



Reducing body dissatisfaction by means of an evaluative conditioning procedure in undergraduate women: A replication study



Klaske A. Glashouwer^{a,b,*}, Irina Masselman^a, Peter J. de Jong^a

^a Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Grote Kruisstraat 2/1, 9712, TS, Groningen, the Netherlands

^b Department of Eating Disorders, Accare Child and Adolescent Psychiatry, Postbus 660, 9700, AR, Groningen, the Netherlands

ARTICLE INFO

Keywords:

Body image
Body dissatisfaction
Evaluative conditioning
Intervention
Replication

ABSTRACT

The aim was to investigate whether a computer-based evaluative conditioning (EC) procedure using positive social feedback is effective in enhancing body satisfaction. Prior findings in three small-scale studies were mixed showing positive effects in pre-clinical samples, but not in a clinical sample of eating disorder patients. Therefore, our main goal was to replicate the original finding of Martijn et al. (2010) in a well-powered unselected sample of 129 female undergraduates. We assessed the impact of EC on questionnaire measures of body satisfaction as well as on affective ratings of the participants' body pictures used in the task to verify whether the EC procedure was effective in heightening the subjective valence of these pictures. Supporting the validity of the current EC procedure, participants in the experimental condition rated their own pictures after the training as more positive than participants in the control condition. However, this effect of the EC procedure did not transfer to the self-report indices of body satisfaction. In addition, women with relatively high body concern did not profit more from the EC procedure than women with relatively low body concern. Together, these findings suggest that the EC procedure in its present form is not ready for use as an intervention for improving body satisfaction. However, it seems worthwhile to investigate in future studies how the EC procedure can be further strengthened in a way that effects on stimulus ratings eventually also 'spill over' to the level of self-reported body satisfaction.

1. Introduction

In Western societies, a significant part of the female population experiences negative thoughts and feelings towards their own body, in particular with respect to their weight and shape (e.g., Fallon, Harris, & Johnson, 2014; Tiggemann, 2004). A negative body image expresses itself in negative body-related perceptions (e.g., overestimation of body size) and behaviors (e.g., body checking and avoidance behaviors; Cash, 2011; Cash & Pruzinsky, 1990) as well as in negative cognitions and emotions regarding the own body. The high prevalence of negative body image is concerning considering the serious consequences in terms of physical and mental health (e.g., Grogan, 2006; Johnson & Wardle, 2005; Paxton, Neumark-Sztainer, Hannan, & Eisenberg, 2006; Wilson, Latner, & Hayashi, 2013). Although some studies have shown substantial reductions in negative body image following body image interventions (e.g., Butters & Cash, 1987; Griffen, Naumann, & Hildebrandt, 2018; McLean, Paxton, & Wertheim, 2011), a recent meta-analysis showed that once corrections for several sources of bias were

applied, stand-alone interventions for body image only led to small improvements (Alleva, Sheeran, Webb, Martijn, & Miles, 2015). This points to the need to further advance current treatments for negative body image.

As a novel approach, Martijn, Vanderlinden, Roefs, Huijding, and Jansen (2010) successfully used an evaluative conditioning (EC) procedure to enhance participants' body satisfaction. The current study was designed as a replication of the study of Martijn et al. to examine the robustness of the promising effects of this newly developed procedure to improve people's negative body image. EC is defined as "a change in the liking of a stimulus (conditioned stimulus; CS) that results from pairing that stimulus with other positive or negative stimuli (unconditioned stimulus; US)" (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010, p. 390). EC has already been examined in several domains using various paradigms and has also been applied towards the self (De Houwer, Thomas, & Baeyens, 2001; Hofmann et al., 2010). It was shown that positive automatic self-associations can be strengthened by systematic pairings of self-relevant stimuli (CS) with positive

* Corresponding author. Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Grote Kruisstraat 2/1, 9712, TS, Groningen, the Netherlands.

E-mail addresses: k.a.glashouwer@rug.nl (K.A. Glashouwer), i.masselman@rug.nl (I. Masselman), p.j.de.jong@rug.nl (P.J. de Jong).

<https://doi.org/10.1016/j.brat.2019.103435>

Received 21 March 2019; Received in revised form 16 May 2019; Accepted 1 July 2019

Available online 02 July 2019

0005-7967/ © 2019 Elsevier Ltd. All rights reserved.

stimuli (US) (Baccus, Baldwin, & Packer, 2004; Dijksterhuis, 2004; Riketta & Dauenheimer, 2003). Based on these promising findings, subsequent research used an EC procedure to enhance body satisfaction in a female undergraduate student sample (Martijn et al., 2010). In this EC task images of the own body and of other people's bodies were presented in one of four quadrants on a computer screen (cf. Baccus et al., 2004). Upon its presentation, participants had to click as fast as possible on the image. Following this response, the body picture disappeared and was replaced by a short presentation of a face with an emotional expression. In the experimental condition, pictures of the own body (CS) were consistently followed by pictures of smiling faces (US), whereas photographs of control persons were followed by pictures of neutral or frowning faces. The purpose of this EC procedure was to enhance body satisfaction via strengthening the association between participants' appearance and social approval/acceptance.

So far, three small-scale studies have tested the effect of this EC procedure on body image using different designs (Aspen et al., 2015; Glashouwer, Neimeijer, de Koning, Vestjens, & Martijn, 2018; Martijn et al., 2010). The procedure was first tested as a single session training in a controlled laboratory setting among 54 unselected female students who were randomly assigned to the experimental or placebo-control condition (Martijn et al., 2010). In the control condition, participants were shown the same stimuli, but here, all body pictures were randomly followed by pictures of smiling, neutral, and frowning faces, so that the own body was not systematically paired with smiling faces. Results showed that specifically for participants with relatively high body concerns, body satisfaction and general self-esteem increased following the active EC procedure. Subsequently, the procedure was tested in a field experiment among 39 female students at risk for developing an eating disorder (Aspen et al., 2015). This study was designed as a randomized waitlist-controlled trial in which the experimental group received four sessions of the EC training over a 4-week period. The training sessions were administered in a controlled setting under supervision. Women in the experimental group ($n = 22$) showed a decrease in shape and weight concern as well as an increase in self-esteem following the training procedure, as compared to those in the waitlist-control group. Improvements with respect to body image were maintained at 4-week and 12-week follow-ups. Finally, a clinical trial was conducted in which 51 healthy weight adolescent girls with an eating disorder were randomly assigned to the experimental condition or a placebo-control condition (Glashouwer et al., 2018). Participants in this study received six online training sessions over a 3-week period. In the control condition, participants were shown the same stimuli as in the experimental condition, but here, a stimulus was always followed by another stimulus from the same category, so that the own body was not paired with smiling faces. Although this latter study used a relatively intense EC procedure (six sessions instead of one or four session), it failed to find any effects of the EC intervention on self-reported body satisfaction, weight and shape concern, general self-esteem, or automatic appearance associations. The study of Glashouwer et al. (2018) differed in several respects from the other two studies, most importantly by using an online training instead of a controlled laboratory situation, allowing participants to wear their own clothes instead of standardized clothes during the photoshoot, and testing an adolescent rather than a young adult sample. Nevertheless, these conflicting results cast doubts on the robustness of the earlier positive findings. Especially, since these earlier studies relied on relatively small sample sizes and were therefore sensitive to chance findings. For an adequate appreciation of the current EC procedure as a tool to enhance body satisfaction it seems therefore critical to replicate the original finding of Martijn et al. (2010) in a well-powered study.

Accordingly, the primary aim of the present study was to conduct an exact replication of the study of Martijn et al. (2010) in a well-powered sample of female participants. To test whether any improvement in body satisfaction could indeed be attributed to the efficacy of the EC procedure to improve the affective evaluation of the CS, the current

study not only assessed the impact of EC on participants' body satisfaction, but also on the affective appreciation of the CS-pictures per se. If the EC procedure would not be effective in modifying the valence of participants' body pictures (CS-pictures), this would also preclude any transfer of positive conditioning effects to more distal outcomes, such as participants' body dissatisfaction.

In short, we tested whether: i) participants in the experimental condition in which pictures of one's own body were systematically paired with stimuli signaling social acceptance (i.e., smiling faces) report higher levels of body satisfaction after the training compared to participants in the placebo-control condition; ii) women with relatively high body concern profit more from the evaluative conditioning intervention than women with relatively low body concern. In line with the original design, we also tested the effect of the training on the secondary outcome measures global self-esteem, state social self-esteem, state performance self-esteem, and mood. In addition to the original design, we assessed the self-reported valence of the stimuli used in the EC task to be able to test the critical prediction that after the training, participants' affective ratings of their own pictures would be more positive in the experimental condition than in the control condition, and that this effect would be especially pronounced for participants with high body concern.

2. Method

2.1. Participants

A total of 131 undergraduate female Psychology students (M age = 20.03 years, $SD = 2.12$) participated in this study in exchange for course credit points. Apart from sex and good comprehension of the English language, there were no further selection criteria (cf. Martijn et al., 2010). During analyses, data from two participants were excluded. For one participant the interval between test sessions exceeded the maximum amount of 8 days, while another participant refused to wear the standardized clothing due to religious reasons. The eventual sample consisted of 129 participants who were randomly assigned to the experimental condition ($N = 67$) or the control condition ($N = 62$). The current study was approved by the Ethical Committee Psychology of the University of Groningen (18052-S) and all individual participants actively gave informed consent before the start of the study.

2.2. Materials and measures

2.2.1. Evaluative conditioning task

Participants were instructed to click on body pictures appearing randomly at the center of one of four quadrants on a computer screen (1920 x 1080 IYAMA, refresh rate = 60 Hz). Once participants had clicked on the stimulus, a second feedback picture was presented at the same location (see Fig. 1). After 400 ms, the feedback picture disappeared and the next trial started immediately. In total, the EC task consisted of 270 trials. The statistical contingency between the stimulus pictures (CS) and feedback pictures (US) differed as a function of the condition that a participant was assigned to. In the experimental condition, pictures of the participant's body were always followed by a smiling face (100%), whereas pictures of control bodies were always followed by either a neutral (50%) or frowning face (50%). In the control condition, the participant's own body pictures and control body pictures were each followed by an equal amount of smiling (33.3%), neutral (33.3%), and frowning (33.3%) faces.

2.2.2. Stimuli

We used two types of CSs; full-body pictures of the participant (90 trials) and full-body pictures of two control women (180 trials). Pictures were taken from the front, left lateral and right lateral, resulting in 9 different CSs. Both the control women and the participants were photographed with a Panasonic DC-FZ82 camera, standing in

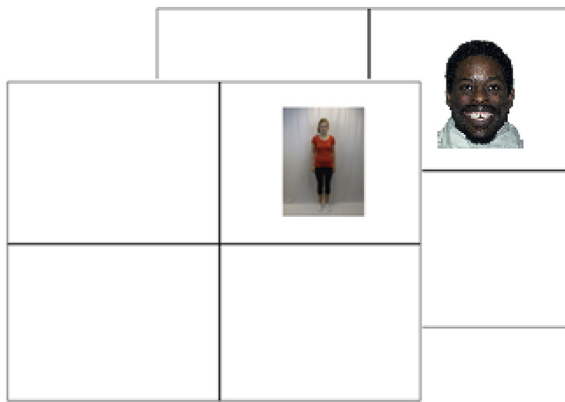


Fig. 1. Example of a trial from the Evaluative conditioning task. After participants clicked on the full-body picture (CS), a smiling, neutral, or frowning face appeared for 400 ms at the same location (US). CS-US contingency depended on the condition that participants were assigned to.

front of a white background and always in the same position (i.e. arms along the body and with a neutral facial expression). The control women had a healthy Body Mass Index (BMI = kg/m²; control 1: BMI = 20.92; control 2: BMI = 21.85), and were comparable to the participants in age (control 1: age = 23 years; control 2: age = 25 years). To help participants recognize their own body during the task, the control women always wore a t-shirt in an opposite color to that of the participant. More specifically, if participants wore a black sports leggings and a black t-shirt, the control women wore the same leggings and a pink t-shirt, and vice versa (as was the case in Martijn et al., 2010). T-shirt color was counterbalanced between participants.

As USs we used the same 45 face pictures (22 male and 23 female faces) from the NimStim Facial Stimuli Set¹ (Tottenham et al., 2009) as were used in Martijn et al. (2010). These USs consisted of an equal amount of happy, neutral, and frowning faces. Although the CSs were presented at random, randomization of the USs was limited due to the reinforcement schedule. However, within each category (i.e. smiling, neutral, frowning) USs were randomly presented.

2.3. Questionnaires

2.3.1. Body satisfaction

Body satisfaction was measured using two items hidden in a questionnaire measuring satisfaction with various life domains, such as housing and social life (cf. Martijn et al., 2010). This life satisfaction questionnaire consisted of 13-items, for which answers were given on a 100 mm VAS scale (“At this moment I feel; 0 = not at all satisfied and 100 = very satisfied”). More specifically, body satisfaction was assessed by the items “At this moment I feel, not at all/very satisfied with my body” and “At this moment I feel, not at all/very satisfied with my weight”. Consistent with the study of Martijn et al. (2010), the two body satisfaction items were highly correlated ($r = .84, p < .001$ at Session 1; $r = .92, p < .001$ at Session 2). Consequently, a mean body satisfaction score was calculated by averaging the scores of these two items.

For exploratory reasons, body image was also measured using the Body Image States Scales (BISS; Cash, Fleming, Alindogan, Steadman, & Whitehead, 2002). The BISS assesses evaluative/affective body image states using 6 items. For each item, participants indicate on a 9-point Likert scale how satisfied they are with a specific aspect of their body (1 = *extremely dissatisfied* to 9 = *extremely satisfied*) or how they

evaluate that aspect of their body in comparison to other people (1 = *a great deal worse* to 9 = *a great deal better*). BISS scores were the mean of the six items after reverse-scoring the three items for which the answer dimensions were positive to negative. Since the BISS was not included in the original procedure, we decided to only administer the BISS towards the end of session 2. The BISS showed a good internal consistency in the present sample (Cronbach's $\alpha = .82$).

2.3.2. State self-esteem

Temporary changes in self-esteem were measured using the State Self-Esteem Scale (SSES; Heatherton & Polivy, 1991). The SSES consists of 20 items, which are answered on a 5-point Likert scale. The SSES measures state self-esteem using three subscales, namely the Appearance, Social and Performance subscale. Each subscale represents a domain that is thought to be sensitive to fluctuations in self-esteem. The Appearance subscale is of particular interest for the present study. An example of an item from this subscale is “At this moment, I feel unattractive; 1 = not at all to 5 = extremely”. The other subscales were included for comparability. After recoding the negatively phrased items, subscale scores can be obtained by summing the relevant items for each domain. Higher subscale scores indicate higher self-esteem in that specific domain and scores on the Appearance subscale can range from 6 to 30. For both test sessions, the internal consistency of the subscales was high (Cronbach's $\alpha > .80$ for all subscales).

2.3.3. Global self-esteem

The Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965) was administered as a measure of global self-esteem. The RSES comprises 10 statements for which participants have to reflect on their current feelings of self-worth. Participants indicate on a 4-point Likert scale how much they agree with each statement (*strongly disagree* – *strongly agree*). After recoding the negatively phrased items, a measure of global self-esteem can be obtained by summing up the 10 items. Scores range from 10 to 40 and higher scores indicate higher global self-esteem. The RSES showed a high internal consistency for both session 1 and 2 (Cronbach's α of .86 and .89, respectively).

2.3.4. Mood

Current mood was assessed by two items which were, just like the body satisfaction items, hidden in the aforementioned questionnaire measuring life satisfaction. Responses on the mood items were given on a 100 mm VAS scale (“At this moment I feel; 0 = not at all cheerful, 100 = very cheerful” and “My mood at this moment is; 0 = very bad, 100 = very good”). Since the mood items were highly correlated ($r = .83, p < .001$ at Session 1; $r = .80, p < .001$ at Session 2), scores on these items were averaged. Higher scores indicate a more positive mood.

2.3.5. Body concern

In line with Martijn et al. (2010), body concern was indexed with the 6th version of the Eating Disorder Examination Questionnaire (EDE-Q; Fairburn & Beglin, 2008). The EDE-Q is widely used as a screening instrument for eating disorders and is considered a valid self-report measure of body concern (Luce & Crowther, 1999; Wilfley, Schwartz, Spurrell, & Fairburn, 1997). It measures eating disordered thoughts and behaviors in the previous 28 days (i.e. 4 weeks). For most items, responses are given on a 7-point Likert scale that either measures symptom frequency (*No days* – *Every day*) or symptom severity (0 = *not at all* – 6 = *markedly*). The EDE-Q consists of four subscales, namely Restraint (5 items), Shape concern (8 items), Weight concern (5 items), and Eating concern (5 items). The internal consistency of the subscales was high (Cronbach's $\alpha > .80$ for all subscales). Subscale scores can be calculated by averaging the scores of items that belong to each subscale. Since Shape concern and Weight concern subscales were highly correlated ($r = .89, p < .001$), participants' scores on these subscales were averaged to obtain a measure of body concern (BC; cf. Martijn et al.,

¹ The MacBrain Face Stimulus Set was developed by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at. nlt7@columbia.edu for information about the stimulus set.

2010). In addition, the total EDE-Q score was calculated by averaging the four subscales. A higher total score indicates more severe eating disorder pathology.

2.3.6. Stimulus valence

Participants rated all body and face stimuli (i.e., CSs and USs) that were used in the evaluative conditioning task. Stimuli were presented one-by-one with a Visual Analogue Scales (VAS) directly below each picture. CSs (i.e., 3 full-body pictures of the participant and 3 full-body pictures of each of the two control women) were randomly presented and rated on two different dimensions. First, CSs were rated on a *negative-positive* dimension. Next, the same 9 stimuli were rated on an *unattractive-attractive* dimension. Finally, participants were randomly presented with the 45 US pictures (i.e., face pictures), which they rated on a *negative-positive* dimension.

2.4. Procedure

2.4.1. Session 1

Consistent with the study of Martijn et al. (2010), participants were informed that the current study aimed to measure the reliability of certain clinical questionnaires. Prior to testing, they were also informed that pictures would be taken of them in standardized clothing and that these pictures would later be part of two computer tasks (i.e., the evaluative conditioning task and the rating of the stimuli). All participants were tested individually. Upon arrival at the lab, participants were seated behind a computer on which they filled out the questionnaires. All questionnaires were presented in the following order: life satisfaction questionnaire, SSES, EDE-Q, and RSES. Between the life satisfaction questionnaire and SSES, participants were presented with two additional questionnaires that were included in order to keep the procedure similar to that of Martijn et al. (2010). After the questionnaires, participants were asked to change into standardized clothing provided by the researcher. Size labels were covered with black tape. After participants had posed for three pictures, they were asked to change back into their own clothing and were sent home.

2.4.2. Session 2

After 7 days (± 1 day), participants were invited back to the lab and randomly assigned to either the experimental or control condition. Participants first performed the EC task. Subsequently, participants once again filled out the questionnaires with the exception of the EDE-Q, which was not administered again. The order of the questionnaires was kept consistent over test sessions. Next, the BISS was administered and the VAS ratings of the USs and CSs. Finally, participants were asked to guess what the purpose of the study was, in order to measure hypothesis awareness. In addition, participants estimated the percentage of smiling, neutral, and frowning faces that followed their own and the control pictures (for each picture type, estimates had to round up to 100%). These estimates were used as a measure of contingency awareness. Participants' weight and length were measured, before they were sent home. Both test sessions took approximately 30 min.

2.5. Statistical analysis

2.5.1. Primary and secondary outcome measures

To test the effect of the evaluative conditioning intervention on body satisfaction two separate RM-ANOVAs were conducted with VAS body satisfaction or appearance self-esteem as dependent variable, Time (pre-intervention, post-intervention) as within-subjects factor, and Condition (experimental, control) and Body concern (low, high) as between-subjects factors. We decided to conduct two separate RM-ANOVAs instead of MANOVAs to enhance comparability with Martijn et al. (2010). Participants were divided into a low and high BC group by means of a median split on EDE-Q Body concern (cf. Martijn et al., 2010). The two-way interaction of Condition x Time was critical to test

whether participants who received the experimental training report higher levels of body satisfaction after the training than individuals in the control group. The three-way interaction of Condition x Body concern x Time was critical to test whether women with relatively high body concern profit more from the intervention than women with relatively low body concern. Four similar RM-ANOVAs were conducted with the secondary outcome measures as dependent variables (global self-esteem, mood, performance state self-esteem, social state self-esteem). To correct for multiple testing, alpha criterion was set at .025 ($p = .05/2$) for the primary analyses. Alpha criterion for the secondary outcome measures was set at .05.

2.5.2. Stimulus valence

To test the effect of the training on participants' affect towards their own full-body pictures, a RM-ANOVA was conducted with Condition (experimental, control) and Body concern (low, high) as between-subjects factors, and CS type (own picture, control picture) and VAS dimension (positive, attractive) as within-subjects factors. The two-way interaction of Condition x CS type was critical to test the assumption that after the training participants' affect towards their own full-body pictures is more positive in the experimental condition than in the control condition. The three-way interaction of Condition x Body concern x CS type was critical to test whether such an effect would be particularly prominent in women with high body concern. Alpha criterion for these tests was set at .05.

2.5.3. Power calculation

A priori power analysis indicated that a sample size of 98 participants would provide a power of .80 to detect an effect of the same size as Martijn et al. (2010) ($\eta_p^2 = .091$) for the primary outcome variables with an alpha level of .025. Our sample size can therefore be considered satisfactory.

3. Results

3.1. Participant characteristics

Table 1 provides an overview of the means and standard deviations for the group characteristics and outcome measures at all assessment points. Length and weight measurements in the lab indicated that mean BMI for the total sample was within the range indicative for a healthy body weight ($M = 22.5$, $SD = 3.05$), although 25 individuals did exceed what would be considered a healthy BMI ($BMI > 25$) and 11 participants were considered underweight ($BMI < 18.5$). For 3 participants, the total EDE-Q score fell above 4, which is indicative for an eating disorder (control condition: $N = 1$; experimental condition: $N = 2$). As planned, we distinguished a low and high body concern (BC) group by means of a median split on the EDE-Q Body concern scale. Participants with a body concern score equal to or below 1.19 (scores ranged from 0 to 5.94) were assigned to the low BC group, while the rest was assigned to the high BC group. Low and high BC individuals were evenly distributed over conditions, even though we did not control for this beforehand (control condition: low BC, $N = 31$; high BC, $N = 31$; experimental condition: low BC, $N = 34$, high BC, $N = 33$).

A 2 Condition (experimental, control) x 2 BC (low, high) ANOVA was conducted to check for group differences in BMI. Furthermore, multiple non-parametric tests (Kruskal-Wallis H) were performed to compare Conditions and BC groups on age and general eating disorder pathology. Results indicated that participants in the experimental and control condition did not significantly differ in age ($\chi^2(1) = 0.78$, $p = .38$), BMI ($F(1, 124) = 1.23$, $p = .27$), and total EDE-Q score ($\chi^2(1) = 0.026$, $p = .87$). The BC groups were comparable in mean age ($\chi^2(1) = 0.84$, $p = .36$), but not in BMI ($F(1, 124) = 10.8$, $p = .001$) and total EDE-Q ($\chi^2(1) = 90.3$, $p < .001$). As expected, the group high in BC scored on average higher on BMI and eating disorder symptoms than the group low in BC. However, these differences were

Table 1

Means and standard deviations at all assessment points per group.

	Experimental condition				Control condition			
	Low BC (N = 34)		High BC (N = 33)		Low BC (N = 31)		High BC (N = 31)	
	Session 1 M (SD)	Session 2 M (SD)	Session 1 M (SD)	Session 2 M (SD)	Session 1 M (SD)	Session 2 M (SD)	Session 1 M (SD)	Session 2 M (SD)
Age	20.00 (2.02)	-	19.85 (2.32)	-	20.32 (2.26)	-	19.97 (1.94)	-
BMI	21.14 (2.66)	-	23.30 (3.34)	-	22.16 (2.85)	-	23.43 (2.85)	-
EDE-Q	0.46 (0.34)	-	2.12 (1.15)	-	0.39 (0.27)	-	2.31 (1.05)	-
VAS Body Satisfaction	74.10 (16.95)	76.81 (12.32)	44.50 (25.10)	44.39 (25.31)	74.29 (15.17)	72.87 (15.87)	38.52 (22.58)	38.29 (24.07)
Appearance SSES	23.50 (3.16)	24.00 (2.74)	17.30 (4.91)	17.58 (5.27)	23.94 (2.86)	23.71 (3.13)	16.84 (4.40)	16.48 (4.81)
Social SSES	27.38 (4.01)	27.82 (4.33)	22.70 (4.75)	24.15 (4.61)	27.42 (3.96)	27.32 (5.07)	21.74 (5.56)	22.81 (6.10)
Performance SSES	26.44 (4.43)	27.68 (4.58)	26.00 (4.33)	25.85 (4.76)	28.16 (4.14)	28.55 (4.10)	23.65 (6.32)	23.94 (5.65)
RSES	18.15 (4.61)	18.53 (4.08)	21.78 (4.70)	21.94 (5.02)	17.55 (4.16)	17.84 (4.52)	22.68 (5.20)	23.29 (5.35)
VAS Mood	65.14 (16.83)	70.40 (12.91)	57.41 (18.84)	58.14 (19.77)	70.03 (14.08)	68.53 (16.87)	54.37 (20.26)	55.52 (21.44)
BISS	-	38.41 (4.99)	-	28.61 (8.85)	-	36.84 (6.30)	-	26.29 (8.39)

Note. EDE-Q = Eating Disorder Examination Questionnaire (range 0–6, higher scores indicate more eating disordered thoughts and behaviors); VAS Body Satisfaction (range 0–100, higher scores indicate higher body satisfaction); Appearance SSES = State Self-Esteem Scale (range 6–30, higher scores indicate more appearance state self-esteem); Social SSES = State Self-Esteem Scale (range 7–35, higher scores indicate more social state self-esteem); Performance SSES = State Self-Esteem Scale (range 7–35, higher scores indicate more performance state self-esteem); RSES = Rosenberg Self-Esteem Scale (range 10–40, higher scores indicate more self-esteem); VAS Mood (range 0–100, higher scores indicate a more positive mood); BISS = Body Image States Scale (range 6–54, higher scores indicate higher body satisfaction); Session 1 = before the evaluative conditioning intervention; Session 2 = after the evaluative conditioning intervention.

consistent over conditions (BMI: Condition $\times$ BC, $F(1, 124) = 0.73$, $p = .39$; total EDE-Q: low BC group, $\chi^2(1) = 0.28$, $p = .60$; high BC group, $\chi^2(1) = 0.95$, $p = .33$).

3.1.1. Manipulation check

In general, participants completed the EC task in a conscientious manner (M RT = 523 ms, $SD = 213$ ms, $range = 322$ – 1650 ms; mean % of trials > 3 s = $< 1\%$). In addition, we checked the assumed valence of the USs (i.e. face pictures) in a RM-ANOVA with the negative-positive ratings as dependent variable and US type (happy, neutral, frowning) as within-subjects factor. Since the sphericity assumption was violated, we performed a Greenhouse-Geisser correction on the degrees of freedom (Greenhouse & Geisser, 1959). As expected, there was a main effect of US type ($F(1.31, 163) = 1084$, $p < .001$, $\eta_p^2 = .90$). That is, participants rated the happy faces as most positive ($M = 85.4$, $SD = 11.8$), the frowning faces as most negative ($M = 18.8$, $SD = 12.0$) and the neutral faces in between ($M = 44.6$, $SD = 9.59$). There were no systematic differences between conditions in the valence ratings of the USs (Main effect Condition: $F(1, 125) = 0.004$, $p = .95$, $\eta_p^2 = .000$; Condition $\times$ US type: $F(1.31, 163) = 0.49$, $p = .54$, $\eta_p^2 = .004$; Condition $\times$ BC $\times$ US type: $F(1.31, 163) = 1.51$, $p = .23$, $\eta_p^2 = .012$). However, we did find a main effect of BC on US ratings ($F(1, 125) = 6.73$, $p = .011$, $\eta_p^2 = .051$) indicating that individuals low in BC on average rated the faces more positive than individuals high in BC. This effect did not differ between US types (BC $\times$ US type: $F(1.31, 163) = 0.31$, $p = .64$, $\eta_p^2 = .002$). For an overview of the means and standard deviations of the ratings per group, see Table 2.

Table 2

Means and standard deviations of US ratings post-intervention per group.

	Experimental condition		Control condition	
	Low BC (N = 34)	High BC (N = 33)	Low BC (N = 31)	High BC (N = 31)
Happy faces	86.09 (12.75)	86.29 (10.29)	85.99 (10.53)	82.99 (13.66)
Neutral faces	47.38 (9.13)	40.87 (11.61)	45.06 (5.96)	44.90 (9.87)
Frowning faces	21.20 (15.14)	15.72 (9.88)	19.77 (11.10)	18.47 (10.76)

Note. Happy faces (range 0–100, higher scores indicate a higher mean valence rating for happy faces); Neutral faces (range 0–100, higher scores indicate a higher mean valence rating for neutral faces); Frowning faces (range 0–100, higher scores indicate a higher mean valence rating for frowning faces).

3.2. Primary outcome measures of body satisfaction

First, we checked whether the experimental and control condition differed in baseline levels of VAS body satisfaction and appearance SSES, which was not the case (VAS body satisfaction: $t(127) = -0.68$, $p = .50$, two-tailed; appearance SSES: $t(127) = -0.067$, $p = .95$, two-tailed). Subsequently, we performed the planned RM-ANOVAs on VAS body satisfaction and appearance SSES. In contrast to our expectations the two-way interactions of Condition $\times$ Time and three-way interactions of Condition $\times$ BC $\times$ Time were not significant for both body satisfaction indices (VAS body satisfaction: Condition $\times$ Time, $F(1, 125) = 0.75$, $p = .39$, $\eta_p^2 = .006$; Condition $\times$ BC $\times$ Time, $F(1, 125) = 0.67$, $p = .42$, $\eta_p^2 = .005$; appearance SSES: Condition $\times$ Time, $F(1, 125) = 1.81$, $p = .18$, $\eta_p^2 = .014$; Condition $\times$ BC $\times$ Time, $F(1, 125) = 0.010$, $p = .92$, $\eta_p^2 = .000$). As expected, we did find significant main effects for Body concern (VAS body satisfaction: $F(1, 125) = 97.8$, $p < .001$, $\eta_p^2 = .44$; appearance SSES, $F(1, 125) = 103$, $p < .001$, $\eta_p^2 = .45$) indicating that participants high in BC reported lower body satisfaction and less appearance related self-esteem compared to participants low in BC. No significant main effects were evident for Condition (VAS body satisfaction: $F(1, 125) = 1.40$, $p = .24$, $\eta_p^2 = .011$; appearance SSES: $F(1, 125) = 0.28$, $p = .60$, $\eta_p^2 = .002$), for Condition $\times$ BC (VAS body satisfaction: $F(1, 125) = 0.39$, $p = .54$, $\eta_p^2 = .003$; appearance SSES, $F(1, 125) = .41$, $p = .52$, $\eta_p^2 = .003$), or for BC $\times$ Time (VAS body satisfaction: $F(1, 125) = 0.11$, $p = .74$, $\eta_p^2 = .001$; appearance SSES, $F(1, 125) = 0.13$, $p = .72$, $\eta_p^2 = .001$). These outcomes indicate that individuals in the experimental condition did not increase in their levels of body satisfaction after the training compared to the individuals in the control condition. In addition, these findings do not support that women with relatively high body concern profited more from the intervention than women with relatively low body concern.

For exploratory purposes, we tested whether there was a difference in total BISS scores between the experimental and control condition after the intervention. A 2 Condition (experimental, control) $\times$ 2 Body concern (low, high) ANOVA was performed on post-intervention BISS scores. Again, the main effect of Condition on total BISS scores ($F(1, 125) = 2.30$, $p = .13$, $\eta_p^2 = .018$) and the interaction of Condition $\times$ BC were not significant ($F(1, 125) = 0.084$, $p = .77$, $\eta_p^2 = .001$), but the main effect of BC was ($F(1, 125) = 62.9$, $p < .001$, $\eta_p^2 = .34$).

3.3. Secondary outcome measures

At baseline, individuals in the experimental and control condition did not differ on any of the secondary outcome measures (global self-esteem: $t(127) = 0.19, p = .85$; social SSES: $t(127) = -0.53, p = .60$; performance SSES: $U = 2029, z = -.23, p = .82$; VAS mood, $U = 1970, z = -.51, p = .61$). Next, we performed 4 separate RM-ANOVAs on mood and the self-esteem indices (i.e. global self-esteem, social SSES, and performance SSES). Again, the two-way interactions of Condition x Time and three-way interactions of Condition x BC x Time were not significant (global self-esteem: Time x Condition: $F(1, 125) = 0.13, p = .72, \eta_p^2 = .001$; Time x Condition x BC: $F(1, 125) = 0.30, p = .59, \eta_p^2 = .002$; social SSES: Time x Condition: $F(1, 125) = 0.59, p = .44, \eta_p^2 = .005$; Time x Condition x BC: $F(1, 125) = 0.015, p = .90, \eta_p^2 = .000$; performance SSES: Time x Condition: $F(1, 125) = 0.12, p = .73, \eta_p^2 = .001$; Time x Condition x BC: $F(1, 125) = 1.22, p = .27, \eta_p^2 = .010$; VAS mood: Time x Condition, $F(1, 125) = 0.83, p = .36, \eta_p^2 = .007$; Time x Condition x BC: $F(1, 125) = 1.06, p = .30, \eta_p^2 = .008$). For mood and performance SSES the normality assumption was violated. Therefore, we repeated the analyses with the non-parametric Friedman Test. The findings were consistent with the results from the RM-ANOVAs. Thus, individuals in the experimental condition did not report increased levels of global self-esteem, social state self-esteem, performance state self-esteem or mood after the training compared to the individuals in the control condition. In addition, these findings do not support that women with relatively high body concern profited more from the intervention than women with relatively low body concern.

3.4. Stimulus valence

The within-group covariance matrices for the planned RM-ANOVA were unequal. Therefore, Pillai's trace criterion was used for this analysis (Pillai, 1955). Consistent with our expectation, we found that the two-way interaction of Condition x CS type was significant ($F(1, 125) = 6.22, p = .014, \eta_p^2 = .047$). Participants in the experimental condition rated their own pictures more positively and the control pictures more negatively than participants in the control condition (see Fig. 2). The three-way interaction of Condition x BC x CS type was not significant ($F(1, 125) = 0.38, p = .54, \eta_p^2 = .003$), indicating that

although participants rated their own pictures more positively after the intervention, this effect was not stronger for women with relatively high body concern. The effect of the intervention on the ratings was independent of VAS dimension (positive, attractive) (Condition x CS type x Dimension, $F(1, 125) = 0.14, p = .71, \eta_p^2 = .001$; Condition x BC x CS type x Dimension, $F(1, 125) = 0.88, p = .35, \eta_p^2 = .007$). For an overview of the means and standard deviations of the ratings per group, see Table 3.

3.5. Post-hoc analyses

3.5.1. Awareness check

Awareness data from 2 participants were missing. It was not possible to conduct a similar MANOVA analysis on contingency awareness as Martijn et al. (2010) performed, since assumption violation could only be solved by excluding the contingency aware participants (outliers) from the analysis. In Fig. 3, mean estimates of the percentage smiling, neutral, and frowning faces are presented as a function of CS type (self, other) and Condition (experimental, control). On average, participants in the experimental condition estimated that their own pictures were more often followed by smiling faces ($M = 54.5, SD = 31.9$), compared to the control pictures ($M = 20.3, SD = 16.8$). In line with this, they estimated that neutral and frowning faces more often followed the control pictures than participant pictures. For the control condition, all estimates (regardless of US and CS type) were in between 30 and 40%, with higher estimates being found for the percentage frowning faces that had followed the participants' pictures (39.0%) and control pictures (35.6%). These results indicate that the participants in the experimental condition were not entirely unaware of the statistical contingency between the USs and CSs and were overall more contingency aware than the sample of Martijn et al. (2010).

Since the EC procedure might not have worked in contingency or hypothesis aware individuals, we repeated the analyses for the primary outcome measures excluding data of participants that had correctly detected that their own pictures were in 100% of the trials followed by a smiling picture ($N = 11$) and/or showed awareness of the hypothesis ($N = 11$; cf. Martijn et al., 2010). However, this did not change the outcomes, i.e. the two-way interactions of Condition x Time and three-way interactions of Condition x BC x Time were not significant for both body satisfaction indices (VAS body satisfaction: Condition x Time, $F(1,$

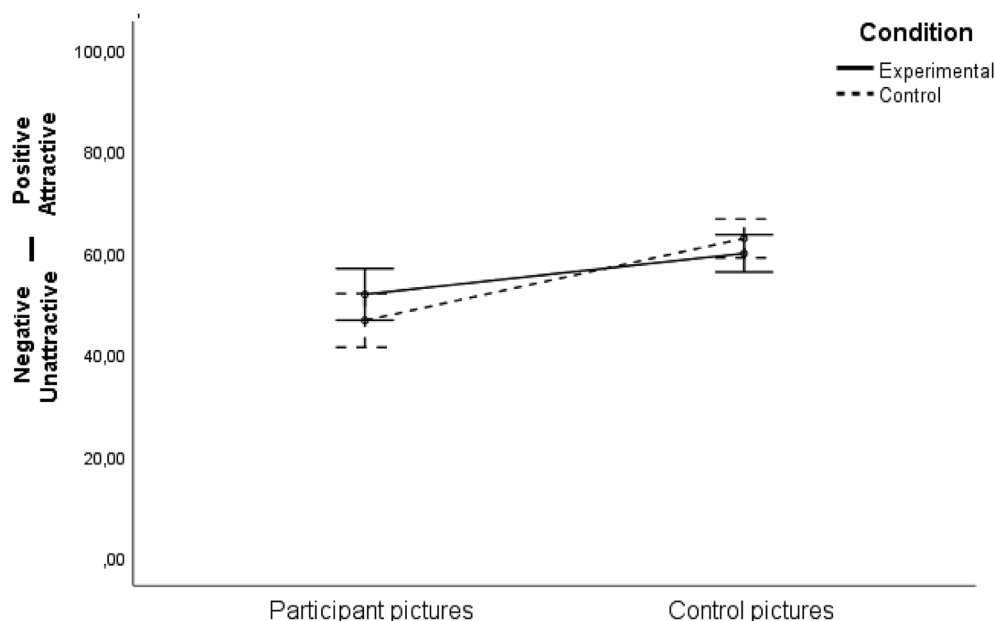


Fig. 2. Mean valence + attractiveness rating post-intervention per CS type and as a function of condition (" 95% CI). Higher scores indicate a more positive rating of the CS after the intervention (VAS scale ranges from 0 to 100).

Table 3
Means and standard deviations of CS ratings post-intervention per group.

		Experimental condition		Control condition	
		Low BC (N = 34)	High BC (N = 33)	Low BC (N = 31)	High BC (N = 31)
Participant pictures	Valence	63.64 (15.40)	40.47 (19.16)	61.14 (16.40)	33.44 (19.32)
	Attractiveness	63.44 (16.11)	39.02 (16.87)	60.20 (14.57)	31.90 (20.77)
Control pictures	Valence	62.92 (14.95)	60.52 (18.83)	66.98 (14.80)	62.06 (14.80)
	Attractiveness	60.31 (16.58)	55.68 (18.38)	62.48 (15.67)	59.78 (14.57)

Note. Participant pictures (range 0–100, higher scores indicate a higher mean valence or attractiveness rating for the participants' own pictures); Control pictures (range 0–100, higher scores indicate a higher mean valence or attractiveness rating for the control pictures).

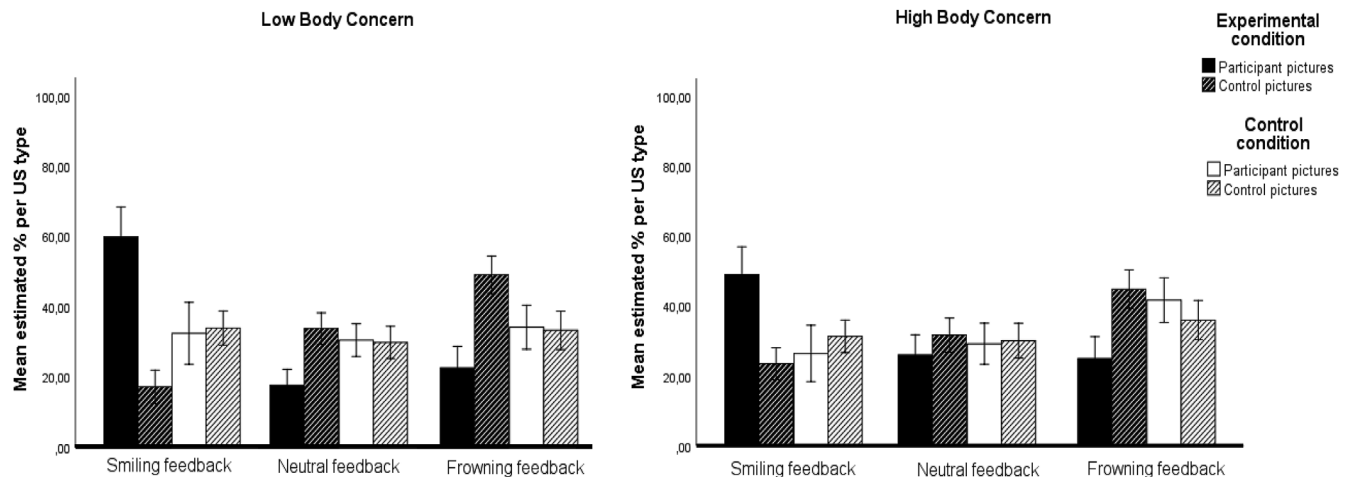


Fig. 3. Contingency awareness per condition and body concern group. Scores represent the mean estimated percentage of each US type as a function of CS type ("95% CI). Higher scores indicate a higher estimate of the percentage US.

101) = 0.009, $p = .93$, $\eta_p^2 = .000$; Condition $\times$ BC $\times$ Time, $F(1, 101) = 0.001$, $p = .97$, $\eta_p^2 = .000$; appearance SSES: Condition $\times$ Time, $F(1, 101) = 1.19$, $p = .28$, $\eta_p^2 = .012$; Condition $\times$ BC $\times$ Time, $F(1, 101) = 0.39$, $p = .53$, $\eta_p^2 = .004$). So, even in a 'contingency unaware' sample, similar to that of Martijn et al. (2010), we did not find an effect of the intervention on the body satisfaction indices.

3.5.2. Body concern scale split at 1.70

The median body concern in the present sample (1.19) was considerably lower than the median of 1.70 that Martijn et al. (2010) used to assign participants into low and high BC groups. Consequently, our high BC group was not comparable to theirs, which is especially problematic considering they only found an effect of the intervention in the high BC group. Therefore, we re-divided the participants over the BC groups; all participants with a body concern score equal to or below 1.70 were assigned to the low BC group and the remainder of the participants were assigned to the high BC group. Next, we repeated the RM-ANOVAs for the primary outcome variables. Again, results were consistent with our prior findings (VAS body satisfaction: Condition $\times$ Time: $F(1, 125) = 0.60$, $p = .44$, $\eta_p^2 = .005$; Condition $\times$ BC $\times$ Time: $F(1, 125) = 0.12$, $p = .73$, $\eta_p^2 = .001$; appearance SSES: Condition $\times$ Time: $F(1, 125) = 2.13$, $p = .15$, $\eta_p^2 = .017$; Condition $\times$ BC $\times$ Time: $F(1, 125) = 0.41$, $p = .53$, $\eta_p^2 = .003$). Thus, also if we used the same criterion to divide high and low BC as was used in Martijn et al. (2010), we were unable to replicate the previous findings that the EC procedure was specifically effective for participants high in body concern.

3.5.3. Regression model with body concern as a continuous variable

The use of median split as a technique for creating groups has been widely criticized, since it results in a considerable loss of measurement information and increases the chance of type II errors (e.g., Cohen, 1983). Therefore, we repeated the primary analyses with body concern

as a continuous variable using two separate stepwise hierarchical regression analyses. Since the interaction of Condition $\times$ BC was included as a predictor, we centered the continuous predictors (i.e. baseline VAS body satisfaction, baseline appearance SSES and BC) to avoid problems with multi-collinearity. Preliminary analyses led to the exclusion of data from 2 participants, who were identified as outliers. In Step 1 either baseline VAS body satisfaction or baseline appearance SSES was entered into the model. As expected, pre-intervention body satisfaction was a significant predictor of post-intervention body satisfaction (VAS body satisfaction: $R_{\text{change}}^2 = .73$, $F_{\text{change}}(1, 125) = 335$, $p < .001$; appearance SSES: $R_{\text{change}}^2 = .72$, $F_{\text{change}}(1, 125) = 313$, $p < .001$). In Step 2, Condition was entered in the model as a predictor. However, this did not significantly increase the variance explained by the model (VAS body satisfaction: $R_{\text{change}}^2 = .003$, $F_{\text{change}}(1, 124) = 1.46$, $p = .23$; appearance SSES: $R_{\text{change}}^2 = .006$, $F_{\text{change}}(1, 124) = 2.44$, $p = .12$). In Step 3, BC was entered into the model as a continuous variable, which significantly improved the model (VAS body satisfaction: $R_{\text{change}}^2 = .022$, $F_{\text{change}}(1, 123) = 11.0$, $p < .001$; appearance SSES: $R_{\text{change}}^2 = .012$, $F_{\text{change}}(1, 123) = 5.71$, $p = .018$). In step 4, the interaction of Condition $\times$ BC was entered. In contrast with our expectations, this did not significantly improve the model (VAS body satisfaction, $R_{\text{change}}^2 = .001$, $F_{\text{change}}(1, 122) = 0.60$, $p = .44$; appearance SSES, $R_{\text{change}}^2 = .000$, $F_{\text{change}}(1, 122) = 0.029$, $p = .86$). In the final model, only baseline levels of body satisfaction and BC were significantly associated with body satisfaction after the intervention.

3.5.4. Covariation bias

Since it has been previously found that individuals with a negative body image tend to overestimate negative social feedback (so-called covariation bias; Alleva, Lange, Jansen, & Martijn, 2014), we explored whether higher levels of BC in participants in the placebo-condition were related to overestimation of the percentage of frowning faces that

followed their own pictures. First, we calculated a covariation bias (CB) index by subtracting the estimate for the control pictures from the estimate for the participant pictures for each US type. Subsequently, we used the non-parametric Wilcoxon Signed Rank Test to test whether CB-index significantly differed from zero. Results indicated a significant CB for frowning faces ($Z = 292$, $p = .003$) and smiling faces ($Z = 41.5$, $p = .001$), but not for neutral faces ($Z = 81.0$, $p = .84$). Next, we examined whether the CB-index correlated significantly with body concern. Because body concern was not normally distributed in our sample (distribution was strongly left-skewed), this was tested non-parametrically using Spearman's rho correlation coefficient. Consistent with the findings of [Allewaert et al. \(2014\)](#), we found a significant positive correlation of body concern with the CB-index for frowning faces ($r(60) = .32$, $p = .012$) and a significant negative correlation of body concern with the CB-index for smiling faces ($r(60) = -.30$, $p = .021$). Body concern did not significantly correlate with CB-index for neutral faces ($r(60) = -.16$, $p = .24$). These results indicate that body concerned individuals indeed tend to overestimate negative social feedback, while underestimating the frequency of positive social feedback.

4. Discussion

The current study investigated the robustness of an EC procedure developed by [Martijn et al. \(2010\)](#) as a tool to enhance body satisfaction in a sample of female undergraduate students. Supporting the validity of the current EC procedure, we found that participants in the experimental condition rated their own pictures after the training as more positive/attractive and the control pictures as more negative/unattractive than participants in the control condition. However, the latter effect was not stronger for participants with high body concern. In addition, we did not find an effect of the EC procedure on self-report indices of body satisfaction, i.e. individuals in the experimental condition did not increase in their levels of body satisfaction after the training compared to the individuals in the control condition. Moreover, women with relatively high body concern did not profit more from the EC procedure than women with relatively low body concern. Finally, we obtained similar results for the secondary outcome measures global self-esteem, state social self-esteem, state performance self-esteem, and mood.

The present findings are not consistent with preclinical studies showing a positive effect of this EC procedure on body image and self-esteem in women with high body concern ([Aspen et al., 2015](#); [Martijn et al., 2010](#)). However, the outcomes are in line with a recent clinical study in which the EC procedure did not result in a change of body image and self-esteem measures in adolescents girls with an eating disorder ([Glashouwer et al., 2018](#)). To get more insight into potential explanations for these null-findings, we conducted several post-hoc tests. First of all, we noticed that the median level of body concern was considerably lower in our sample than in the sample of [Martijn et al. \(2010\)](#). Therefore, we repeated the primary analyses using the same criterion to divide high and low BC groups as [Martijn et al. \(2010\)](#). In addition, we also repeated the primary analyses with body concern as a continuous variable to overcome the limitations of median split analyses. However, for both post-hoc analyses results remained unchanged. Second, it seems as if our sample showed somewhat higher contingency and hypothesis awareness than the study of [Martijn et al. \(2010\)](#). Although, it was shown that contingency awareness actually is related to stronger EC effects ([Hofmann et al., 2010](#)), we repeated the primary analyses without the aware participants to be consistent with [Martijn et al.](#) Again, results remained the same.

In addition to the original design, we checked whether the active EC procedure resulted in a change in valence of the body pictures (CSs). This was indeed the case. After the training, participants in the experimental condition rated their own pictures as more positive and the control pictures as more negative than participants in the control condition. Although the effect size was small, it can be considered as

already quite impressive that such a brief training resulted in a change in CS valence, all the more considering that it is generally harder to change existing attitudes than to form new attitudes towards neutral objects ([De Houwer et al., 2001](#)). In contrast to our expectations, this effect was not especially pronounced in individuals with high body concern. Although body dissatisfied individuals might have more room for improvement of their CS ratings, one could also reason that body dissatisfied individuals are actually more resistant to change since the valence of their own body is negative and therefore potentially more difficult to change. From the latter perspective, it can be considered reassuring that individuals high in body dissatisfaction on average did not improve less in terms of CS ratings than individuals low in body dissatisfaction. Together, these findings indicate that the EC procedure worked as intended, but that effects might not have been strong enough to 'spill over' to self-report measures of body satisfaction. In prior studies showing positive effects of this EC procedure on body image, the subjective valence of the CS-pictures was not measured ([Aspen et al., 2015](#); [Martijn et al., 2010](#)). Therefore, it remains unclear whether the effects of the EC procedure in these studies were stronger and therefore also transferred to self-report measures of body satisfaction, or that prior positive findings may just represent a fragile effect.

Although USs (face pictures) in the present study were identical to USs used in [Martijn et al. \(2010\)](#), we decided to create new CS pictures of control bodies to make the pictures of the own and the control bodies visually comparable. Since two-third of the trials comprised control bodies, this difference might have impacted the results. A first possibility is that the CS pictures of control bodies in the sample of [Martijn et al.](#) might have been perceived as more attractive than the CS pictures of control bodies in our sample. The latter can have led to a bigger contrast in the EC procedure of [Martijn et al. \(2010\)](#) compared to the one used in the present study. Because the effect of the EC procedure might be for an important part driven by the contrast between (attractive) control bodies being followed by negative feedback and own body pictures being followed by positive feedback, a larger contrast between the attractiveness of the participants' own bodies versus the control bodies might result in stronger EC effects. A second possibility is that the CS pictures of control bodies in our sample were evaluated as more attractive leading to more negative comparison tendencies in our population, in particular in the high body concern group. The latter might have increased noise and thereby decreased the chance to find effects on body satisfaction in our study. Future studies should more carefully investigate the impact of the control bodies used in this training task. It could even be an option to adapt the EC procedure excluding the control bodies from the task, for example by replacing these pictures with a more neutral control category. In addition, it seems relevant to consider that participants in the control condition generally overestimated the percentage of negative feedback that followed their own body CSs, and underestimated the percentage of positive feedback that followed their own body CSs. Such covariation bias in the control condition may result in a negative evaluative conditioning effect, and thus partly drive the difference in appreciation of the CSs between the control and experimental conditions. However, the finding that the covariation bias was stronger in individuals with relatively strong body concerns, whereas the experimental effect on the CSs was not relatively pronounced in participants with high body concerns, seems to indicate that covariation bias was not a major factor in the current EC effects. Another methodological point of concern is the valence of the USs. The rating of positive USs (smiling faces) on average was higher than the rating of own body CSs, which indicates that there indeed was room for positive change in valence of the own body CSs. However, US ratings might differ on an individual level. This implies that the EC procedure perhaps can be improved by rating USs and CSs before the intervention. That way, for each individual the most positive USs can be selected as feedback for the own body pictures in order to maximize the difference in valence and therefore the effect of the EC procedure. Finally, another way to boost the effect of the EC

procedure could be to increase contingency awareness of the participants. A meta-analysis showed that in general, higher levels of contingency awareness are related to stronger EC effects (Hofmann et al., 2010). Since a large part of the present sample in the experimental condition were not aware of the contingency between CSs and USs, increasing this awareness could increase the effect of the EC procedure.

5. Conclusions

Our study did not provide evidence for the effectiveness of an EC procedure developed by Martijn et al. (2010) as a tool to enhance body satisfaction in undergraduate women. Despite positive findings in small-scale preclinical samples, we did not find any positive effects of the EC procedure on body satisfaction or self-esteem in the present well-powered sample. As an important addition to prior studies, we assessed the impact of the EC procedure on the affective appreciation of the CS-pictures used in the training. In support of our aim to strengthen the association between participants' appearance and social approval, we found that the active EC procedure resulted in a change in valence of the body pictures (CSs) in the expected direction. Together, these findings suggest that the EC procedure in its present form is not ready for use as an intervention for improving body satisfaction. However, it seems worthwhile to investigate in future studies how the EC procedure can be further strengthened in a way that effects on CS ratings eventually also 'spill over' to the level of self-reported body satisfaction.

Conflicts of interest

All authors declare that they have no conflicts of interest.

Data

The raw data supporting the conclusions of this manuscript are available through <https://hdl.handle.net/10411/QMPIVW>.

Acknowledgements

This study was supported by a Veni grant [451-15-026] awarded by the Netherlands Organisation for Scientific Research. The authors would like to thank Carolien Martijn for sharing materials and providing all details necessary for setting up the study; Anna Nitzsche who helped with the data collection; and the students who volunteered to participate in this study.

References

- Allewa, J. M., Lange, W. G., Jansen, A., & Martijn, C. (2014). Seeing ghosts: Negative body evaluation predicts overestimation of negative social feedback. *Body Image*, 11, 228–232.
- Allewa, J. M., Sheeran, P., Webb, T. L., Martijn, C., & Miles, E. (2015). A meta-analytic review of interventions designed to improve body image. *PLoS One*, 10, e0139177. <https://doi.org/10.1371/journal.pone.0139177>.
- Aspen, V., Martijn, C., Allewa, J., Nagel, J., Perret, C., Purvis, C., et al. (2015). Decreasing body dissatisfaction using a brief conditioning intervention. *Behaviour Research and Therapy*, 69, 93–99. <https://doi.org/10.1016/j.brat.2015.04.003>.
- Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science*, 15, 498–502. <https://doi.org/10.1111/j.0956-7976.2004.00708.x>.
- Butters, J. W., & Cash, T. F. (1987). Cognitive-behavioral treatment of women's body-image dissatisfaction. *Journal of Consulting and Clinical Psychology*, 55, 889–897. <https://doi.org/10.1037/0022-006X.55.6.889>.
- Cash, T. F. (2011). Cognitive-behavioral perspectives on body image. In T. F. Cash, & L. Smolak (Eds.), *Body image: A handbook of science, practice, and prevention* (pp. 39–47). New York, NY: Guilford Press.
- Cash, T. F., Fleming, E. C., Alindogan, J., Steadman, L., & Whitehead, A. (2002). Beyond body image as a trait: The development and validation of the Body Image States Scale. *Eating Disorders. J. Prev. Treat.*, 10, 103–113. <https://doi.org/10.1080/10640260290081678>.
- Cash, T. F., & Pruzinsky, T. E. (1990). *Body images: Development, deviance, and change*. New York: Guilford Press.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7(3), 249–253.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853–869. <https://doi.org/10.1037/0033-2909.127.6.853>.
- Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology*, 86, 345–355. <https://doi.org/10.1037/0022-3514.86.2.345>.
- Fairburn, C. G., & Beglin, S. J. (2008). Eating disorder examination questionnaire (EDE-Q 6.0). In C. G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders* (pp. 309–314). New York, NY: Guilford Press.
- Fallon, E. A., Harris, B. S., & Johnson, P. (2014). Prevalence of body dissatisfaction among a United States adult sample. *Eating Behaviors*, 15, 151–158. <https://doi.org/10.1016/j.eatbeh.2013.11.007>.
- Glashouwer, K. A., Neimeijer, R. A. M., de Koning, M., Vestjens, M., & Martijn, C. (2018). Evaluative conditioning as a body image intervention for adolescents with eating disorders. *Journal of Consulting and Clinical Psychology*, 86(12), 1046–1055. <https://doi.org/10.1037/ccp0000311>.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95–112.
- Griffen, T. C., Naumann, E., & Hildebrandt, T. (2018). Mirror exposure therapy for body image disturbances and eating disorders: A review. *Clinical Psychology Review*, 65, 163–174. <https://doi.org/10.1016/j.cpr.2018.08.006>.
- Grogan, S. (2006). Body image and health: Contemporary perspectives. *Journal of Health Psychology*, 11(4), 523–530. <https://dx.doi.org/10.1177/1359105306065013>.
- Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, 60(6), 895–910. <https://doi.org/10.1037/0022-3514.60.6.895>.
- Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, 136(3), 390–421. <https://doi.org/10.1037/a0018916>.
- Johnson, F., & Wardle, J. (2005). Dietary restraint, body dissatisfaction, and psychological distress: A prospective analysis. *Journal of Abnormal Psychology*, 114(1), 119–125. <https://doi.org/10.1037/0021-843X.114.1.119>.
- Luce, K. H., & Crowther, J. H. (1999). The reliability of the eating disorder examination—self-report questionnaire version (EDE-Q). *International Journal of Eating Disorders*, 25(3), 349–351.
- Martijn, C., Vanderlinden, M., Roefs, A., Huijding, J., & Jansen, A. (2010). Increasing body satisfaction of body concerned women through evaluative conditioning using social stimuli. *Health Psychology*, 29, 514–520. <https://doi.org/10.1037/a0020770>.
- McLean, S. A., Paxton, S. J., & Wertheim, E. H. (2011). A body image and disordered eating intervention for women in midlife: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 79, 751–758. <https://doi.org/10.1037/a0026094>.
- Paxton, S. J., Neumark-Sztainer, D., Hannan, P., & Eisenberg, M. E. (2006). Body dissatisfaction prospectively predicts depressive mood and low self-esteem in adolescent girls and boys. *Journal of Clinical Child and Adolescent Psychology*, 35, 539–549.
- Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *The Annals of Mathematical Statistics*, 26(1), 117–121.
- Riketta, M., & Dauenheimer, D. (2003). Manipulating self-esteem with subliminally presented words. *European Journal of Social Psychology*, 33(5), 679–699.
- Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy. Meas. Package*, 61, 52.
- Tiggemann, M. (2004). Body image across the adult life span: Stability and change. *Body Image*, 1, 29e41. [https://doi.org/10.1016/S1740-1445\(03\)00002e0](https://doi.org/10.1016/S1740-1445(03)00002e0).
- Tottenham, N., Tanaka, J., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168, 242–249. <https://doi.org/10.1016/j.psychres.2008.05.006>.
- Wilfley, D. E., Schwartz, M. B., Spurrell, E. B., & Fairburn, C. G. (1997). Assessing the specific psychopathology of binge eating disorder patients: Interview or self-report? *Behaviour Research and Therapy*, 35(12), 1151–1159.
- Wilson, R. E., Latner, J. D., & Hayashi, K. (2013). More than just body weight: The role of body image in psychological and physical functioning. *Body Image*, 10, 644–647. <https://doi.org/10.1016/j.bodyim.2013.04.007>.