Broderick et al., 2008; Houtveen & Oei, 2007), and (6) cognitive heuristics, such as the peak-end effect

(Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). The peak-end effect assumes that the retrospective evaluation of an experience is predominantly determined by two distinctive moments, the one with the highest intensity (peak) and the final (end) part of the episode, with relative duration neglect, meaning that the actual duration of the experience has a limited influence on the global retrospective evaluation. The influence of this heuristic on symptom memory was confirmed not only in the laboratory (Bogaerts et al., 2012; Kahneman et al., 1993), but also in naturalistic settings, such as during medical examinations (Redelmeier & Kahneman, 1996; Redelmeier, Katz, & Kahneman, 2003) and childbirth (Chajut, Caspi, Chen, Hod, & Ariely, 2014). Finally, psychological factors may inflate both concurrent (catastrophizing, Lefebvre & Keefe, 2002; Sohl & Friedberg, 2008; anxiety, Suls & Howren, 2012) and retrospective (depression, Suls & Howren, 2012) symptom ratings. Also negative affectivity (NA) has been found to be strongly related to memory distortions for symptoms (Levine & Safer, 2002; Safer, Levine, & Drapalski, 2002). Similarly, trait NA is associated with a tendency to attend more to somatic information (Stegen, Van Diest, Van de Woestijne, & Van den Bergh, 2001) and to interpret it as threatening (Stegen, Van Diest, Van de Woestijne, & Van den Bergh, 2000), as well as to overreport symptoms during recall (Larsen, 1992).

Even though memory processes play an important role in biased symptom reporting, only a limited number of studies has explored memory for bodily symptoms among patients with MUS or non-consulting high habitual symptom reporters (HSR). Bogaerts et al. (2012) investigated the peak-end effect in the memory for dyspnea in patients with medically unexplained dyspnea (MUD) and healthy participants. Dyspneic experience was experimentally induced with two rebreathing trials: one ended at the peak of dyspnea, while in the other a recovery phase was added to assure a milder end. The expected peak-end effect was observed among healthy participants, but not in the MUD group. Because the patients

displayed a slower recovery in self-reported dyspnea which could not be accounted for by differences in respiratory physiology, it is suggested that perceptual-cognitive processing of aversive sensations among patients with MUD differs from healthy people. Investigating a non-consulting high HSR group, Houtveen and Oei (2007) conducted a diary study and found that, compared to averaged concurrent symptom reports, both high and low HSR reported experiencing more symptoms during recall. However, only high HSR showed a gradual increase in estimation of experienced symptoms with longer time frames. Moreover, biased recall in high HSR was not related to the hypothesized sources of bias, i.e. the peak-end effect and symptom variability. Taken together, delayed recovery in symptom reports but not in physiological dysfunction, as well as bias in retrospective symptom reporting, suggest distorted and less detailed perceptual-cognitive processing of symptom experiences in persons with MUS.

In view of this limited set of findings, the present study aimed to advance our understanding of the role of perceptual-cognitive biases affecting retrospective symptom reports. The primary goals of this study were to investigate whether the retrospective symptom reports are subject to recall biases leading to increased symptom reporting and whether such biases are larger for high HSR. To this end, retrospective memory for two experimentally induced and well-controlled aversive bodily sensations, i.e. pain and dyspnea, was examined. We selected participants high and low on HSR and administered the two aversive sensations within subject to examine the generality of the findings across symptom types. In one study (StudyPain), a painful experience was induced by means of the cold pressor task (CPT), while in the other (StudyDyspnea), dyspnea was induced via a rebreathing paradigm (Read, 1967). Each study consisted of two trials in order to investigate the peak-end effect: One terminated at peak distress, while the other included an additional recovery phase to end at a less distressing level. Participants rated their concurrent symptom

levels while being exposed to aversive stimuli during the trials, which were followed by three retrospective ratings of induced symptoms and affective responses. Based upon the arguments described above, the following hypotheses were tested in each study separately: (1) Retrospective symptom ratings were expected to be higher than averaged concurrent ratings, with this effect being more pronounced in high HSR; (2) Recalled symptom reports were expected to increase over time in high HSR, but not in low HSR; (3) According to the peak-end rule, short trials were expected to be retrospectively rated as more intense than the long trials. However, this effect was hypothesized to be present in low, but not high, HSR (Bogaerts et al., 2012; Houtveen & Oei, 2007). Possible differences between the two symptom types were investigated in an exploratory manner, thus no specific hypotheses were formulated regarding these differences.

Methods

Participants

Forty-eight healthy students (all women), aged 18 – 27 years, participated in both experiments in return for two course credits or 15 euros. They were selected after screening for habitual symptom reporting via the Checklist for Symptoms in Daily Life (CDS; Wientjes & Grossman, 1994). Predefined cut-off scores were used to select high ($\geq$ 100; n = 24) and low ($\leq$ 75; n = 24) habitual symptom reporters (HSR). Cut-off scores were based on upper and lower quartiles of the scores on this questionnaire found in large samples from the same population (Bogaerts et al., 2008). Prior to the experiment, participants completed the CDS a second time; only participants who still met the cut-offs were included.

Exclusion criteria were any self-reported chronic illness (e.g. pulmonary, cardiovascular, gastrointestinal, neuromuscular diseases), acute illnesses, fever or headache, major psychiatric condition, diabetes, recent arm fracture or wrist sprain prior to participating, earlier frostbite, and pregnancy. The experimental protocol was approved by

the Multidisciplinary Ethical Committee of the Faculty of Psychology and Educational Sciences of the University of Leuven.

Measures

Habitual symptom reporting. Habitual symptom reporting was assessed using the adapted version of The Checklist for Symptoms in Daily Life (CDS; Wientjes & Grossman, 1994). Participants rated how often they experienced 39 listed symptoms in the past year on a 5-point Likert scale (*never*, *seldom*, *sometimes*, *often*, *very often*). The total score (range: 39 – 195) was used to select high/low HSR; reliability (Cronbach's α) exceeded .95 in our sample.

Negative affectivity. Trait and state Negative Affectivity (NA) were assessed with the Dutch version of the Positive and Negative Affect Schedule. The PANAS consists of 20 positive and negative adjectives for which participants had to indicate (on a 5-point Likert scale ranging from *not at all* to *very much*) to which extent they felt that way in general (trait) or now (state). Good reliability and validity have been reported (Engelen, De Peuter, Victoir, Van Diest, & Van den Bergh, 2006; Watson, Clark, & Tellegen, 1988).

State symptom checklist. At baseline and after every symptom induction trial in both studies, a state symptom checklist was administered. Participants had to rate to which extent they experienced each of 12 symptoms now (baseline) or during the past trial on a 5-point rating scale ranging from 0 (*not at all*) to 4 (*very much*). This symptom list included: *chest tightness*, *pounding of the heart*, *stomach or abdominal cramps*, *headache*, *fatigue*, *not able to breathe deeply*, *rapid heartbeat*, *nausea*, *dizziness*, *muscular pain*, *dyspnea*, *pain*. State symptom checklists showed acceptable and good internal consistency, with Cronbach's alphas ranging from .70 to .86.

State anxiety and threat. A numerical rating scale (NRS) was used to evaluate the level of anxiety (1 = *not anxious at all*, 9 = *very anxious*) at the baseline and after every symptom induction trial in both studies. Additionally, after every symptom induction trial, a

NRS concerning the threat value of each trial (1 = *not threatening at all*, 9 = *very threatening*) was administered.

Concurrent symptom ratings. During each symptom induction trial, concurrent symptom ratings were collected on a 0-100 computerized scale. The scale was presented as a vertical bar in the middle of the screen. Different levels of the experienced pain/dyspnea, based on a modified Borg scale (Borg, 1982), were verbally described on its right side: *none* (0), *very slight* (10), *slight* (20), *moderate* (30), *fairly severe* (40), *severe* (50), *very severe* (60), *very severe* (70), *very severe* (80), *very*, *very severe* (90), *intolerable* (100). In StudyPain, perceived pain was rated continuously with a scroll wheel (sampling every second), while in StudyDyspnea perceived dyspnea was rated every 10s (after auditory cue) with a mouse click.

Retrospective symptom ratings. In both studies, the retrospective evaluations of symptoms experienced during each trial were collected at three moments: immediately after each trial (immediate rating), at the end of the experimental session (delayed rating) and in a two weeks follow-up (follow-up rating). Participants indicated the average symptom level (StudyPain: pain/ StudyDyspnea: dyspnea) experienced during the trial (*How much pain/ dyspnea have you experienced on average during this trial?*) on a visual analog scale (10cm) ranging from 0 (*no pain/ dyspnea*) to 100 (*maximum pain/ dyspnea*). The follow-up ratings for both studies were collected on a single occasion two weeks after the last study.

Apparatuses and Physiological Recordings

StudyPain: Pain induction – the cold pressor task (CPT). During the two trials of the cold pressor task (CPT), participants immersed their hand in a Plexiglas box (Julabo®) filled with 18L of water. The water temperature was controlled by an electric immersion cooler, type FT200, and a bath circulator, type ED-19A. This ensured that the temperature could be either maintained at a constant level (12°C) or increased by 2°C in 60s. The changes

in water circulation during temperature manipulation were unnoticeable for the participant. In contrast with Kahneman et al. (1993), who used 14°C to induce pain and 15°C to induce discomfort reduction, temperatures in the current experiment were set at 12°C and 14°C (± 0.3°C), after a pilot study showing a detectable change in discomfort with these temperatures, which was not observed with 14°C and 15°C. The Plexiglas box was placed upon a trolley adjustable in height to provide comfortable access. Before each CPT, participants were asked to hold both hands in the second box (type FT200 Julabo®), in which water was kept at room temperature (20.5°C ± 0.3°C). A 2-minute baseline was used to ensure that the skin temperature of the participants was similar before each trial.

StudyDyspnea: Dyspnea induction – the rebreathing paradigm. Two trials of the standardized rebreathing paradigm (Read, 1967, see also Bogaerts et al., 2012) were used to induce the sensations of dyspnea. During the trials, participants wore a nose clip and breathed through a mouthpiece, connected to the rebreathing bag via a wide vinyl tube and a Y-valve ending on a pneumotachograph (Fleisch no. 2, Lausanne, Switzerland) measuring airflow. The valve allowed to switch between room air and the rebreathing bag, which was initially filled with 5-liter gas mixture of 5% CO_2 and 95% O_2. Breathing in this closed hyperoxic system led to a progressive rise of PCO_2, of minute ventilation and of dyspnea, defined as uncomfortable feeling of not having enough air, an urge to breathe, or a feeling of having more difficulty in breathing. Fractional end-tidal concentration of CO_2 ($FetCO_2$) was determined using an infrared CO_2 monitor (POET RC, Criticare Systems Inc., Waukesha, WI). The exhaled air was sampled close to the mouthpiece. The data from the pneumotachograph and the CO_2 monitor were sampled at 20Hz and stored on a computer. All data were stored and analyzed offline to determine the following parameters: minute ventilation (MV) in L/min and $FetCO_2$ in %. The manipulation of the valve was undetectable

by the participants to make sure that they depended exclusively on the experienced bodily changes to give their concurrent ratings.

Procedure

Selected participants were invited to participate in two experiments examining the relationship between bodily sensations and well-being. One day before the laboratory sessions, participants completed the trait questionnaires (CSD, PANAS) online. Upon arrival in the laboratory, participants signed the informed consent, were informed about the procedure and completed the questionnaires (state symptoms checklist, PANAS and anxiety). Participants completed both studies on two consecutive days and the order of the studies and the trials within study was counterbalanced across participants.

In StudyPain, two CPT trials were administered, one on each hand, with a 7-minute intertrial interval. Each trial began with a baseline period, during which participants immersed both hands in room-temperature water. For the short trial, baseline was followed by a cold phase (60 seconds in 12°C water) after which participants could withdraw their hand. For the long trial, baseline was followed by the same cold phase (60 seconds in 12°C) with an additional recovery phase (60 seconds) during which the temperature increased to 14°C (unknown to the participants). The order of trials (short/long, to the dominant/nondominant hand) was counterbalanced across participants.

In StudyDyspnea, participants went through two rebreathing trials, the order of which was counterbalanced across participants. A short trial consisted of a baseline (60 seconds of room-air breathing) and a rebreathing phase (150 seconds). After 150-second rebreathing, the trial was stopped and participants could breathe freely outside of the rebreathing system. In a long trial, the baseline (60 seconds) and rebreathing phases (150 seconds) were followed by an additional recovery phase (150 seconds), initiated by unobtrusively switching the valve to

room air. Full recovery between the trials was ensured by a 15-minute intertrial interval. Respiration was measured throughout each trial.

Average pain/dyspnea experienced during the trials was rated after each trial (immediate rating), followed by the ratings of state affect, symptoms and anxiety. After two trials of each study, the average pain/dyspnea ratings were repeated (delayed rating). Two weeks after the second session participants completed an online questionnaire regarding retrospective ratings for both trials in each of the studies (follow-up rating).

Data Analyses

Manipulation check: Concurrent symptom ratings. Concurrent symptom ratings were analyzed to verify the effect of symptom induction and to examine whether symptom reports differed between high and low HSR. Concurrent ratings were divided into equal time segments of 10 seconds (StudyPain) or 30 seconds (StudyDyspnea) in order to acquire a detailed picture of the somatic experience. Separate repeated-measures analyses of variance (ANOVAs) were conducted on self-reported symptom ratings (pain/dyspnea) during each trial as dependent variables, with Group and Order of trials as between-subject factors and Time segment as within-subject variable. Moreover, the effect of dyspnea induction on respiratory behavior was investigated by separate repeated measures ANOVAs on MV and $FetCO_2$ (per 30 seconds) during each trial as dependent variables, with Group and Order of trials as between-subject factors and Time segment as within-subject variable.

Testing hypotheses: Retrospective symptom ratings. To assess symptom memory in both studies, separate repeated measures ANOVAs were performed on symptom ratings (pain/dyspnea) as dependent variables, with Group (high/low HSR) and Order of trials as between-subjects factors, Trial (short trial/long trial) and Moment of symptom assessment (averaged concurrent/immediate/delayed/follow-up) as within-subject factors. The averaged concurrent symptom scores were averaged across the actual symptom reports given during

pain and dyspnea inductions per trial. Planned contrasts were used to examine specific time and group effects (*C1*: averaged concurrent vs. all retrospective ratings to test Hypothesis 1; *C2*: retrospective ratings during experimental session (immediate, delayed) vs. follow-up ratings to test Hypothesis 2), as well as trial effects (*C3*: averaged concurrent vs. immediate ratings to test Hypothesis 3). Greenhouse–Geisser corrections were applied when the sphericity assumption was violated.

Affective responses. In order to investigate the differences in affective responses to different bodily stimuli, data regarding affective states during both studies were analyzed together. Repeated measures ANOVAs were carried out on state symptom, NA and anxiety ratings as dependent variables, with Induction (StudyPain/StudyDyspnea) and Moment of measurement (baseline/short trial/long trial) as within-subject factors and Group (high/low HSR) and Order of studies as between-subject factors. Another repeated measures ANOVA was conducted on threat value as dependent variable, with Induction (StudyPain/StudyDyspnea) and Moment of measurement (short/long trial) as within-subject factors and Group (high/low HSR) and Order of studies as between-subject factors. All analyses were conducted with SPSS 22.0.

Results

Sample characteristics

Low HSR reported less habitual symptoms than high HSR (low: $M = 57.13$, $SE = 2.06$; high: $M = 114.75$, $SE = 2.61$; $t(46) = -17.33$, $p < .001$). High HSR also reported higher trait NA levels than low HSR, $t(38.04) = -7.34$, $p < .001$.

StudyPain

Manipulation check: Concurrent pain ratings. No significant group-related differences were observed for either pain ratings or their change over time during the short

trial, while during the long trial high HSR tended to report more pain, $F(1, 44) = 2.98$, $p = .09$, $\eta_p^2 = .06$ (see Fig. 1).

Testing hypotheses: Retrospective pain ratings. Hypothesis 1 (retrospective symptom reporting): Main effects showed that high HSR reported overall more pain than low HSR, $F(1, 44) = 4.80$, $p = .03$, $\eta_p^2 = .10$ (Fig. 2, left panel) and that pain was higher when rated retrospectively than concurrently, *C1*: $F(1, 44) = 58.99$, $p < .001$, $\eta_p^2 = .57$. Moreover, the latter effect was stronger in high HSR than in low HSR participants, *C1* for Group × Moment: $F(1, 44) = 6.63$, $p = .01$, $\eta_p^2 = .13$; Group × Moment: $F(2.43, 106.96) = 3.53$, $p = .03$, $\eta_p^2 = .07$. Hypothesis 2 (increase in retrospective symptom reporting over time): Retrospective ratings did not further increase over time, *C2*: $F(1, 44) = 1.35$, $p = .25$, $\eta_p^2 = .03$. Hypothesis 3 (peak-end effect): Even though the concurrently rated pain was lower in the short trial than in the long trial, pain rating in the immediate rating was higher for the short, compared with the long, trial, *C3*: $F(1, 44) = 17.38$, $p < .001$, $\eta_p^2 = .28$ (Fig. 3, left panel); Trial × Moment: $F(1.91, 83.99) = 4.69$, $p = .01$, $\eta_p^2 = .10$. This confirmed the peak-end effect, but no group differences appeared for this interaction, *C3*: $F(1, 44) = 1.22$, $p = .28$, $\eta_p^2 = .03$; Trial × Group × Moment: $F(1.91, 83.99) = .75$, $p = .47$, $\eta_p^2 = .02$.

StudyDyspnea

Manipulation check: Concurrent dyspnea ratings. Group-related differences were observed for concurrent symptom ratings in both trials, with high HSR reporting more dyspnea during the short, $F(1, 41) = 9.53$, $p < .01$, $\eta_p^2 = .19$, and the long trial, $F(1, 41) = 3.97$, $p = .05$, $\eta_p^2 = .09$ (Fig. 4). Moreover, this difference became stronger over time during the short trial (Group × Time segment: $F(1.50, 61.40) = 5.81$, $p < .01$, $\eta_p^2 = .12$). No group differences were observed for either $FetCO_2$ or MV during the rebreathing trials.

Testing hypotheses: Retrospective dyspnea ratings. Hypothesis 1 (retrospective symptom reporting): Retrospective dyspnea ratings were higher compared to averaged concurrent ones, *C1*: $F(1, 43) = 126.63$, $p < .001$, $\eta_p^2 = .75$ (Fig. 2, right panel); Moment of symptom assessment: $F(2.29, 98.55) = 64.45$, $p < .001$, $\eta_p^2 = .60$. However, the expected group differences were not found for dyspnea ratings, $F(1, 43) = 1.70$, $p = .20$, $\eta_p^2 = .04$. Hypothesis 2 (increase in retrospective symptom reporting over time): A Group × Moment of symptom assessment interaction emerged, $F(2.29, 98.55) = 3.02$, $p = .05$, $\eta_p^2 = .07$, revealing that high HSR gave higher follow-up ratings compared to immediate and delayed ratings than low HSR (*C2*: $F(1, 43) = 6.39$, p = .02, $\eta_p^2 = .13$). Hypothesis 3 (peak-end effect): Even though the concurrently rated dyspnea did not differ between the trials, dyspnea in the immediate rating was higher for the short, compared with the long, trial, *C3*: $F(1, 43) = 28.09$, $p < .001$, $\eta_p^2 = .40$ (Fig. 3, right panel); Trial × Moment: $F(2.09, 89.90) = 10.37$, $p < .001$, $\eta_p^2 = .19$. This confirmed the peak-end effect, but no group differences appeared for this interaction, *C3*: $F(1, 43) = .81$, $p = .37$, $\eta_p^2 = .02$; Trial × Group × Moment: $F(2.09, 89.90) = .40$, $p = .68$, $\eta_p^2 = .01$.[1]

[1] The effect of the order of studies was additionally investigated by including Order of studies as a between-subject factor to the repeated measures ANOVAs reported above. The abovementioned effects did not change. A small tendency towards sensitization was observed for the concurrent and retrospective pain ratings, when the pain induction was preceded by the dyspnea induction (concurrent pain ratings: Time segment × Order of studies in short trial, $F(1.62, 64.77) = 3.59$, $p = .04$, $\eta_p^2 = .08$, and in long trial, $F(2.45, 97.87) = 4.60$, $p = .01$, $\eta_p^2 = .10$; retrospective pain ratings: Order of studies, $F(1, 40) = 4.00$, $p = .05$, $\eta_p^2 = .09$). A tendency towards habituation was found for concurrent ratings of dyspnea in the short trial (Time segment × Order of studies, $F(1, 37) = 3.94$, $p = .06$, $\eta_p^2 = .10$). However, no interactions between the order of studies and the group variable emerged.

Affective responses

High HSR had higher state NA than low HSR, $F(1, 44) = 9.71$, $p < .01$, $\eta_p^2 = .18$. The dyspnea induction resulted in higher state NA after the trials than the pain induction (Induction × Moment: $F(2, 88) = 3.80$, $p = .03$, $\eta_p^2 = .08$). State symptom and anxiety ratings followed a similar pattern, as can be seen in Table 1. No group differences regarding the threat value of the trials were observed. However, the short rebreathing trial was overall more threatening than the long trial, while the threat value of pain induction trials did not differ (Induction × Moment: $F(1, 44) = 5.18$, $p = .03$, $\eta_p^2 = .11$). Moreover, mediation analyses (see Supplementary Online Material) showed that for three out of four trials (except for short pain trial) state NA during the trial was a significant mediator of the association between the habitual symptom reporting and retrospective symptom ratings.

Discussion

The present study aimed to investigate the course of retrospective memory for two distinct aversive bodily sensations, i.e. pain and dyspnea, across a 2-week period and whether individual differences in habitual symptom reporting would moderate this course over time. These sensations were experimentally induced in the laboratory to individuals scoring high or low on habitual symptom reporting (HSR). Concurrent symptom ratings collected during the inductions served as a reference point for comparisons with retrospective ratings collected at three fixed time points after the symptom induction. Consistent with previous research, it was found that retrospective memory for symptoms is inaccurate and that the course of bias over time differs between the groups. In addition, differences between the two aversive bodily sensations were also observed.

The manipulation checks indicated a successful pain and dyspnea induction. Nonetheless, it was also found that the groups differed in their concurrent symptom perception, with high HSR reporting more symptoms than low HSR, especially during

dyspnea induction. However, physiological responses to this induction, as measured by MV and $FetCO_2$, were not different. These observations are in line with the findings of Bogaerts et al. (2010) who showed increased concurrent dyspnea ratings in MUD patients compared to healthy controls, despite the lack of differences at the physiological level.

Considering retrospective symptom ratings, two important findings emerged: (1) symptoms are retrospectively biased, and (2) the pattern of bias in symptom recall differs between low and high HSR. With regards to the first finding, retrospective ratings for both pain and dyspnea were overall higher than averaged concurrent evaluations. This accords with previous research on the inaccuracy of symptom memory in general, showing that the overall retrospective evaluation of the experience is usually more aversive than the averaged concurrent reports (e.g. Broderick et al., 2008; Giske et al., 2010; Houtveen & Oei, 2007). It should be noted, however, that previous studies investigating this discrepancy were mostly based on diary protocols and included longer time frames (starting from one day recollections). To our knowledge, our study is the first to show that the global evaluation is biased already immediately after the symptom episode. This discrepancy is often explained by peak/saliency effects (Kahneman et al., 1993; Miron-Shatz, Stone, & Kahneman, 2009; Stone et al., 2005) which suggests that the peak intensity of the experience, due to its aversive and threat-signaling connotation, is more heavily weighted during retrospective assessment of averaged symptoms while other symptom-free moments tend to be disregarded. Findings related to trial differences support this idea: during immediate ratings, the short trials (ending at aversive peak) were consistently rated as more intense than the long trials, including gradual recovery. Interestingly, this difference was found not only when the averaged concurrent ratings of both trials did not differ (dyspnea induction), but also when the long trial caused more concurrent pain than the short one. However, in contrast to earlier findings showing a lack of the peak-end heuristic in high symptom reporters (patients: Bogaerts et al.,

2012; non-clinical HSR: Houtveen & Oei, 2007), no group differences emerged for the peak-end related findings. While this may be due to methodological differences (diary vs. experimental study, peak-end effect measured by forced-choice preferences for the short or long trial vs. by actual symptom ratings), it could also suggest that this heuristic is used to the same extent by both groups. Further work is needed to elucidate this issue.

Secondly, the pattern of bias in retrospective symptom reporting differed between the groups. In pain induction, group differences emerged for immediate evaluations in that high HSR reported having experienced more pain than low HSR, while in dyspnea induction both high and low HSR reported similar levels of dyspnea. This pattern of results could be caused by the differences in symptom intensity and affective reactions to symptom manipulations. In the current study, pain induction was associated with less state symptoms, and lower state NA and anxiety levels than dyspnea induction. We assume that the lower intensity and threat value of pain induction allowed for the increased influence of the trait characteristics, such as habitual symptom reporting or NA. This interpretation is in line with previous findings showing that trait NA and HSR influence somatic complaints especially when symptoms are ambiguous or low in intensity (Bogaerts et al., 2008; De Peuter et al., 2007; Larsen, 1992; Stegen et al., 1998, 2001; Van Diest et al., 2005).

The differences were also observed for the pattern of retrospective symptom reporting over time: while there was no increase in recalled pain over the course of two weeks, dyspnea reports increased over time in high HSR, but not in low HSR group. This corroborates the findings from the diary study by Houtveen and Oei (2007), who observed a rise in retrospective symptoms reporting over time in high HSR participants, but it is currently unclear why increasing bias in symptom recall over time in the high HSR group was specific for dyspnea. One reason could be related to the ratio of sensory and affective processing of bodily information used during retrospective recall. Sensory and affective aspects of somatic

information are processed in parallel (Leventhal, Brown, Shacham, & Engquist, 1979), but a focus on one component may decrease attention to the other. In contrast to sensory processing of bodily signals, which leads to more detailed perception and reduced symptom ratings (Cioffi, 1991; Crane & Martin, 2003), affective processing of usually unpleasant and aversive symptom experience may lead to negative affect biasing both symptom perception and retrospective memory (Bogaerts et al., 2008; Michael & Burns, 2004). It is plausible that high HSR, who also have elevated NA, focus relatively more on the affective aspects of the somatic experience, reducing the influence of the actual sensory input. This way, the level of NA and anxiety experienced during symptom induction could affectively color the memory of experience, resulting in reporting bias. This explanation is to some extent supported by the findings from the mediation analyses, which showed that the relationship between the habitual symptom reporting and retrospective symptom ratings was mediated by the state NA. Moreover, because the pain induction was not as distressing as the dyspnea induction, the negative biasing of the affective component might have been attenuated in this condition. Further research is needed to explore the processing styles adopted by high HSR and their possible influence on symptom memory bias in equally distressing somatic conditions.

The findings from this study make several contributions to the current literature. First, as one of the first studies it has investigated memory for symptoms in high habitual symptom reporters by means of experimentally controlled symptom inductions. The inclusion of both concurrent symptom ratings and physiological responses during controlled symptom inductions together with retrospective evaluations of the same experience extend our knowledge of symptom perception and memory, without being undermined by possible physiological differences. Second, symptom memory was found to be substantially biased already immediately after the end of experience, with relatively little change thereafter. This is important for studies assessing the accuracy of symptom memory with methods such as the

Ecological Momentary Assessment (EMA, Shiffman, Stone, & Hufford, 2008) or the Experience Sampling Method (ESM, Larson & Csikszentmihalyi, 1983). In the majority of studies using EMA or ESM, such immediate ratings would be considered concurrent and relatively unbiased, while our data show that, by then, biases have had most of their effect already. Finally, by investigating two types of symptoms, i.e. pain and dyspnea, in two different groups, our findings point to the role of particular characteristics of the aversive bodily sensations that, in interaction with individual differences, determine retrospective memory for symptoms.

The present study has some limitations. First, the current study used a healthy female HSR sample, which may limit the generalizability of the findings to a clinical MUS population. However, given that our findings concerning dyspnea perception were largely in line with the previous study using the rebreathing paradigm in patients with MUD (Bogaerts et al., 2012), it could be expected that other cognitive processes (i.e. memory) are comparable in those two groups. Second, different concurrent assessment procedures (continuous vs. every 10 seconds) were applied in both studies. Nonetheless, because the participants were inquired to rate an average symptom experience, the influence of the trial duration and frequency of assessment should not have a great impact on the ratings.

In conclusion, the present study documents retrospective memory inaccuracy for symptoms, it replicates the peak-end bias in two different bodily sensations, and it extends our understanding of symptom memory in habitual symptom reporting. The observed increase of retrospective dyspnea reporting over time in high HSR corroborates the role of perceptual-cognitive and memory processes underlying HSR and MUS. Future research is needed to more narrowly specify the precise mechanisms underlying the observed symptom memory distortions in high HSR.

Acknowledgments

The authors want to thank Anne-Marie Verlinden for her assistance in data collection. The study was funded by Grant OT/10/027 from the University Research Council of University of Leuven.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington: VA: American Psychiatric Publishing.

Barsky, A. J., Orav, E. J., & Bates, D. W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Archives of General Psychiatry, 62*, 903–910. doi:10.1001/archpsyc.62.8.903

Bogaerts, K., Millen, A., Li, W., De Peuter, S., Van Diest, I., Vlemincx, E., … Van den Bergh, O. (2008). High symptom reporters are less interoceptively accurate in a symptom-related context. *Journal of Psychosomatic Research*, *65*, 417–724. doi:10.1016/j.jpsychores.2008.03.019

Bogaerts, K., Van Eylen, L., Li, W., Bresseleers, J., Van Diest, I., De Peuter, S., … Van den Bergh, O. (2010). Distorted symptom perception in patients with medically unexplained symptoms. *Journal of Abnormal Psychology*, *119*, 226–234. doi:10.1037/a0017780

Bogaerts, K., Wan, L., Van Diest, I., Stans, L., Decramer, M., & Van den Bergh, O. (2012). Peak-end memory bias in laboratory-induced dyspnea: a comparison of patients with medically unexplained symptoms and healthy controls. *Psychosomatic Medicine*, *74*, 974–981. doi:10.1097/PSY.0b013e318273099c

Borg, G. (1982). Ratings of perceived exertion and heart rates during short-term cycle exercise and their use in a new cycling strength test. *International Journal of Sports Medicine, 3*, 153–158. doi:10.1055/s-2008-1026080

Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *PAIN*, *139*, 146–157. doi:10.1016/j.pain.2008.03.024

Carson, A. J., Ringbauer, B., Stone, J., McKenzie, L., Warlow, C., & Sharpe, M. (2000). Do medically unexplained symptoms matter? A prospective cohort study of 300 new

referrals to neurology outpatient clinics. *Journal of Neurology, Neurosurgery & Psychiatry, 68,* 207–210. doi:10.1136/jnnp.68.2.207

Chajut, E., Caspi, A., Chen, R., Hod, M., & Ariely, D. (2014). In Pain Thou Shalt Bring Forth Children: The Peak-and-End Rule in Recall of Labor Pain. *Psychological Science, 25,* 1–6. doi:10.1177/0956797614551004

Cioffi, D. (1991). Sensory awareness versus sensory impression: Affect and attention interact to produce somatic meaning. *Cognition & Emotion, 5,* 275–294. doi:10.1080/02699939108411041

Crane, C., & Martin, M. (2003). Illness schema and level of reported gastrointestinal symptoms in irritable bowel syndrome. *Cognitive Therapy and Research, 27,* 185–203. doi:10.1023/A:1023561226173

De Peuter, S., Put, C., Lemaigre, V., Demedts, M., Verleden, G., & Van den Bergh, O. (2007). Context-evoked overperception in asthma. *Psychology & Health, 22,* 737–748. doi:10.1080/14768320601151702

Eich, E., Reeves, J. L., Jaeger, B., & Graff-Radford, S. B. (1985). Memory for pain: Relation between past and present pain intensity. *PAIN, 23,* 375–380. doi:10.1016/0304-3959(85)90007-7

Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., & Van den Bergh, O. (2006). Verdere validering van de Positive and Negative Affect Schedule (PANAS) en vergelijking van twee nederlandstalige versies. *Gedrag & Gezondheid, 34,* 89–102.

Everts, B., Karlson, B., Währborg, P., Abdon, N.-J., Herlitz, J., & Hedner, T. (1999). Pain Recollection after Chest Pain of Cardiac Origin. *Cardiology, 92,* 115–120. doi:10.1159/000006958

Feine, J. S., Lavigne, G. J., Thuan Dao, T. ., Morin, C., & Lund, J. P. (1998). Memories of chronic pain and perceptions of relief. *PAIN*, *77*, 137–141. doi:10.1016/S0304-3959(98)00089-X

Gedney, J. J., & Logan, H. (2004). Memory for stress-associated acute pain. *The Journal of Pain*, *5*, 83–91. doi:10.1016/j.jpain.2003.11.005

Giske, L., Sandvik, L., & Røe, C. (2010). Comparison of daily and weekly retrospectively reported pain intensity in patients with localized and generalized musculoskeletal pain. *European Journal of Pain*, *14*, 959–965. doi:10.1016/j.ejpain.2010.02.011

Houtveen, J. H., & Oei, N. Y. L. (2007). Recall bias in reporting medically unexplained symptoms comes from semantic memory. *Journal of Psychosomatic Research*, *62*, 277–282. doi:10.1016/j.jpsychores.2006.11.006

Hunter, M., Philips, C., & Rachman, S. (1979). Memory for pain. *PAIN*, *6*, 35–46. doi:10.1016/0304-3959(79)90138-6

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a Better End. *Psychological Science*, *4*, 401–405. doi:10.1111/j.1467-9280.1993.tb00589.x

Khan, A. A., Khan, A., Harezlak, J., Tu, W., & Kroenke, K. (2003). Somatic symptoms in primary care: etiology and outcome. *Psychosomatics*, *44*, 471–478. doi:10.1176/appi.psy.44.6.471

Larsen, R. J. (1992). Neuroticism and selective encoding and recall of symptoms: Evidence from a combined concurrent-retrospective study. *Journal of Personality and Social Psychology*, *62*, 480–488. doi:10.1037/0022-3514.62.3.480

Larson, R., & Csikszentmihalyi, M. (1983). The Experience Sampling Method. *New Directions for Methodology of Social & Behavioral Science*, *15*, 41–56.

Lefebvre, J. C., & Keefe, F. J. (2002). Memory for Pain: The Relationship of Pain Catastrophizing to the Recall of Daily Rheumatoid Arthritis Pain. *The Clinical Journal of Pain, 18*, 56–63. doi:10.1097/00002508-200201000-00009

Leventhal, H., Brown, D., Shacham, S., & Engquist, G. (1979). Effects of preparatory information about sensations, threat of pain, and attention on cold pressor distress. *Journal of Personality and Social Psychology, 37*, 688–714. doi:10.1037//0022-3514.37.5.688

Levine, L. J., & Safer, M. A. (2002). Sources of Bias in Memory for Emotions. *Current Directions in Psychological Science, 11*, 169–173. doi:10.1111/1467-8721.00193

Linton, S. J., & Melin, L. (1982). The accuracy of remembering chronic pain. *PAIN, 13*, 281–285. doi:10.1016/0304-3959(82)90017-3

Meek, P. M., Lareau, S. C., & Anderson, D. (2001). Memory for symptoms in COPD patients: how accurate are their reports? *European Respiratory Journal, 18*, 474–481. doi:10.1183/09031936.01.00083501

Michael, E. S., & Burns, J. W. (2004). Catastrophizing and pain sensitivity among chronic pain patients: moderating effects of sensory and affect focus. *Annals of Behavioral Medicine, 27*, 185–194. doi:10.1207/s15324796abm2703_6

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of yesterday's emotions: does the valence of experience affect the memory-experience gap? *Emotion, 9*, 885–891. doi:10.1037/a0017823

Nimnuan, C., Rabe-Hesketh, S., Wessely, S., & Hotopf, M. (2001). How many functional somatic syndromes? *Journal of Psychosomatic Research, 51*, 549–557. doi:10.1016/S0022-3999(01)00224-0

Read, D. J. C. (1967). A clinical method for assessing the ventilatory response to carbon dioxide. *Australasian Annals of Medicine, 16*, 20–32.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *PAIN*, *66*, 3–8. doi:10.1016/0304-3959(96)02994-6

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *PAIN*, *104*, 187–194. doi:10.1016/S0304-3959(03)00003-4

Rief, W., & Broadbent, E. (2007). Explaining medically unexplained symptoms-models and mechanisms. *Clinical Psychology Review*, *27*, 821–841. doi:10.1016/j.cpr.2007.07.005

Safer, M. A., Levine, L. J., & Drapalski, A. L. (2002). Distortion in Memory for Emotions: The Contributions of Personality and Post-Event Knowledge. *Personality and Social Psychology Bulletin*, *28*, 1495–1507. doi:10.1177/014616702237577

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. doi:10.1146/annurev.clinpsy.3.022806.091415

Smith, W. B., & Safer, M. A. (1993). Effects of present pain level on recall of chronic pain and medication use. *PAIN*, *55*, 355–361. doi:10.1016/0304-3959(93)90011-D

Sohl, S. J., & Friedberg, F. (2008). Memory for fatigue in chronic fatigue syndrome: relationships to fatigue variability, catastrophizing, and negative affect. *Behavioral Medicine*, *34*, 29–38. doi:10.3200/BMED.34.1.29-38

Stegen, K., Neujens, A., Crombez, G., Hermans, D., Van de Woestijne, K. P., & Van den Bergh, O. (1998). Negative affect, respiratory reactivity, and somatic complaints in a CO_2 enriched air inhalation paradigm. *Biological Psychology*, *49*, 109–122. doi:10.1016/S0301-0511(98)00030-1

Stegen, K., Van Diest, I., Van de Woestijne, K. P., & Van den Bergh, O. (2000). Negative affectivity and bodily sensations induced by 5.5% CO_2 enriched air inhalation: Is there a

bias to interpret bodily sensations negatively in persons with negative affect? *Psychology & Health, 15*, 513–525. doi:10.1080/08870440008402010

Stegen, K., Van Diest, I., Van de Woestijne, K. P., & Van den Bergh, O. (2001). Do persons with negative affect have an attentional bias to bodily sensations? *Cognition & Emotion, 15*, 813–829. doi:10.1080/02699930143000211

Stone, A. A., Schwartz, J. E., Broderick, J. E., & Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Personality & Social Psychology Bulletin, 31*, 1340–1346. doi:10.1177/0146167205275615

Suls, J., & Howren, M. B. (2012). Understanding the Physical-Symptom Experience: The Distinctive Contributions of Anxiety and Depression. *Current Directions in Psychological Science, 21*, 129–134. doi:10.1177/0963721412439298

Van den Bergh, O., Bogaerts, K., & Van Diest, I. (in press). Symptom Perception, Awareness and Interpretation. In J. D. Wright (Ed.), *The International Encyclopedia of Social and Behavioral Sciences* (2nd Edition). Oxford, UK: Elsevier Ltd.

Van Diest, I., De Peuter, S., Eertmans, A., Bogaerts, K., Victoir, A., & Van den Bergh, O. (2005). Negative affectivity and enhanced symptom reports: differentiating between symptoms in men and women. *Social Science & Medicine, 61*, 1835–1845. doi:10.1016/j.socscimed.2005.03.031

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070. doi:10.1037//0022-3514.54.6.1063

Wientjes, C. J. E., & Grossman, P. (1994). Overreactivity of the psyche or the soma? Interindividual associations between psychosomatic symptoms, anxiety, heart rate, and end-tidal partial carbon dioxide pressure. *Psychosomatic Medicine, 56*, 533–540.

Figure 1. Mean values and standard errors of concurrent pain ratings (0-100) for high and low habitual symptom reporters (HSR) during the short (left) and the long trial (right).

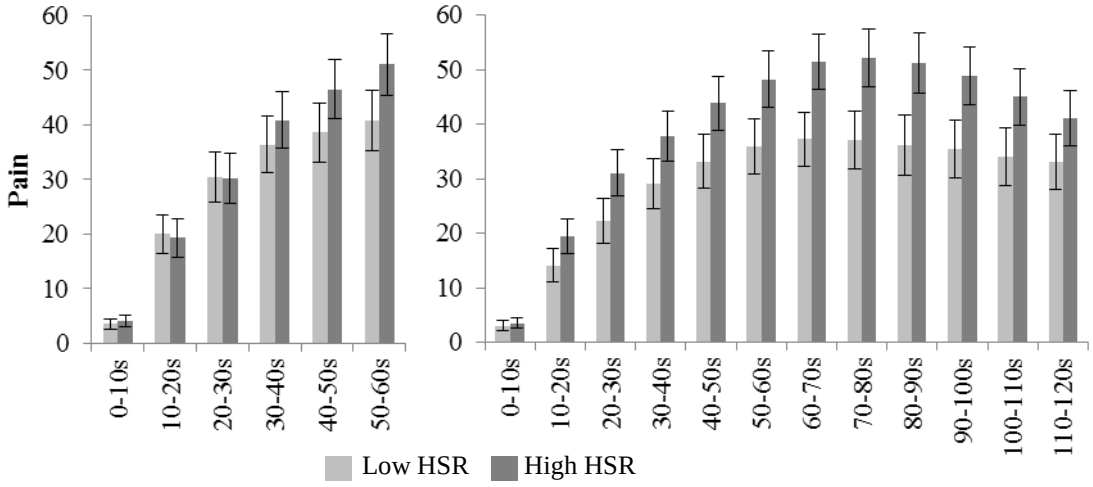

Figure 2. Mean averaged concurrent and retrospective pain ratings (0-100, left) and dyspnea ratings (0-100, right) for high and low habitual symptom reporters (HSR). Whiskers denote standard errors.

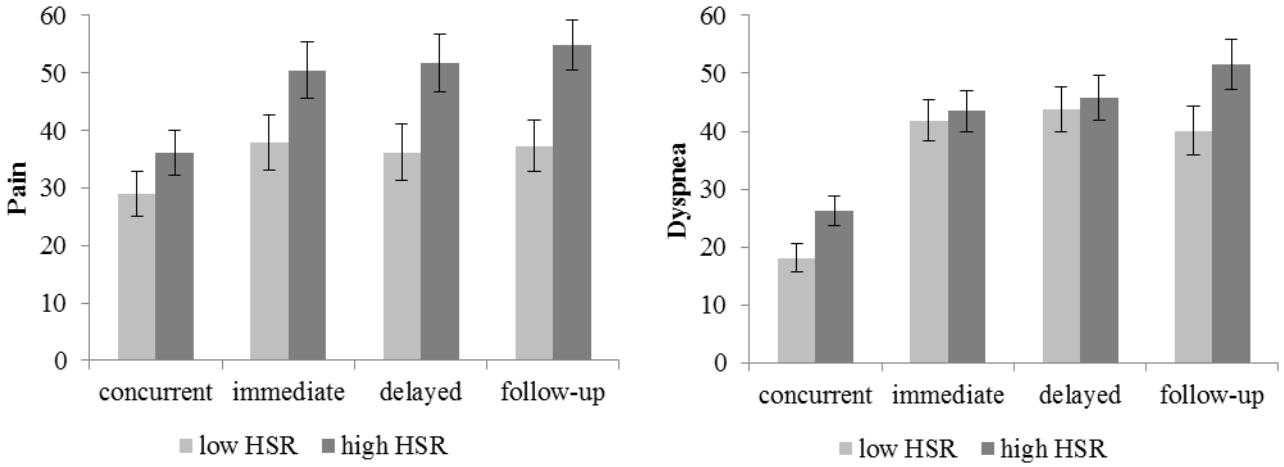

Figure 3. Mean averaged concurrent and immediate pain ratings (left) and dyspnea ratings (right) for short and long trial. Whiskers denote standard errors.

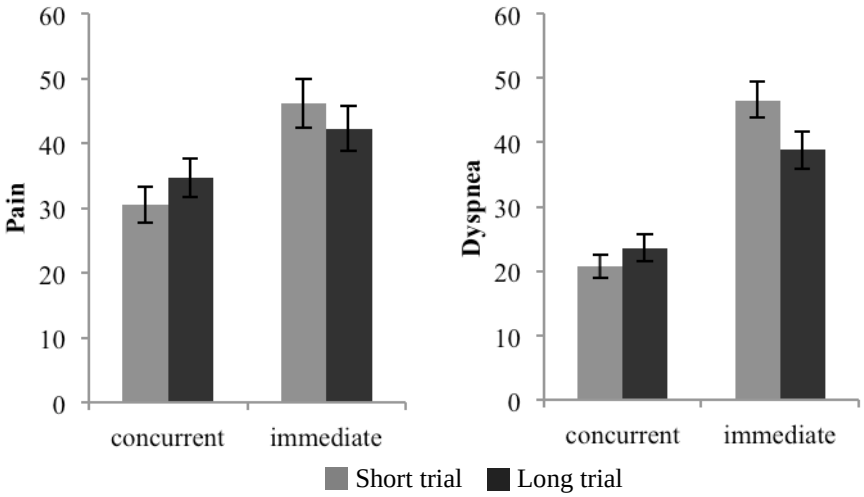

Figure 4. Mean values and standard errors of concurrent dyspnea (0-100), mean fractional end-tidal concentration of CO_2 ($FetCO_2$) and minute ventilation for high and low habitual symptom reporters (HSR) in baseline, rebreathing and recovery phase for the short (left) and the long trial (right).

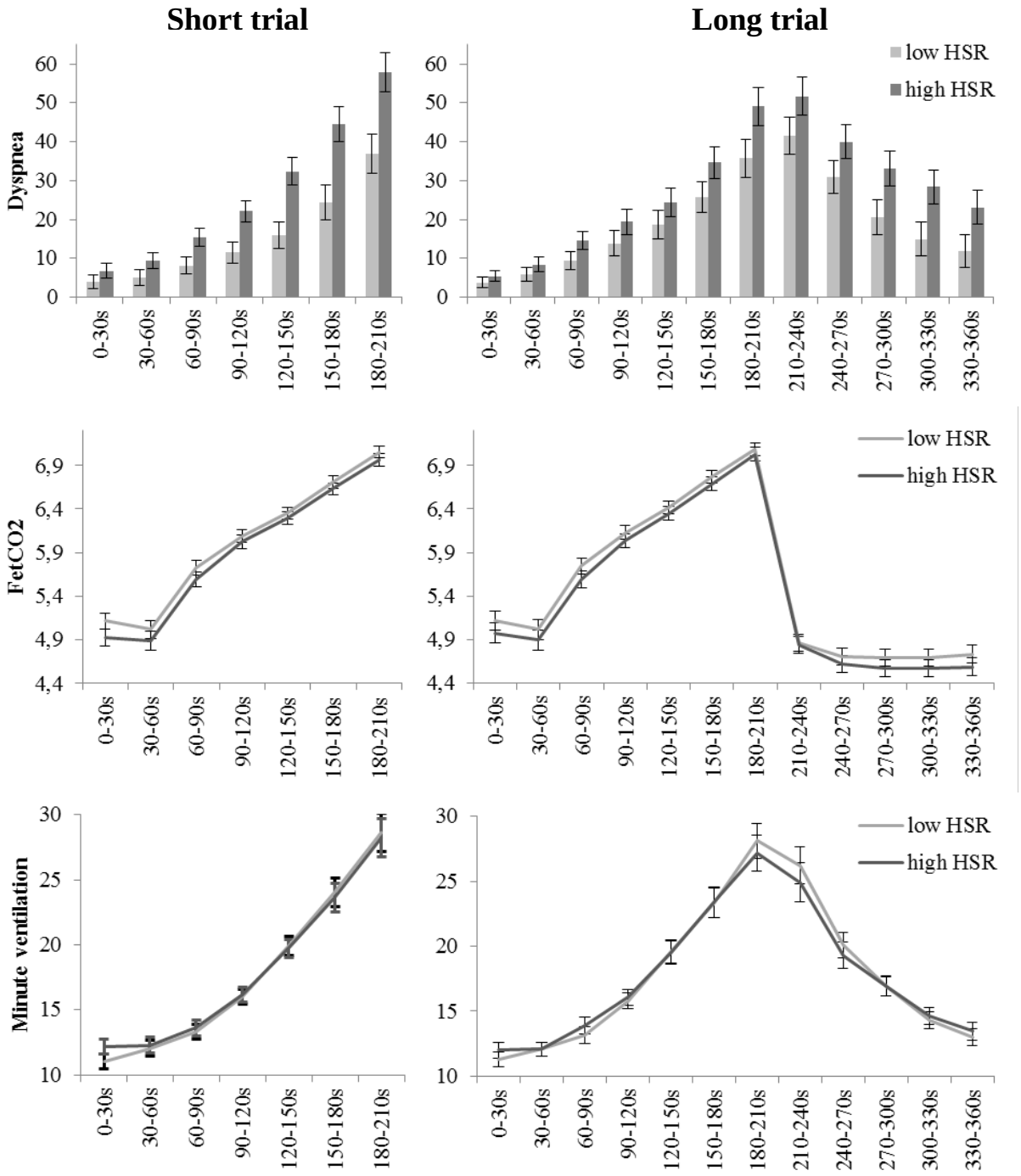

Table 1. Means and SDs for state symptoms, state NA, state anxiety and threat value, and significant effects of Repeated Measures ANOVA for all dependent variables.

Measure	Group	StudyPain			StudyDyspnea			Significant effects (F;df)
		Baseline	Short trial	Long trial	Baseline	Short trial	Long trial	
State symptoms	low HSR							IND*** (78.39; 1) MOM*** (27.87; 1.59) INDxMOM*** (41.74; 1.53) HSR*** (35.91; 1)
	M	1.38	2.67	2.75	1.67	8.50	8.08	
	SD	1.31	2.08	2.17	1.43	4.94	4.93	
	high HSR							
	M	7.25	6.83	6.54	7.42	15.13	13.21	
	SD	5.24	5.24	3.99	4.68	6.02	6.46	
State NA	low HSR							IND*** (20.49; 1) MOM*** (15.70; 2) INDxMOM* (3.80; 2) HSR** (9.71; 1)
	M	12.38	13.17	13.17	13.33	15.00	14.79	
	SD	2.53	3.13	3.47	3.07	4.60	3.62	
	high HSR							
	M	15.21	16.96	16.71	16.13	20.21	19.50	
	SD	5.35	6.04	5.69	5.17	6.78	6.54	
State anxiety	low HSR							IND*** (18.87; 1) MOM*** (50.72; 2) MOMxOrder* (3.78; 2) INDxMOM** (5.67; 1.77) HSR* (7.30; 1)
	M	1.67	2.58	2.58	2.04	3.79	3.71	
	SD	0.82	1.67	1.79	1.23	1.96	1.99	
	high HSR							
	M	2.58	3.71	4.04	2.88	5.25	4.67	
	SD	1.50	2.20	2.22	1.94	2.13	2.16	
Threat value	low HSR							IND* (6.50; 1) MOM** (7.40; 1) INDxMOM* (5.18; 1) Order* (7.06; 1)
	M		3.54	3.17		4.96	4.25	
	SD		2.25	2.04		1.88	2.09	
	high HSR							
	M		4.54	4.54		5.42	4.50	
	SD		2.59	2.50		2.17	2.47	

Note: HSR = Habitual symptom reporting; IND = Induction; MOM = Moment of measurement.
*p<.05, **p<.01, ***p<.001.