Northumbria Research Link

Citation: Irvine, Kamila, Invine, Andrew R, Maalin, Nadia, Mccarty, Kris, Cornelissen, Katri, Tovee, Martin J. and Cornelissen, Piers (2020) Using immersive virtual reality to modify body image. *Body Image*, 33. pp. 232-243. ISSN 1740-1445

Published by: Elsevier

URL: <https://doi.org/10.1016/j.bodyim.2020.03.007>
<<https://doi.org/10.1016/j.bodyim.2020.03.007>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/42726/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria
University
NEWCASTLE

UniversityLibrary

Using immersive virtual reality to modify body image

Kamila R. Irvine ^{a 1}, Andrew R. Irvine ^{b, c}, Nadia Maalin ^c, Kristofor McCarty ^a, Katri K. Cornelissen ^a,
Martin J. Tovée ^a, & Piers L. Cornelissen ^a

^a *Department of Psychology, Faculty of health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK*

^b *Virtual Research Innovations Ltd.*

^c *School of Psychology, College of Social Science, University of Lincoln, Lincolnshire, LN6 7TS, UK*

¹ *Present address: School of Psychology, College of Social Science, University of Lincoln, Lincolnshire, LN6 7TS, UK*

Number of references: 113

Number of tables: 4

Number of figures: 4

Declarations of interests: none

Funding: This work was supported by a PhD studentship from Northumbria University awarded to the first author.

Corresponding author: email address: kirvine@lincoln.ac.uk (K. R. Irvine)

1

2
3
4
5
6
7
8
9
10
11
12
13
14

16

1. Introduction

Body image is a multidimensional construct comprising two related, yet independent modalities: (a) *an attitudinal component* (also known as an affective-cognitive component) which corresponds to the feelings someone has about their body, and (b) *a perceptual component* which captures the size and shape they believe themselves to be (Cash & Deagle, 1997; Schilder, 1935; Slade, 1988) and which probably corresponds to a volumetric representation (Cornelissen et al., 2019). Currently, body image disturbance is a cardinal feature of anorexia nervosa and is included as one of the diagnostic criteria for the disorder in the DSM5 (APA, 2013; Cash & Deagle, 1997; Fairburn, Cooper, & Shafran, 2003). Perceptually, body image disturbance in anorexia typically manifests as over-estimation of one's own body size (Gardner & Brown, 2014) and women with anorexia nervosa often engage in comparing their own body with others (Alleva et al., 2013; Espeset et al., 2012). Attitudinally, body image disturbance manifests as body dissatisfaction (Grogan, 2016; Striegel-Moore & Bulik, 2007). Body image disturbance is one of the most persistent of all the eating disorder symptoms, it predicts onset of unsafe weight loss behaviour, and its severity predicts treatment outcomes and relapse rates (Castro, Gila, Puig, Rodriguez, & Toro, 2004; Freeman, Thomas, Solyom, & Koopman, 1985; Junne et al., 2019; Liechty, 2010).

1.1 Treatments focused on both weight gain and psychological symptoms

Effective treatment of anorexia nervosa has proved elusive and relapse is a significant problem, with rates reported to be as high as 22%-51% (Berkman, Lohr, & Bulik, 2007; Castro et al., 2004; Channon & DeSilva, 1985; Herzog et al., 1999; Keel, Dorer, Franko, Jackson, & Herzog, 2005; Strober, Freeman, & Morrell, 1998). Recently, van den Berg et al. (2019) reported a meta-analysis of randomized controlled trials (RCT) that tested the efficacy of psychological treatments for anorexia nervosa. In this report, psychological treatments included psychodynamic or psychoanalytic therapy, cognitive (behavioural) therapy, interpersonal therapy, family therapy, social skills training, motivational interviewing, as well as combinations of these. Control conditions included dietary advice, psychoeducational interventions and placebo conditions under the generic heading of treatment as usual. The meta-analysis included data from 1,279 participants (761 in treatment groups) who took part in 17 RCTs. van den Berg et al. (2019) found no significant differences between psychological treatments and controls for weight gain, eating disorder pathology, and quality of life. Another recent meta-analysis of 35 RCTs which also focused on weight gain and psychological symptoms as outcomes, was conducted by Murray, Quintana, Loeb, Griffiths, and Le Grange (2019). This meta-analysis included a wider range of treatment options including psychosocial, pharmacological, medical, and complementary/alternative interventions. Compared to controls,

Murray et al. (2019) did find evidence for a significant treatment effect on weight gain at the end of treatment, but this did not persist at follow up. With respect to psychological outcomes, these authors failed to find any significant differences between treatment and control conditions at either time point.

1.2 Treatments focused on distorted body image

Disturbances in body image often persist post-treatment in anorexia nervosa (Bachner-Melman, Zohar, & Ebstein, 2006) and their presence increases the likelihood of relapse (Carter, Blackmore, Sutandar-Pinnock, & Woodside, 2004). Consequently, there has also been an imperative to develop specific treatments for disturbed body image as part of the management of anorexia nervosa (Garner & Garfinkel, 1997). Ziser et al. (2018) reviewed two case reports/series and nine studies that focused solely on body image interventions and which involved a total of 260 patients and 345 controls. A wide variety of techniques had been used in these studies, including mirror exposure, video confrontation, and virtual reality body exposure as well as mixed forms of intervention. The latter incorporated techniques such as psychoeducation, relaxation techniques and cognitive restructuring. Ziser et al. (2018) conclude, albeit cautiously, that such approaches are promising. However, they also caution that it is unclear whether adding such interventions to established treatments for patients with anorexia nervosa is more effective than the established treatments alone.

1.3 Categorical perception

A recent addition to the repertoire of body image specific interventions has been to manipulate the categorical boundary for what participants judge, subjectively, to be a thin versus a fat body. Categorical perception is a phenomenon whereby a continuous sequence of equal physical changes in a stimulus can be perceived and grouped as two different categories of stimuli (Liberman, Harris, Hoffman, & Griffith, 1957; Scott & Evans, 2010). Recent research has shown that both faces and bodies can be judged in a categorical manner: e.g., healthy versus unhealthy, or angry versus happy for faces, and thin versus fat for bodies (Calder et al., 1996; Tovée, Edmonds, & Vuong, 2012). Accordingly, Penton-Voak et al. (2013) used feedback-based training to shift the categorical boundary in the recognition of emotion in faces. They used a sequence of face images that morphed smoothly from clearly happy at one end of the continuum to clearly angry at the other end. Images in the middle of the sequence (intermediate in their expression) could be judged either way. Penton-Voak et al. (2013) showed that adolescents at high risk of offending, unlike controls, tend to categorize even the intermediate faces as angry. However, by giving appropriately structured feedback, their angry–happy categorical boundary could be shifted

towards the ‘happy’ end of the spectrum. Penton-Voak et al. (2013) found that this resulted in a decrease in self-reported anger and aggression and in independently rated aggressive behaviour. Gledhill et al. (2017) modified this training technique to shift participants’ subjective categorical boundaries for what constituted a thin versus a fat version of a standard model, towards higher BMIs. Here, a person’s categorical boundary is the point in the BMI spectrum at which bodies change from being categorised by the participant as thin to being categorised as fat (see Figure 1 for conceptual illustration).

Intriguingly, this perceptual shift in judgements of *others* was accompanied by reductions in participants’ psychological concerns about their *own* body shape, weight and eating, and these effects persisted for two weeks post-training. In a second experiment, Gledhill et al. (2017) found similar effects in a sample of women with a history of anorexia, although the perceptual changes took longer to emerge during the training period. Post-training, the reductions in the anorexics’ psychological concerns about their own bodies persisted up to a month from initial testing. In a very similar study, Szostak (2018) used the same paradigm to manipulate participants’ interpretation of body size, and to encourage their interpretation of thinness over heaviness in normal-sized bodies, in women with heightened body shape, weight and eating concerns. Again, the training successfully shifted participants’ categorical boundaries towards larger bodies, and this was retained at 2-week follow-up. Furthermore, in the intervention group the shift in threshold was accompanied by more favourable attractiveness and ideal ratings of heavier stimulus bodies at 2-week follow-up, in contrast to baseline, along with lower levels of dissatisfaction with the participants’ own bodies. The success of these two studies demonstrates that this approach may be an important additional technique to support recovery in eating disorders.

1.4 Potential advantages of virtual reality

In the last two decades, researchers have explored the use of virtual reality in a variety of psychological treatments. It has proved useful as a safe alternative to interventions that would otherwise require imagination, memory or in-vivo exposure, for example in the treatment of anxiety disorders and phobias (Carl et al., 2019; Freeman, Reeve, Robinson & Ehlers, 2017; Maples-Keller et al., 2017). Virtual reality has also been used to investigate body image and treat eating disorders and obesity (Fonseca-Baeza, Corno, & Baños, 2018; Gutiérrez-Maldonado, Wiederhold, & Riva, 2016; Perpiñá et al., 1999; Perpiñá, Botella, & Baños, 2003; Riva, 1998; Riva, Bacchetta, Baruffi, & Molinari, 2001; Riva et al., 2018). Systematic reviews of this work have shown promising results, and indicate that large controlled trials of clinical samples are warranted (Clus, Larsen, Lemey & Berrouiguet, 2018; De Carvalho, Dias, Duchesne, Nardi & Appolinario, 2017; Ferrer-Garcia, Gutiérrez-

Maldonado, & Riva, 2013; Koskina, Campbell, & Schmidt, 2013). The essence of virtual reality is the immersive, interactive relationship between the user and the virtual environment, producing a sense of presence (Freeman et al., 2017) or “being there” (Gorini, Capideville, De Leo, Mantovani, & Riva, 2011; Riva & Waterworth, 2003; Riva, 1998; Riva, Wiederhold & Mantovani, 2019). Moreover, virtual reality has also been shown to be as effective as reality in inducing emotional responses (Felnhofer et al., 2015; Ferrer-García, Gutiérrez-Maldonado, Caqueo-Urizar, & Moreno, 2009; Vincelli, 1999). It is these compelling qualities of virtual reality that we wish to exploit for the current study. Ultimately, we propose that the training effects reported by Gledhill et al. (2017) and Szostak (2018) may be stronger and more persistent if carried out in virtual reality as opposed to standard 2D stimulus presentation. Evidence consistent with this idea comes from a wide variety of domains including: comparisons of 2D versus 3D views for endoscopic keyhole surgery (Fraser, Allen, Anand & Schwartz, 2009), post-stroke upper limb retraining in virtual reality (Levin, Snir, Liebermann, Weingarden, & Weiss, 2013), and complex multi-variate data exploration in 2D and VR (Millais, Jones, & Kelly, 2018), all of which suggest that performance may be enhanced in virtual reality compared to standard 2D image presentation.

1.5 The current study

As a first step in this direction, our aim is to test the feasibility of running the training paradigm described by Gledhill et al. (2017) and Szostak (2018) in virtual reality. If successful, this would open the door to: (a) RCTs which investigate the effects of this training on people who have anorexia nervosa, and (b) formal comparisons of the efficacy of this perceptual training paradigm, comparing conventional 2D stimulus presentations with those presented in 3D in virtual reality.

In the current study, we investigated a sample of women with heightened, sub-clinical body image concerns. We set out to replicate as closely as possible the methodology of Gledhill et al. (2017). However, one striking feature of the paradigms described by Penton-Voak et al. (2013), Gledhill et al. (2017), and Szostak (2018), is the use of very brief stimulus presentations. Specifically, the images of the faces/bodies to be judged were presented on flat screens for 150ms and were then masked. Yet, no a-priori justification was given for this aspect of the task design. Moreover, when piloting our paradigm in VR, participants commented frequently that such short presentations made it very difficult for them to make their judgements. This raised the possibility that what is effective and compelling for 2D stimulus presentations may not necessarily hold for VR. Furthermore, literature on social comparison theory suggests that women evaluate and compare themselves against other

women in real-life situations with the potential for improving their attractiveness (Bailey & Ricciardelli, 2010; Cundall & Guo, 2015; Festinger, 1954; Gervais, Holland, & Dodd, 2013; Leahey et al., 2007). Even if it is a “glance-over,” this process may take time, and certainly longer than the brief 150ms presentations used previously. Therefore, in the current study, we added a third condition, referred to as the ‘new intervention condition’, where the stimuli were presented to the participant without a time limit.

In summary, we asked women with high body image concerns to categorise a series of body stimuli in order to ascertain, and then modify, their subjective thin/fat categorical boundaries in one of three conditions: (a) an original intervention condition with stimulus presentation timings matching those of Gledhill et al. (2017), (b) a new intervention condition with untimed stimulus presentation, and (c) a control condition with timings matching Gledhill et al. (2017). Our main predictions were that we should be able to replicate the results of Gledhill et al. (2017) and show that: (a) the original intervention condition in VR, with brief presentations, should increase participants’ categorical boundaries towards fatter bodies; (b) no such shift should occur for the control condition; (c) a shift in perceptual categorical boundary should be associated with a post-training reduction in psychological concerns about body shape, weight, and eating. Finally, we also expected that the new intervention condition, with unrestricted viewing times, should at least match, but hopefully exceed the effects seen with the old intervention condition.

2. Method

2.1 Participants

To estimate appropriate sample sizes for the current study, we used data from Gledhill et al. (2017). Specifically, we used the means and standard deviations from both the perceptual boundary and psychometric outcomes (i.e. the global score for the EDE-Q questionnaire) obtained from the intervention and control groups on the last day of training (Day 4) and 10 days post-training (Day 14). We used PROC POWER (SAS v9.4) to estimate sample sizes sufficient to distinguish between the means for the intervention group and controls at each time point, with a power of 0.9 at an alpha of .05. For the perceptual training data on Days 4 and 14, the integer sample size estimates were 19 and 20, respectively. For the EDE-Q, the corresponding estimates were 20 and 19, respectively. Therefore, for the current study, we used 20 participants per comparison group.

All experimental procedures and methods for participant recruitment for this study were approved by the ethics committee at Northumbria University. Participants were recruited from undergraduate and postgraduate student populations at Northumbria University and the University of Lincoln. To be eligible to take part in this study, participants had to be female (CIS/as assigned at birth), aged 18-35, fluent in spoken English and have no history of eating disorders. Additionally, potential participants were asked to complete an online version of the Body Shape Questionnaire (BSQ-16b; Evans & Dolan, 1993). The questionnaire measures attitudes towards body shape and size. A score of 66 and above indicates ‘marked concern with shape’ (Taylor, 1987) and is considered clinically concerning. However, to be consistent with the study design of Gledhill et al. (2017), we accepted participants who scored 60 or above, i.e., ‘substantial, yet sub-clinical body shape concerns’ (Taylor, 1987). We collected online BSQ responses from 333 individuals, of whom 95 met the criterion for study inclusion (i.e., BSQ of 60 or above). Of those, 33 participants did not respond to invitations to attend the training, 62 attended the training and 2 did not fully complete the training. Participants who completed the pre-screen questionnaire were awarded research participation points for their time. Participants who completed the full training received a £20 gift voucher as well as points. The consenting participants were randomly assigned to one of three training conditions, and their characteristics are reported in Table 1. The right-hand columns of Table 1 show the output of pairwise comparisons between the three subgroup means, adjusted for multiple comparisons, using the permutation method in PROC MULTTEST (SAS v9.4).

2.2 Materials

2.2.1 Psychometric and anthropometric measures

We administered a set of well-established, validated, self-report questionnaires to assess participants’ attitudes towards their body shape/size, weight and eating, as well as their tendency towards depression, and their self-esteem. The following questionnaires were used:

The Eating Disorders Examination Questionnaire (EDE-Q; Fairburn & Beglin, 1994) is a self-report version of the Eating Disorders Examination (EDE) interview (Fairburn & Beglin, 1993). The questionnaire contains four subscales: (a) the Restraint (EDE-Q res) subscale contains 5 items which measure the restrictive nature of eating; (b) the Eating Concern (EDE-Q eat) subscale contains 5 items which measure the preoccupation with food and social eating; (c) the Shape Concern (EDE-Q SC) subscale contains 8 items which measure dissatisfaction with body shape; (d) and the Weight Concern (EDE-Q WC) subscale contains 5 items which

measure dissatisfaction with body weight. Participants report how many days of the past four weeks they have experienced an item, e.g., ‘Have you been deliberately trying to limit the amount of food you eat to influence your shape or weight (whether or not you have succeeded)?’ on a 7-point response scale from 0 indicates (*no days*) to 6 (*every day*). A global score of overall disordered eating behaviour is also calculated by averaging the appropriate items, and frequency data on key behavioural features are recorded. Cronbach’s alpha for this measure was .86 across all participants.

The 16-item Body Shape Questionnaire (BSQ-16b; Evans & Dolan, 1993) was used to assess size and shape concerns, e.g., ‘Have you been so worried about your shape that you have been feeling you ought to diet?’ Items are rated along a 6-point response scale, from 1 (*never*) to 6 (*always*). Items are summed for a total score. Cronbach’s alpha for this measure was .92 across all participants.

The Beck Depression Inventory (BDI; Beck, Steer, Ball, & Rainieri, 1996) was used to measure levels of depressive symptomatology. It is a behavioural checklist that contains 21 items such as ‘loss of interest,’ ‘sadness,’ and ‘self-dislike.’ Each item is rated on a 4-point scale, ranging from 0 (*no symptom of depression*) to 3 (*severe expression of a depressive symptom*). Items are summed for a total score. Cronbach’s alpha for this measure was .87 across all participants.

The Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965) was used to assess self-esteem by reflection on current feelings. The 10 items are rated on a 4-point scale (0-3) from 0 (*strongly disagree*) to 3 (*strongly disagree*). Five of the items have positively worded statements, e.g., ‘On the whole I am satisfied with myself’ and five are worded negatively, e.g., ‘At times I think that I am no good at all.’ Items are summed for a total score. Cronbach’s alpha for this measure was .86 across all participants.

Participants’ body mass index (BMI) was calculated from their weight and height measures with a set of calibrated scales and a stadiometer.

2.2.2 Stimulus creation

Using a base model provided by the Manuel Bastioni Laboratory we created 15 3D models for our stimulus set in Unreal Engine v4.18. We followed Gledhill et al. (2017) who, based on a number of pilot experiments, selected 15 models whose BMIs increased from 15.45 to 33.70 in equal steps of 1.30 BMI units. We therefore replicated this BMI range for our stimuli. The models’ height was set to 1.65m, which is the average

height of Caucasian women aged 25 in the Health Survey for England (HSE) 2008 dataset. The models' BMIs were computed from calibration equations predicting BMI from waist and hip circumference, age, sex, height and ethnicity in the Health Survey for England databases (Cornelissen, 2016). Figure 2 shows all 15 stimuli in front view with the relationship between the BMIs of the stimuli and the World Health Organisation's BMI classification (World Health Organisation, 2003). The models' heads were masked by white ovoids to hide facial features, ensuring that participants attended only to body shape.

2.2.3 The perceptual training paradigm

Like the Gledhill et al. (2017) study, perceptual training was carried out over 4 consecutive days. At the start of each day, a baseline measurement was obtained to establish the categorical boundary for each participant on that day. It comprised three presentations of each of the 15 models in a randomised order, totalling 45 trials. Responses were used to calculate the categorical boundary at which participants shifted from perceiving thinness to perceiving fatness. The control and original intervention conditions (as well as the timed baselines for these conditions) were set up to match as closely as possible the control and intervention conditions in the 2D study by Gledhill et al. (2017). Specifically, the sequence of events visible in the virtual reality headset with respect to the structure of the fixation cross, the visual angle subtended by the figure in the scene and the size and structure of the mask. We also matched the timings of all events. Accordingly, each trial began with a central fixation cross which was shown for 1500-2500ms (randomly jittered). The cross was then replaced by a body model (for 150ms), followed immediately by a visual noise mask (for 150ms). Two virtual boxes then appeared in front of the participant, one read 'fat' and the other 'thin,' thereby presenting the participant with a 2-alternative forced choice. The participant was asked to respond using the Oculus Touch controller to tap the box of their choice.

The new intervention condition was designed to be optimised for virtual reality presentation. Specifically, to take full advantage of the potentially improved presentation (3D, life-sized), a quick presentation akin to that of Gledhill et al. (2017) was felt to be too abrupt. Therefore, time constraints for stimulus presentation and the visual noise mask were removed. Instead, a fixation cross appeared first (1500-2500ms, randomly jittered). This was followed by the body model along with the response boxes, all of which remained visible for as long as it took participants to make a decision. This sequence of events felt much more natural, and therefore VR-friendly. This unlimited exposure was applied to both baseline and training parts of the new intervention.

In the training sequence, in accordance with Gledhill et al. (2017), Bodies 1, 2, 14 and 15 were presented once. Bodies 3-5 and 11-13 were presented twice, and Bodies 6-10 were presented 3 times each, totalling 31 trials. The training phase on each day differed from the baseline measurement in that participants were given feedback on every trial once they had responded. The feedback was displayed in large writing in front of the participant, e.g., ‘Correct! That body was fat’ or ‘Incorrect! That body was thin.’ The nature of the feedback in both the original and new intervention conditions was inflationary. It was designed to shift participants’ categorical boundary by two bodies further along the sequence shown in Figure 1, based on their baseline categorical boundary for the day. This way, participants were re-trained to judge bodies near their categorical boundary which they previously judged as ‘fat’ to be ‘thin.’ The nature of the feedback for the control condition was not inflationary. Instead, it was consistent with their categorical boundary for the day as measured at baseline and was intended to reinforce this. In all cases, once the training sequence had been administered, a post-training categorical boundary was calculated. See Figure 3 for illustration of the training sequence.

2.3 Procedure

Consenting participants attended the perceptual training sessions over 4 consecutive days. On Day 1, participants’ weight and height were measured, and the questionnaires were completed. The questionnaires were completed in randomised order. Participants were then instructed on how to wear the Oculus Rift headset, adjust the focus to ensure the stimulus was clearly visible, and shown how to use the Touch Controllers. They then carried out the first baseline measurement, followed by the first training sequence. On Days 2 and 3, participants carried out the perceptual baseline and training only. On Day 4, the participants carried out the baseline and training sequences and then completed the questionnaires. On Day 14, at follow-up, participants carried out the baseline sequence only and completed the questionnaires. On full completion of the study, the aims of the study were outlined to participants, they were verbally debriefed, they were also given a debrief sheet and gift card.

2.4 Analysis pipeline

The main analyses of the experimental data included the following steps:

- Calculation and tabulation of univariate descriptive statistics for participants’ characteristics and their psychometric performance.

- Computation of linear mixed effects model of participants' categorical boundaries explained by experimental condition (i.e., control, original intervention, new intervention), training (i.e., baseline measurement for the day versus post-intervention), and measurement time point (i.e., Day 1, Day 2, Day 3, Day 4, Day 14).
- Data reduction of the psychometric responses, using principal component analysis, to produce one latent variable. High scores on this latent variable, henceforth referred to as PSYCH, reflect a combination of high body image concerns, pathological attitudes to eating, low self-esteem, and depressed mood. Low scores reflect the opposite; confidence in one's body, the absence of eating pathology, high self-esteem, and the absence of depressive thoughts.
- Computation of linear mixed effects model of participants' PSYCH responses explained by experimental condition (i.e., control, old intervention, new intervention), and measurement time point (i.e., Day 1, Day 4, Day 14).

3. Results

3.1 Perceptual training

We wanted to test: (a) whether we could replicate the perceptual training effects reported by Gledhill et al. (2017) when carried out in virtual reality, using the same brief presentation of stimulus and mask; (b) whether we could obtain a perceptual training effect in virtual reality using the new intervention condition with longer stimulus presentation times; (c) whether we could confirm an absence of perceptual training with the control condition.

To do this, we fitted a linear mixed effects model for the perceptual training data using PROC MIXED in SAS v9.4 (SAS Institute, North Carolina, USA). The model included fixed effects for: (a) time (i.e., test Day: 1, 2, 3, 4 and 14), (b) intervention (pre-training baseline threshold vs post-training threshold), and (c) condition (original intervention, control intervention, new intervention), and we tested all possible two-way and three-way interactions. We tested both individual slope and intercept variation for each subject and specified an 'unstructured' variance-covariance matrix for the random effects. For dummy coding, original intervention (for condition), post-training (for intervention), and Day 14 (for time) were treated as the reference levels.

The -2 Log Likelihood for an intercept only model was 2155.2, compared to 1944.0 for the final model. Subject variance was 3.80, which was statistically significant ($z = 5.10, p < .0001$). For the type III tests of fixed

effects, we found significant effects of time $F(4,456) = 15.51, p < .0001$, training $F(1,456) = 130.68, p < .0001$, and condition $F(2,57.2) = 5.12, p = .009$. We also found significant two-way interactions between time and condition $F(8, 456) = 7.02, p < .0001$, and intervention and condition $F(4, 456) = 12.53, p < .0001$, but no significant two-way interaction between intervention and time, or three-way interaction between time, intervention and condition. The effect size for the random effect can be estimated from the intraclass correlation (ICC, Snijders & Bosker, 2012) and this was .64. With respect to the fixed effects, $f^2 = .21$ for the overall model (Lorah, 2019), suggesting that together, intervention, time, and condition explain 21% of the variance in perceptual training relative to the unexplained variance in perceptual training. Guidelines for interpretation of f^2 indicate that .02 is a small effect, .15 is a medium effect, and .35 is a large effect (Cohen 1992), indicating that the present effect is medium.

Figure 4a shows plots of the least squares means (LSmeans) for the categorical boundary, derived from our linear mixed effects model, as a function of training day. The data are plotted separately for the different conditions: control intervention (dark gray), new intervention (white), and original intervention (black), and in each case, split further according to whether the measurement was pre-training (dashed) or post-training (solid lines). Figure 4b shows the LSmeans difference between baseline and post-training thresholds as a function of training day, separately for the three groups. The error bars represent the 95% CIs for these differences. Figure 4 shows a statistically robust effect of training on each training day for the new intervention condition because none of the 95% CIs straddle zero. For the original intervention, only Days 3 and 4 showed a robust training effect. For the control intervention, there were no statistically robust differences between pre- and post-training on any day. Finally, to test whether any perceptual training effects persisted over time, we carried out three planned comparisons between the baseline categorical boundary thresholds measures on Day 1 and Day 14. This was calculated as Day 14 – Day 1, so that positive values represent increases in BMI at the categorical boundary. Pairwise comparisons were adjusted for multiple comparisons using the simulate option for the LSmeans command in PROC MIXED (SAS v9.4). For the new intervention, we found a significant increase of 2.8 BMI units at the categorical boundary, $t(456) = 7.10, p < .0001, p_{adjusted} < .0001$, Hedges' $g = 1.19$, between Day 1 ($M = 24.12$) and follow-up at Day 14 ($M = 26.92$). (To express effect size for t-tests we used Hedges' g instead of Cohen's d , owing to the relatively small sample size [see e.g., Lakens, 2013]). For the original intervention, we found a more modest increase of 0.98 BMI units at the categorical boundary, $t(456) = 2.47, p = .01, p_{adjusted} = .04$, Hedges' $g = 0.52$, between Day 1 ($M = 23.73$) and follow-up at Day 14 ($M = 24.70$). In comparison, the

difference of -0.39 for the control intervention was not statistically significant, $t(457) = 0.99, p = .32, p \text{ adjusted} = .69$, Hedges' $g = 0.19$; Day 1 $M = 24.77$; follow-up at Day 14 $M = 24.38$. This suggests that both the original and new training paradigms in virtual reality cause accumulating shifts in participants' categorical boundaries towards heavier bodies, which were retained for 2 weeks. However, the training effect for the new training paradigm at Day 14, with longer exposure of the stimuli than was the case in Gledhill et al. (2017), was significantly larger than that for the original training paradigm ($M = 2.22, t(102.3) = 3.03, p = .003, p \text{ adjusted} = .01$, Hedges' $g = 1.06$).

3.2 Psychometric measures

In Gledhill et al. (2017), only changes in EDE-Q over the course of training and follow-up are reported. BSQ was not measured (except to identify putative participants with high body image concerns) and neither was RSE nor BDI. For the current study, Table 2 shows the Pearson correlations between the four psychometric measures, across the entire sample, at each of the three measurement points, on Days 1, 4, and 14. It is clear that on each day there are statistically significant, substantial correlations between all four tasks. Given that we want to know whether perceptual training has a knock on effect on participants' psychological state, and in the face of these correlations, we decided to use principal component analysis (PCA) to compress these four psychometric outcome variables into a single latent variable (cf. Cornelissen, Bester, Cairns, Tovée, & Cornelissen, 2015; Cornelissen, McCarty, Cornelissen, & Tovée, 2017; Irvine et al., 2019). Low scores on this latent variable, henceforth referred to as PSYCH, reflect a combination of high body image concerns, low self-esteem, and depressed mood. High scores reflect the opposite; confidence in one's body, high self-esteem, and the absence of depressive thoughts. To do this, we used PROC FACTOR in SAS v9.4 (SAS Institute, North Carolina, US) to carry out a principal component analysis. We used the factor score from the component in our statistical models. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (which indicates the degree of diffusion in the pattern of correlations) was 0.70 suggesting a middling acceptable sample. Bartlett's test of sphericity was statistically significant ($X^2 = 328.6, df = 6, p < .0001$). One principal component (PC) had Eigen values greater than Kaiser's criterion of 1 (i.e., 2.68) which explained 67% of the variance. The scree plot showed an inflexion, i.e., Cattell's criterion, which also justified retaining just one PC. The factor loadings on PSYCH for EDE-Q, BSQ, BDI, and RSE were: .82, .79, .83, and -.83 respectively.

We used PROC MIXED in SAS v9.4 (SAS Institute, North Carolina, USA) to fit a linear mixed effects model to participants' PSYCH scores measured on Days 1, 4 and 14. Individual variation was permitted at the intercept level for each subject by including a random effect for the intercept. For dummy coding, original intervention (for condition) and Day 14 (for time) were treated as the reference levels. The -2 Log Likelihood for an intercept only model was 420.1 compared to 385.0 for the final model. Subject variance was 0.70, which was statistically significant ($z = 4.83, p < .0001$). The type III tests of fixed effects were: time $F(2,114) = 22.78, p < .0001$; condition $F(2,114) = 2.61, p = .07$; condition $\times$ time $F(4,114) = 1.48, p = .21$, respectively. For the random effect of the model, ICC was .76, and for the fixed effects, $f^2 = .11$, the latter indicating a small to medium effect size. Table 3 shows the means and standard deviations of the standardized factor scores for PSYCH, separately for condition and time. In addition, Table 3 includes the planned comparisons between Day 1 and Day 4, as well as between Day 1 and Day 14. This shows that over the course of two weeks, there is no significant influence of the control intervention on PSYCH scores. However, both intervention methods in virtual reality produced significant reductions in PSYCH scores over the same time period.

A key question is whether the reductions we observed in psychological concerns in our sample may have been clinically meaningful (Kazis, Anderson, & Meenan, 1989). Clearly, we have not tested a patient group in this study. Nevertheless, participants were selected on the basis of BSQ scores which showed 'substantial, yet sub-clinical body shape concerns' (Taylor, 1987). Therefore, we followed the spirit of Kazis et al. (1989) to ask whether the statistically robust effects we found for PSYCH may at least be consistent with clinically meaningful effects. With respect to the EDE-Q, Bardone-Cone et al. (2010) operationalise recovery in eating-disordered patients as a reduction in all four subscale scores to within 1SD of age-matched community norms. Mond et al. (2006) reported such norms for the 18-22 age group based on a sample of 1186 women. These are replicated in Table 4, along with the mean EDE-Q subscale scores on Day 14 for each group. We can conclude that the new intervention produced a reduction in all 4 EDE-Q subscale scores that was clinically meaningful when defined this way. However, this result needs to be treated with caution. While there were no statistically significant group differences in the global EDE-Q scores between the groups on Day 1, nevertheless there was a trend whereby the scores for Original intervention $>$ Control $>$ New intervention. Therefore, it is possible that it was easier for the EDE-Q subscale scores for the New intervention group to fall into the normative range on Day 14, because they were the lowest on average to start with. Finally, in order to facilitate comparison with Gledhill et al. (2017),

Table 1 in Supplementary Materials, shows the comparison between Days 1 and 4 and Days 1 and 14 for the BSQ, RSE, BDI, EDE-Q, and the four EDE-Q sub-scales.

4. Discussion

Given that persistent disturbances in body image increase the likelihood of relapse in anorexia nervosa (Carter et al., 2004), efforts have been made to develop specific treatments to reduce body image distortion. Here, we tested whether the perceptual training technique used by Gledhill et al. (2017) and Szostak (2018) can be transferred successfully to immersive virtual reality. We asked women with high body image concerns to view images of a standard female CGI model in virtual reality. The model varied in BMI across trials in the task, and participants made subjective judgements whether each image was thin or fat. Based on these responses, we ascertained each participant's thin/fat categorical boundary. Over four days, participants were allocated to one of two intervention conditions: (a) an original intervention condition with stimulus presentation timings matching those of Gledhill et al. (2017), (b) a new intervention condition with untimed stimulus presentation, or (c) a control condition with timings matching Gledhill et al. (2017). We found that the perceptual training in virtual reality successfully shifted participants' categorical boundaries towards higher BMI bodies in both intervention groups, but not the controls. Moreover, this effect was stronger for the participants who were presented stimuli with no time constraints (new intervention) as compared to those who were given very brief presentations (original intervention). With respect to participants' PSYCH scores, representing a composite of their attitudes towards body shape, weight, eating, depressive symptoms and self-esteem, both new and original interventions, but not the controls, showed significant reductions in psychological concerns and these were of similar magnitudes. Intriguingly, a break-down of the sub-scale scores for the EDE-Q showed that the reductions in participants' concerns extended beyond their own body shape and weight and included their attitudes towards eating. These findings therefore confirm that the perceptual training technique, reported by Gledhill et al. (2017) and Szostak (2018), can successfully be translated into immersive virtual reality.

4.1 Differential effects due to speed of presentation in the perceptual training

In the current study, the increases in mean BMI at the categorical boundary between Days 1 and 4 for the new intervention and the old intervention were 1.95 and 0.52 respectively. The corresponding differences for Days 1 and 14 were 2.80 and 0.97 BMI units. Clearly, the perceptual training effect was stronger for the new intervention with unrestricted presentation times than the old intervention with brief presentations. In Gledhill et

al. (2017), using briefly presented 2D stimuli, the increases for the intervention condition between Days 1 and 4 and 1 and 14 were 1.45 and 1.54 BMI units, respectively. This implies that for *brief* presentations there is a distinct advantage to presenting stimuli in 2D rather than 3D in virtual reality.

There are two reasons why brief presentations in virtual reality may have produced smaller training outcomes than either brief presentations in 2D or longer presentations in virtual reality. The first has to do with possible fixation disparities that may exist in virtual reality when the fixation cross is replaced by the body, depending on which depth plane on the body participants were attempting to fixate. In order to correct these disparities, and obtain an optimal view of the body, small vergence eye movements would be required. For vergence steps of 4°, the latency of the left and right eye movements can be as long as ~250ms (Yaramothu, Jaswal, & Alvarez, 2019), which is nearly twice the duration of our brief stimulus presentations. While it is unlikely that any residual fixation disparity may have been as large as 4°, nevertheless, such delays may have contributed to poorer performance in the old intervention in virtual reality. This problem would not exist for the 2D displays in Gledhill et al. (2017), because the fixation cross and body stimulus were presented in exactly the same depth plane.

A second reason for these discrepancies may have to do with the distinction between 2D and 3D object recognition. This is a profound difference owing to out-of-plane pose variation because, as an object rotates in 3D, different parts of it are seen (Elder, 2018). Two major theories for how the brain handles this problem have been proposed. In the 3D model-based theory, the brain uses 2D features of the image projection of an object to estimate the 3D geometry and pose of the object, allowing comparison with stored 3D object models (e.g., Biederman 1987). In the alternative view-based theory (e.g., Bülthoff & Edelman 1992), the brain maintains models for the 2D appearance of remembered objects over a sampling of 3D poses. The 2D image appearance of an object segmented from a new image can then be compared to stored 2D views to identify both the identity and pose of the new object (Elder, 2018). Either theory implies a considerable addition to the brain's computational load when dealing with 3D images. Indeed, Turnbull, Driver, and McCarthy (2004) presented a patient, DM, whose visual agnosia still permitted reliable extraction of 2D information from the visual world but compromised his ability to extract pictorial information about 3D object-structure. This confirms that aspects of 3D vision can be dissociated from 2D vision. Moreover, Van Orden and Broyles (2000) compared human performance on several 2D and 3D display formats across four visuospatial tasks. They found that 2D plan or side-view displays yielded performance as good or better than any other display system for speed and altitude judgment tasks. This

implies that, when time is limited, there may well be situations where 2D views can be judged more reliably than 3D ones. Ultimately, however, further research is needed to adjudicate between these possibilities.

4.2 Changes to the categorical boundary

Does the perceptual training have a purely perceptual basis? It could function as a form of exposure therapy, desensitising the participants to women's bodies of different BMI levels (Koskina et al., 2013; Trentowska, Svaldi, & Tuschen-Caffier, 2014). Alternatively, it could function by expanding the range of images with which the participant is familiar, thus placing their own BMI in context. In short, broadening the participants' visual diet. Consistent with both explanations, our results suggest that longer stimulus presentation was more effective in creating a shift in the categorical boundary. However, a purely perceptual explanation would suggest that the feedback component of the training is unnecessary and the control condition without feedback should be as effective as the other two conditions. The fact that it is relatively ineffective in changing the categorical boundary and any subsequent psychological change suggests that the involvement of cognitive mechanisms is necessary for the training to work.

It is likely that the training does not significantly impact the judgement of the absolute size of the bodies (i.e. their physical dimensions) as although women with high body image concerns over-estimate their own body size, they are as accurate as controls in estimating the size of other women's bodies (Gledhill et al., 2019; Thaler et al., 2018). However, women with high body concerns may *categorise* a body as being fat at a lower BMI than controls (Gledhill et al., 2017), and it is this categorisation process – i.e., a participant's decision criterion for thin versus fat - that is being modified, rather than the perceptual accuracy of size estimation. This modification of the categorical boundary does not seem to be a passive process mediated through adaptation but instead needs the active engagement of the cognitive processes such as in making a 2-AFC (two forced-choice) judgement. Simple presentation of the stimuli without feedback on the accuracy of each judgement is insufficient to recalibrate the position of the categorical boundary.

4.3 Changes to psychometric performance

The training programme is designed to shift the categorical boundary at which *other* women's bodies are judged to be fat rather than thin. An obvious question is why should this effect our participant's concerns about their *own* body and eating? Socio-cultural theories of body image, such as the Tripartite Influence Model and the Dual Pathway Model, suggest associations between the existence of appearance-related norms and the making of

social comparisons in order to explain the development of dissatisfaction with one's body image (Festinger, 1954; Polivy & Herman, 2002; Stice 2002; Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999). When we judge our own bodies, we do so within the framework of appearance-related norms (Stice & Shaw, 2002; Tiggemann, 2011). Women who have high levels of body dissatisfaction, such as our sample, are thought to internalise body ideals to a great extent (Smolak & Thompson, 2009) and are prone to comparing themselves against such standards (Schaefer & Thompson, 2014; Thompson et al., 1999). Thus, comparison of their own size with respect to the norms for size can lead to dissatisfaction, so that recalibrating these benchmarks may lead to a reduction in body image dissatisfaction. As body dissatisfaction is a core feature of eating disordered concerns and behaviour, reducing body dissatisfaction may, in turn, reduce the eating disordered concerns indexed by the EDE-Q (Brechan & Kvalem, 2015; Sirirassamee, Phoolsawat, & Limkhunthammo, 2018).

Concretely, the average actual BMI in the new intervention group was 26.59, and their average initial categorical boundary was lower than this, at a BMI of 24.12. According to their own categorisation of the standard model, these participants would therefore categorise themselves as 'fat,' which would be consistent with their high body dissatisfaction scores. However, at follow-up, this group's categorical boundary had increased to an average BMI of 26.92. This would have meant that they now compared more favourably to the standard model, because they were, on average, slimmer than those models that they now considered to be 'fat.' Such a shift could potentially explain the changes in their psychological scores, as downward comparisons have been shown to enhance self-evaluation (Bailey & Ricciadelli, 2010). As levels of satisfaction with one's own body stem from discrepancies between ideal and actual body size, narrowing or shifting this gap may therefore have led to increased body satisfaction and change related attitudes, as illustrated by the changes in PSYCH, BSQ and EDE-Q scores. Additionally, the cognitive-behavioural model (Cash, 2012) proposed that the development of body image perception and attitudes are shaped, amongst others, by interpersonal experiences, while Bouton (2011) suggested the importance of learning processes in the development of body image and eating pathology. The feedback we provided to our participants is an example of such experience and/or learning. In the new intervention condition, we have demonstrated that feedback has changed the perceptual boundary, i.e., it has changed how shape/size was interpreted, which also led to observed changes in body shape concerns, weight concerns and eating attitudes as measured by the BSQ and EDE-Q.

4.4 Limitations

The stimulus timing between the Gledhill et al. (2017) 2D intervention and the current ‘original intervention’ condition in virtual reality was matched as closely as possible. In addition, we matched the 15 CGI models in the current study to the same heights, hip and waist circumferences and BMIs as those in Gledhill et al. (2017). However, because the CGI model itself differed between the two studies, it is not legitimate to make a direct statistical comparison between them. Therefore, given the present encouraging results, it is clear that we need to make a formal comparison using a factorial design crossing 2D with 3D presentation in VR with brief versus unrestricted presentation times, all based on the same CGI model. In a similar vein, the feasibility and efficacy of this perceptual training needs to be tested with a clinical sample in an appropriately powered randomised control trial. It is possible that the concerns of women with eating disorders may be harder to modify, as they tend to be more severe than the kinds of participants recruited to the current study. Nevertheless, if successful, this approach may represent a new way to treat persistent body image disturbance, deliverable perhaps alongside established treatments, such as cognitive behavioural therapy.

References

- Alleva, J., Jansen, A., Martijn, C., Schepers, J., & Nederkoorn, C. (2013). Get your own mirror: Investigating how strict eating disordered women are in judging the bodies of other eating disordered women. *Appetite*, 68, 98–104. <https://doi.org/10.1016/j.appet.2013.04.015>
- Bachner-Melman, R., Zohar, A. H., & Ebstein, R. P. (2006). An examination of cognitive versus behavioral components of recovery from anorexia nervosa. *The Journal of Nervous and Mental Disease*, 194, 697–703. <https://doi.org/10.1097/01.nmd.0000235795.51683.99>
- Bailey, S. D., & Ricciardelli, L. A. (2010). Social comparisons, appearance related comments, contingent self-esteem and their relationships with body dissatisfaction and eating disturbance among women. *Eating Behaviors*, 11, 107–112. <https://doi.org/10.1016/j.eatbeh.2009.12.001>
- Bardone-Cone, A. M., Harney, M. B., Maldonado, C. R., Lawson, M. A., Robinson, D. P., Smith, R., & Tosh, A. (2010). Defining recovery from an eating disorder: Conceptualization, validation, and examination of psychosocial functioning and psychiatric comorbidity. *Behaviour Research and Therapy*, 48, 194–202. <https://doi.org/10.1016/j.brat.2009.11.001>
- Bardone-Cone, A. M., Hunt, R. A., & Watson, H. J. (2018). An overview of conceptualizations of eating disorder recovery, recent findings, and future directions. *Current Psychiatry Reports*, 20, 79. <https://doi.org/10.1007/s11920-018-0932-9>
- Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of Personality Assessment*, 67, 588–597. https://doi.org/10.1207/s15327752jpa6703_13
- Berkman, N. D., Lohr, K. N., & Bulik, C. M. (2007). Outcomes of eating disorders: A systematic review of the literature. *International Journal of Eating Disorders*, 40, 293–309. <https://doi.org/10.1002/eat.20369>
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.
- Birmingham, C. L., Su, J., Hlynsky, J. A., Goldner, E. M., & Gao, M. (2005). The mortality rate from anorexia nervosa. *International Journal of Eating Disorders*, 38, 143–146. <https://doi.org/10.1002/eat.20164>

- Bouton, M. E. (2011). Learning and the persistence of appetite: Extinction and the motivation to eat and overeat. *Physiology and Behavior*, 103, 51–58. <https://doi.org/10.1016/j.physbeh.2010.11.025>.
- Brechan, I., & Kvalem, I. L. (2015). Relationship between body dissatisfaction and disordered eating: Mediating role of self-esteem and depression. *Eating Behaviors*, 17, 49–58. <https://doi.org/10.1016/j.eatbeh.2014.12.008>
- Bruch, H. (1962). Perceptual and conceptual disturbances in anorexia nervosa. *Psychosomatic Medicine*, 24, 187–194.
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 60–64.
- Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, 3, 81–117. <https://doi.org/10.1080/713756735>
- Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, R., Emmelkamp, P., Asmundson, G. J. G., Carlbring, P., & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders*, 61, 27–36. <https://doi.org/10.1016/j.janxdis.2018.08.003>
- Carter, J. C., Blackmore, E., Sutandar-Pinnock, K., & Woodside, D. (2004). Relapse in anorexia nervosa: A survival analysis. *Psychological Medicine*, 34, 671–679. <https://doi.org/10.1017/S0033291703001168>
- Cash, T. F. (2012). Cognitive-behavioral perspectives on body image. In Cash, T. F. (Eds.) *Encyclopaedia of body image and human appearance* (pp. 334–342). London: Academic Press.
- Cash, T. F. (1995). *What do you see when you look in the mirror? Helping yourself to a positive body image*. New York: Bantam Books.
- Cash, T. F., & Deagle, E. A. (1997). The nature and extent of body-image disturbances in anorexia nervosa and bulimia nervosa: A meta-analysis. *International Journal of Eating Disorders*, 22, 107–126. [https://doi.org/10.1002/\(SICI\)1098-108X\(199709\)22:2%3C107::AID-EAT1%3E3.0.CO;2-J](https://doi.org/10.1002/(SICI)1098-108X(199709)22:2%3C107::AID-EAT1%3E3.0.CO;2-J)

- Casper, R. C., Halmi, K. A., Goldberg, S. C., Eckert, E. D., & Davis, J. M. (1979). Disturbances in body image estimation as related to other characteristics and outcome in anorexia nervosa. *The British Journal of Psychiatry*, 134, 60–66. <https://doi.org/10.1192/bjp.134.1.60>
- Castro, J., Gila, A., Puig, J., Rodriguez, S., & Toro, J. (2004). Predictors of rehospitalization after total weight recovery in adolescents with anorexia nervosa. *International Journal of Eating Disorders*, 36, 22–30. <https://doi.org/10.1002/eat.20009>
- Cesa, G. L., Manzoni, G. M., Bachetta, M., Castelnovo, G., Conti, S., Gaggioli, A., Mantovani, F., Molinari, E., Cárdenas-Lopez, G., & Riva, G. (2013). Virtual reality for enhancing the cognitive behavioural treatment of obesity with binge eating disorder: randomized controlled study with one-year follow-up. *Journal of Medical Internet Research*, 15, e113. <https://doi.org/10.2196/jmir.2441>
- Channon, S., & De Silva, W. (1985). Psychological correlates of weight gain in patients with anorexia nervosa. *Journal of Psychiatric Research*, 19, 267–271. [https://doi.org/10.1016/0022-3956\(85\)90027-5](https://doi.org/10.1016/0022-3956(85)90027-5)
- Clus, A., Larsen, M. E., Lemey, C., & Berrouguet, S. (2018). The use of virtual reality in patients with eating disorders: systematic review. *Journal of medical Internet Research*, 20, e157. <https://doi.org/10.2196/jmir.7898>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. <https://doi.org/10.1037//0033-2909.112.1.155>
- Cornelissen, K. (2016). *What does it mean to have distorted body image in anorexia nervosa?* (Doctoral thesis). Retrieved from <http://nrl.northumbria.ac.uk/30330/>
- Cornelissen, K. K., Bester, A., Cairns, P., Tovée, M. J., & Cornelissen, P. L. (2015). The influence of personal BMI on body size estimations and sensitivity to body size change in anorexia spectrum disorders. *Body Image*, 13, 75–85. <https://doi.org/10.1016/j.bodyim.2015.01.001>
- Cornelissen, K. K., McCarty, K., Cornelissen, P. L., & Tovée, M. J. (2017). Body size estimation in women with anorexia nervosa and healthy controls using 3D avatars. *Scientific Reports*, 7, 15773. <https://doi.org/10.1038/s41598-017-15339-z>

- Cornelissen, K. K., Widdrington, H., McCarty, K., Pollet, T. V., Tovée, M. J., & Cornelissen, P. L. (2019). Are attitudinal and perceptual body image the same or different? Evidence from high-level adaptation. *Body Image, 31*, 35-47. <https://doi.org/10.1016/j.bodyim.2019.08.001>
- Cundall, A., & Guo, K. (2015). Women gaze behaviour in assessing female bodies: The effects of clothing, body size, own body composition and body satisfaction. *Psychological Research, 81*, 1–12. <https://doi.org/10.1007/s00426-015-0726-1>
- De Carvalho, M., Dias, T., Duchesne, M., Nardi, A., & Appolinario, J. (2017). Virtual reality as a promising strategy in the assessment and treatment of Bulimia Nervosa and Binge Eating Disorder: A Systematic Review. *Behavioral Sciences, 7*, 43. <https://doi.org/10.3390/bs7030043>
- Elder, J. H. (2018). Shape from contour: Computation and representation. *Annual Review of Vision Science, 4*, 423–450. <https://doi.org/10.1146/annurev-vision-091517-034110>
- Espeset, E., Gulliksen, K., Nordbø, R., Skårderud, F., & Holte, A. (2012). Fluctuations of body images in anorexia nervosa: Patients’ perception of contextual triggers. *Clinical Psychology & Psychotherapy, 19*, 518–530. <https://doi.org/10.1002/cpp.760>
- Evans, C., & Dolan, B. (1993). Body Shape Questionnaire: derivation of shortened “alternate forms.” *International Journal of Eating Disorders, 13*, 315-321. [https://doi.org/10.1002/1098-108x\(199304\)13:3<315::aid-eat2260130310>3.0.co;2-3](https://doi.org/10.1002/1098-108x(199304)13:3<315::aid-eat2260130310>3.0.co;2-3)
- Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorder psychopathology: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363-370.
- Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A “transdiagnostic” theory and treatment. *Behaviour Research and Therapy, 41*, 509–528. [https://doi.org/10.1016/s0005-7967\(02\)00088-8](https://doi.org/10.1016/s0005-7967(02)00088-8)
- Fairburn, C. G., Cooper, Z., Shafran, R., & Wilson, G. T. (2008). *Eating disorders: A transdiagnostic protocol. Clinical handbook of psychological disorders: A step-by-step treatment manual* (4th Ed.). New York: Guilford Press.

- Fairburn, C. G., & Harrison, P. J. (2003). Eating disorders. *The Lancet*, 361, 407-416.
[https://doi.org/10.1016/s0140-6736\(03\)12378-1](https://doi.org/10.1016/s0140-6736(03)12378-1)
- Felinhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A.-K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human-Computer Studies*, 82, 48-56.
<https://doi.org/10.1016/j.ijhcs.2015.05.004>
- Farrell, C., Shafran, R., & Lee, M. (2006). Empirically evaluated treatments of body image disturbance: A review. *European Eating Disorders Review*, 14, 289-300. <https://doi.org/10.1002/erv.693>
- Ferrer-Garcia, M., Gutiérrez-Maldonado, J., & Riva, G. (2013). Virtual reality based treatments in eating disorders and obesity: A review. *Journal of Contemporary Psychotherapy*, 43, 207-221.
<https://doi.org/10.1007/s10879-013-9240-1>
- Ferrer-García, M., Gutiérrez-Maldonado, J., Caqueo-Urizar, A., & Moreno, E. (2009). The validity of virtual environments for eliciting emotional responses in patients with eating disorders and in controls. *Behavior Modification*, 33, 830-854. <https://doi.org/10.1177/0145445509348056>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
<https://doi.org/10.1177/001872675400700202>
- Fonseca-Baeza, S., Corno, G., & Baños, R. M. (2018). An intervention protocol proposal to modify the body image disturbance using virtual reality. *Calidad de vida y salud*, 11, 48-61.
- Fraser, J. F., Allen, B., Anand, V. K., & Schwartz, T. H. (2009). Three-dimensional neurostereoscopy: subjective and objective comparison to 2D. *Minimally Invasive Neurosurgery*, 52, 25-31.
<https://doi.org/10.1055/s-0028-1104567>
- Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 47, 2393-2400. <https://doi.org/10.1017/s003329171700040x>

- Freeman, R. J., Thomas, C. D., Solyom, L., & Koopman, R. F. (1985). Clinical and personality correlates of body size overestimation in anorexia nervosa and bulimia nervosa. *International Journal of Eating Disorders*, 4, 439–456. [https://doi.org/10.1002/1098-108x\(198511\)4:4<439::aid-eat2260040405>3.0.co;2-b](https://doi.org/10.1002/1098-108x(198511)4:4<439::aid-eat2260040405>3.0.co;2-b)
- Gardner, R. M. (2011). Perceptual measures of body image for adolescents and adults. In Cash, T. F & Smolak, L. (Eds.). *Body image. A handbook of science, practice and prevention* (2nd ed.). New York: The Guilford Press.
- Gardner, R. M., & Brown, D. L. (2014). Body size estimation in anorexia nervosa: a brief review of findings from 2003 through 2013. *Psychiatry Research*, 219, 407-410. <https://doi.org/10.1016/j.psychres.2014.06.029>
- Garner, D. M., & Garfinkel, P. E. (1997). *Handbook of treatment for eating disorders*. New York: Guilford Press.
- Gervais, S. J., Holland, A. M., & Dodd, M. D. (2013). My eyes are up here: The nature of the objectifying gaze toward women. *Sex Roles*, 69, 557–570. <https://doi.org/10.1007/s11199-013-0316-x>
- Gledhill, L. J., Cornelissen, K. K., Cornelissen, P. L., Penton-Voak, I. S., Munafò, M. R., & Tovée, M. J. (2017). An interactive training programme to treat body image disturbance. *British Journal of Health Psychology*, 22, 60–76. <https://doi.org/10.1111/bjhp.12217>
- Gledhill, L. J., George, H. R., & Tovée, M.J. (2019). Perceptual not attitudinal factors predict the accuracy of estimating other women's bodies in both women with anorexia nervosa and controls. *Frontiers in Psychology*, 10, 997. <https://doi.org/10.3389/fpsyg.2019.00997>
- Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: The virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking*, 14, 99–105. <https://doi.org/10.1089/cyber.2010.0100>
- Grogan, S. (2016). *Body image: Understanding body dissatisfaction in men, women and children*. London: Routledge.
- Gutiérrez-Maldonado, J., Wiederhold, B. K., & Riva, G. (2016). Future directions: How virtual reality can further improve the assessment and treatment of eating disorders and obesity. *Cyberpsychology, Behavior, and Social Networking*, 19, 148–153. <https://doi.org/10.1089/cyber.2015.0412>

- Heilbrun, A. B., & Friedberg, L. (1990). Distorted body image in normal college women: Possible implications for the development of anorexia nervosa. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 389-401. [https://doi.org/10.1002/1097-4679\(199007\)46:4<398::aid-jclp2270460404>3.0.co;2-5](https://doi.org/10.1002/1097-4679(199007)46:4<398::aid-jclp2270460404>3.0.co;2-5)
- Herzog, D. B., Dorer, D. J., Keel, P. K., Selwyn, S. E., Ekeblad, E. R., Flores, A. T., Greenwood, D. N., Burwell, R. A., & Keller, M. D. (1999). Recovery and relapse in anorexia nervosa and bulimia nervosa: A 7.5-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 829-837. <https://doi.org/10.1097/00004583-199907000-00012>
- Irvine, K. R., McCarty, K., McKenzie, K., Pollet, T.V., Cornelissen, K. K., Tovée, M. J., & Cornelissen, P. L. (2019). Distorted body image influences body schema in individuals with negative bodily attitudes. *Neuropsychologia*, 122, 38-50. <https://doi.org/10.1016/j.neuropsychologia.2018.11.015>
- Jarry, J. L., & Cash, T. F. (2012). Cognitive-behavioral approaches to body image change. In Cash, T. F. & Smolak, L. (Eds.). *Body image: A handbook of science, practice and prevention* (pp. 415-424). New York: Guilford.
- Junne, J., Wild, B., Resmark, G., Giel, K.E., Teufel, M., Martus, P., Ziser, K., Friederich, H.C., de Zwaan, M., Löwe, B., Dinkel, A., Herpertz, S., Burgmer, M., Tagay, S., Rothermund, E., Almut Zeeck, A., Herzog, W., & Zipfel, S. (2019). The importance of body image disturbances for the outcome of outpatient psychotherapy in patients with anorexia nervosa: Results of the ANTOP-study. *European Eating Disorders Review*, 27, 49–58. <https://doi.org/10.1002/erv.2623>
- Kazis, L.E., Anderson, J.J., & Meenan, R.F. (1989). Effect sizes for interpreting changes in health status. *Medical Care*, 27, S178-S189.
- Keel, P. K., Dorer, D. J., Franko, D. L., Jackson, S. C., & Herzog, D. B. (2005). Postremission predictors of relapse in women with eating disorders. *The American Journal of Psychiatry*, 162, 2263-2268. <https://doi.org/10.1176/appi.ajp.162.12.2263>
- Koskina, A., Campbell, I. C., & Schmidt, U. (2013). Exposure therapy in eating disorders revisited. *Neuroscience & Biobehavioral Reviews*, 37, 193–208. <https://doi.org/10.1016/j.neubiorev.2012.11.010>

- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1-12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Leahey, T. M., Crowther, J. H., & Mickelson, K. D. (2007). The frequency, nature, and effects of naturally occurring appearance-focused social comparisons. *Behaviour Therapy*, 38, 132-143. <https://doi.org/10.1016/j.beth.2006.06.004>
- Levin, M. F., Snir, O., Liebermann, D. G., Weingarden, H., & Weiss, P.L. (2013). Virtual reality versus conventional treatment of reaching ability in chronic stroke: Clinical feasibility study. *Neurology and Therapy*, 1, 3. <https://doi.org/10.1007/s40120-012-0003-9>
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358-368.
- Liechty, J. M. (2010). Body image distortion and three types of weight loss behaviours among nonoverweight girls in the United States. *Journal of Adolescent Health*, 47, 176-182. <https://doi.org/10.1016/j.adohealth.2010.01.004>
- Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-scale Assessments in Education*, 6, 8. <https://doi.org/10.1186/s40536-018-0061-2>
- Manzoni, G. M., Cesa, G. L., Bacchetta, M., Castelnuovo, G., Conti, S., Gaggioli, A., Mantovani, F., Molinari, E., Cárdenas-Lopez, G., & Riva, G. (2016). Virtual reality-enhanced cognitive-behavioral therapy for morbid obesity: A randomized controlled study with 1 year follow-up. *Cyberpsychology, Behavior, and Social Networking*, 19, 134-140. <https://doi.org/10.1089/cyber.2015.0208>
- Maples-Keller, J. L., Yasinski, C., Manjin, N., & Rothbaum, B. O. (2017). Virtual reality-enhanced extinction of phobias and post-traumatic stress. *Neurotherapeutics*, 14, 554-563. <https://doi.org/10.1007/s13311-017-0534-y>
- Millais, P., Jones, S.L., & Kelly, R.M. (2018). Exploring data in virtual reality: Comparisons with 2D data visualizations. CHI EA '18: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. Paper No. LBW007, 1-6. <https://doi.org/10.1145/3170427.3188537>

- Mölbart, S., Thaler, A., Mohler, B., Streuber, S., Romero, J., Black, M., Zipfel, S., Karnath, H.-O., & Giel, K. E. (2018). Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: attitudinal components rather than visual body size estimation are distorted. *Psychological Medicine*, 48, 642–653. <https://doi.org/10.1017/S0033291717002008>
- Mond, J.M., Hay, P.J., Rodgers, B., & Owen, C. (2006). Eating Disorder Examination Questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy*, 44, 53–62. <https://doi.org/10.1016/j.brat.2004.12.003>
- Murray, S. B., Quintana, D. S., Loeb, K. L., Griffiths, S., & Le Grange, D. (2019). Treatment outcomes for anorexia nervosa: A systematic review and meta-analysis of randomized controlled trials. *Psychological Medicine*, 49, 535–544. <https://doi.org/10.1017/s0033291718002088>
- Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurrin, M., McDonald, S., & Munaf, M. R. (2013). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behaviour. *Psychological Science*, 24, 688–697. <https://doi.org/10.1177/0956797612459657>
- Perpiñá, C., Botella, C., & Baños, R. M. (2003). Virtual reality in eating disorders. *European Eating Disorders Review*, 11, 261–278. <https://doi.org/10.1002/erv.520>
- Perpiñá, C., Botella, C., Banos, R., Marco, H., Alcaniz, M., & Quero, S. (1999). Body image and virtual reality in rating disorders: Is exposure to virtual reality more effective than the classical body image treatment? *CyberPsychology & Behavior*, 2, 149–155. <https://doi.org/10.1089/cpb.1999.2.149>
- Polivy, J., & Herman, C. P. (2002). Causes of eating disorders. *Annual Review of Psychology*, 53, 187–213. <https://doi.org/10.1146/annurev.psych.53.100901.135103>
- Riva, G. (1998). Virtual environment for body image modification: Virtual reality system for the treatment of body image disturbances. *Computers in Human Behaviour*, 14, 477–490. [https://doi.org/10.1016/s0747-5632\(98\)00018-1](https://doi.org/10.1016/s0747-5632(98)00018-1)
- Riva, G., Bacchetta, M., Baruffi, M., & Molinari, E. (2001). Virtual reality- based multidimensional therapy for the treatment of body image disturbances in obesity: A controlled study. *CyberPsychology & Behavior*, 4, 511–526. <https://doi.org/10.1089/109493101750527079>

- Riva, G., & Waterworth, J. A. (2003). Presence and the self: A cognitive neuroscience approach. Retrieved from: <http://www8.informatik.umu.se/~jwworth/Riva-Waterworth.htm>
- Riva, G., Gaudio, S., Serino, S., Dakanalis, A., Ferrer-García, M., & Gutiérrez-Maldonado J. (2018). *Virtual reality for the treatment of body image disturbances in eating and weight disorders*. In: Cuzzolaro M., Fassino S. (eds) *Body image, eating, and weight*. Springer, Cham
- Rosen, J. C. (1997). Cognitive-behavioral body image therapy. In D. M. Garner & P. E. Garfinkel (Eds.). *Handbook of treatment for eating disorders* (2nd ed., pp. 188-205). New York: The Guilford Press.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Sala, L., Mirabel-Sarron, C., Pham-Scottez, A., Blanchet, A., Rouillon, F., & Gorwood, P. (2012). Body dissatisfaction is impaired but the ideal silhouette is unchanged during weight recovery in anorexia nervosa female patients. *Eating and Weight Disorders*, 17, e109-e115. <https://doi.org/10.1007/bf03325334>
- Scott, S., & Evans, S. (2010). A Nature Neuroscience commentary on Chang et al., (2010) Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13, 1428-1432.
- Schaefer, L. M., & Thompson, J. K. (2014). The development and validation of the Physical Appearance Comparison Scale-Revised (PACS-R). *Eating Behaviours*, 15, 209-217. <https://doi.org/10.1016/j.eatbeh.2014.01.001>
- Schilder, P. (1935). *The image and appearance of the human body*. New York, NY: International Universities Press.
- Sirirassamee, T., Phoolsawat, S., & Limkhunthammo, S. (2018). Relationship between body weight perception and weight-related behaviours *Journal of International Medical Research*, 46, 3796–3808. <https://doi.org/10.1177/0300060518780138>
- Slade, P., & Russell, G. (1973). Awareness of body dimensions in anorexia nervosa: Cross-sectional and longitudinal studies. *Psychological Medicine*, 3, 188. <https://doi.org/10.1017/s0033291700048510>
- Slade, P. D. (1988). Body image in anorexia nervosa. *British Journal of Psychiatry*, 153, 20-22.

- Smolak, L., & Thompson, J. K. (2009). *Body image, eating disorders, and obesity in youth. Assessment, prevention, and treatment* (2nd ed.). Washington DC: American Psychological Association.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage Publishing.
- Stice, E. (2002). Risk and maintenance factors for eating pathology: A meta-analytic review. *Psychological Bulletin*, 128, 825-848. <https://doi.org/10.1037/0033-2909.128.5.825>
- Stice, E., & Shaw, H. E. (2002). Role of body dissatisfaction in the onset and maintenance of eating pathology: A synthesis of research findings. *Journal of Psychosomatic Research*, 53, 985-993. [https://doi.org/10.1016/s0022-3999\(02\)00488-9](https://doi.org/10.1016/s0022-3999(02)00488-9)
- Striegel-Moore, R. H., & Bulik, C. M. (2007). Risk factors for eating disorders. *American Psychologist*, 62, 181-198. <https://doi.org/10.1037/0003-066X.62.3.181>
- Strober, M., Freeman, R., & Morrell, W. (1998). The long-term course of severe anorexia nervosa in adolescents: Survival analysis of recovery, relapse, and outcome predictors over 10-15 years in a prospective study. *International Journal of Eating Disorders*, 22, 339-360. [https://doi.org/10.1002/\(sici\)1098-108x\(199712\)22:4<339::aid-eat1>3.0.co;2-n](https://doi.org/10.1002/(sici)1098-108x(199712)22:4<339::aid-eat1>3.0.co;2-n)
- Szostak, N. M. (2018). *Negative body image and cognitive biases to body size*. (Doctoral thesis). Retrieved from: <https://hydra.hull.ac.uk/resources/hull:16444>
- Taylor, M. J. (1987). *The nature and significance of body image disturbance*. (Doctoral thesis). Retrieved from: <https://www.psych.org/tools/bsq/thesis/>
- Thaler, A., Geuss, M. N., Mölbert, S. C., Giel, K.E., Streuber, S., Romero, J., Black, M. J., & Mohler, B. J. (2018). Body size estimation of self and others in females varying in BMI. *PLoS ONE*, 13, e0192152. <https://doi.org/10.1371/journal.pone.0192152>
- Thompson, J. K., Heinberg, L. J., Altabe, M., & Tantleff-Dunn, S. (1999). *Exacting beauty: theory, assessment, and treatment of body image disturbance*. Washington DC: American Psychological Association.

- Tiggemann, M. (2011). Sociocultural perspectives on human appearance and body image. In Cash, T. F. & Smolak, L. (Eds.). *Body image: a handbook of science, practice, and prevention* (2nd ed., pp. 39-47), New York: The Guildford Press.
- Tovée, M. J., Edmonds, L., & Vuong, Q. C. (2012). Categorical perception of human female physical attractiveness and health. *Evolution and Human Behavior*, 33, 85-93. <https://doi.org/10.1016/j.evolhumbehav.2011.05.008>
- Trentowska, M., Svaldi, J., & Tuschen-Caffier, B. (2019). Efficacy of body exposure as treatment component for patients with eating disorders. *Journal of Behavior Therapy and Experimental Psychiatry*, 45, 178-185. <https://doi.org/10.1016/j.jbtep.2013.09.010>
- Turnbull, O. H., Driver, J., & McCarthy, R. A. (2004). 2D but not 3D: Pictorial-depth deficits in a case of visual agnosia. *Cortex*, 40, 723-738. [https://doi.org/10.1016/s0010-9452\(08\)70167-9](https://doi.org/10.1016/s0010-9452(08)70167-9)
- van den Berg, E., Houtzager, L., de Vos, J., Daemen, I., Katsaragaki, G., Karyotaki, E., Cuijpers, P., & Dekker, J. (2019). Meta-analysis on the efficacy of psychological treatments of anorexia nervosa. *European Eating Disorder Review*, 27, 331-351. <https://doi.org/10.1002/erv.2683>
- Van Ordena, K. F., & Broyles, J. W. (2000). Visuospatial task performance as a function of two- and three-dimensional display presentation techniques. *Displays*, 21, 17-24. [https://doi.org/10.1016/s0141-9382\(00\)00024-x](https://doi.org/10.1016/s0141-9382(00)00024-x)
- Vincelli, F. (1999). From imagination to virtual reality: The future of clinical psychology. *CyberPsychology & Behavior*, 2, 241-248. <https://doi.org/10.1089/109493199316366>
- Vocks, S., Legenbauer, T., Wächter, A., Wucherer, M., & Kosfelder, J. (2007). What happens in the course of body exposure? *Journal of Psychosomatic Research*, 62, 231-239. <https://doi.org/10.1016/j.jpsychores.2006.08.007>
- Yaramothu, C. Jaswal, R. S., & Alvarez T. L. (2019). Target eccentricity and form influences disparity vergence eye movement responses: A temporal and dynamic analysis. *Journal of Eye Movement Research*, 12, 7. <https://doi.org/10.16910/jemr.12.4.7>

Ziser, K., Mölbert, S. C., Stuber, F., Giel, K. E., Zipfel, S., & Junne, F. (2018). Effectiveness of body image directed interventions in patients with anorexia nervosa: A systematic analysis. *International Journal of Eating Disorders*, 51, 1121-1127. <https://doi.org/10.1002/eat.22946>

Figure Legends

Figure 1. Middle row: a range of BMI increments to the same CGI model, increasing from left to right. Top row: conceptual illustration of the categorical boundary between bodies subjectively classified as “thin” (mauve) versus “fat” (orange) when measured at baseline. Bottom row: the same categorical boundary now shifted towards heavier bodies as a result of perceptual training.

Figure 2. Illustration of all the CGI models: top row shows models 1 - 5 (BMIs of 15.45 - 20.66), middle row shows models 6 - 10 (BMIs 21.97 – 23.27), and bottom row shows models 11 - 15 (BMIs of 28.49 – 33.70). The WHO BMI categories appear next to each model. SUn = severely underweight, Un = underweight, No = normal, Ov = overweight, and Ob = obese.

Figure 3. Top row shows example training sequence for the timed ‘original intervention’ condition: (a) the fixation cross, (b) model presentation, (c) visual noise mask, (d) categorisation response boxes, and (e) feedback displayed to the participant. Bottom row shows the sequence for the untimed ‘new intervention’ condition: (f) the fixation cross, (g) stimuli and response box presentation, and (h) feedback displayed to the participant. Note that the elevation angle of the boundary between the floor and the back wall changes depending on where the participant is looking.

Figure 4. Plot a shows the mean value of body mass index at the categorical boundary, predicted from the mixed linear model as a function of measurement day. Dark grey circles represent the control group pre-training (dashed lines) and post-training (solid lines). White circles represent the new intervention condition pre-training (dashed lines) and post-training (solid lines). Black circles represent the original intervention pre-training (dashed lines) and post-training (solid lines). Plot b shows the predicted differences between pre- and post-training categorical threshold, with 95% C.I., as a function of training day. Confidence intervals that straddle zero are not statistically significant, at $p < .05$. Dark grey squares represent the control group, white squares represent the new intervention, and black squares represent the original intervention.

Table 1. Participant characteristics and questionnaire data.

	Control		Intervention original		Intervention new		Cont v. Int orig	Cont v. Int new	Int new v. Int orig
	(n = 20)		(n = 20)		(n = 20)				
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>p</i>	<i>p</i>
Age	21.10	4.19	21.25	2.67	21.40	3.56	.96	.99	>.99
BMI	28.19	6.51	25.30	4.14	26.59	4.52	.45	.92	.90
BSQ	67.15	12.46	70.70	10.78	66.20	10.73	.90	1.00	.67
EDE-Q global	3.32	0.79	3.71	0.63	3.14	1.14	.40	.99	.26
RSE	14.35	3.80	13.25	4.58	15.45	4.80	.95	.96	.57
BDI	22.20	9.33	27.10	10.90	18.05	10.69	.55	.70	.06

Note: BSQ = Body Shape Questionnaire, EDE-Q global = Eating Disorders Examination Questionnaire global score, RSE = Rosenberg Self-Esteem Scale, BDI = Beck Depression Inventory; Cont = control, Int orig = original intervention, Int new = new intervention

Table 2. Pearson correlations across the sample between psychometric variables at three time points

		EDE-Q	BSQ	BDI
Day 1	BSQ	.55***		
	BDI	.45***	.36**	
	RSE	-.36**	-.36**	-.70***
Day 4	BSQ	.71***		
	BDI	.47***	.35**	
	RSE	-.48***	-.45***	-.73***
Day 14	BSQ	.77***		
	BDI	.57***	.49***	
	RSE	-.54***	-.54***	-.77***

Note: BSQ = Body Shape Questionnaire, EDE-Q global = Eating Disorders Examination Questionnaire global score, RSE = Rosenberg Self-Esteem Scale, BDI = Beck Depression Inventory. *** $p < .001$; ** $p < .01$; * $p < .05$

Table 3. The effects of perceptual training on PSYCH scores. Note these are displayed as standard scores. *p*-values adjusted for multiple comparisons (*adj p*) were computed using the simulate option for LSmeans in PROC MIXED (SAS v9.4).

Condition	Day 1		Day 4		Day 14		Day 1 v 4			Day 1 v 14		
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	LSmean Difference	<i>p</i> / <i>adj p</i>	Hedges' <i>g</i>	LSmean Difference	<i>p</i> / <i>adj p</i>	Hedges' <i>g</i>
Control	0.26	0.71	0.02	0.85	-0.02	0.99	0.24	.11/.46	0.30	0.28	.06/.27	0.32
Original Intervention	0.66	0.80	0.22	0.88	-0.09	1.13	0.44	.003/.02	0.52	0.75	<.0001/<.0001	0.77
New Intervention	0.0	0.94	-0.39	1.00	-0.67	1.19	0.39	.009/.05	0.40	0.67	<.0001/<.0001	0.65

Table 4. EDE-Q global and subscale scores on Day 14, along with 95% CI and community standards

	Control (<i>n</i> = 20)		Intervention Original (<i>n</i> = 20)		Intervention New (<i>n</i> = 20)		Community Standard		
	<i>M</i>	<i>95% CI</i>	<i>M</i>	<i>95% CI</i>	<i>M</i>	<i>95% CI</i>	<i>M</i>	<i>SD</i>	<i>M + 1SD</i>
EDE-Q Global	3.04	2.58-3.50	3.23	2.84-3.62	2.54*	1.94-3.14	1.59	1.32	2.91
EDE-Q Res	2.53*	1.79-3.27	2.74	2.12-3.36	2.45*	1.75-3.15	1.29	1.41	2.70
EDE-Q Eat	2.08	1.48-2.68	2.27	1.78-2.76	1.68*	1.09-2.27	0.87	1.13	2.00
EDE-Q Sc	3.94*	3.39-4.49	4.20	3.72-4.67	3.22*	2.45-3.98	2.29	1.68	3.97
EDE-Q Wc	3.61	3.09-4.13	3.70	3.23-4.17	2.83*	2.17-3.49	1.89	1.60	3.49

*Indicates score within norms as defined by Mond et al. (2006). EDE-Q = Eating Disorders Examination Questionnaire.

1)
15.45
SUn

2)
16.75
Un

3)
18.06
Un

4)
19.36
No

5)
20.66
No

6)
21.97
No

7)
23.27
No

8)
24.57
No

9)
25.88
Ov

10)
27.18
Ov

11)
28.49
Ov

12)
29.79
Ov

13)
31.09
Ob

14)
32.40
Ob

15)
33.70
Ob

Timed original intervention

1500-2500ms (jittered)

150ms

150ms

No time limit

Feedback

Untimed new intervention

1500-2500ms (jittered)

No time limit

Feedback

Kamila R. Irvine: conceptualisation, methodology, investigation, formal analysis, writing – original & draft, writing – review & editing, project administration, funding acquisition

Andrew R. Irvine: software, conceptualisation, writing – original & draft, writing – review & editing

Nadia Maalin: investigation, writing – original & draft, writing – review & editing

Kristofor McCarty: conceptualisation, supervision, writing – review & editing

Katri K. Cornelissen: methodology, conceptualisation, supervision, writing – review & editing

Martin J. Tovée: conceptualisation, writing – original & draft, writing – review & editing

Piers L. Cornelissen: conceptualisation, methodology, formal analysis, writing – original & draft, writing – review & editing

Supplementary Materials

In order to facilitate comparison with the results of Gledhill et al. (2017), we report raw means and standard deviations for the BSQ, the EDE-Q and its four subscales, the RSE and BDI, on days 1, 4, and 14, reported separately for each experimental condition. In addition we also report statistical comparisons between these means for days 1 and 4, as well as days 1 and 14. These were computed using PROC MIXED (SAS v9.4), and the p-values adjusted for multiple comparisons were calculated using the simulate command in the LSmeans option of PROC MIXED.

It is clear that both the new and original interventions produce significant, strong reductions in BSQ scores by the end of training on day 4. The original intervention also shows effects on RSE and BDI, while the new intervention shows modest effects on the shape concern sub-scale of the EDE-Q and on BDI. Strong effects emerge across the board for both original and new intervention by day 14, with the exception of the eating restraint sub-scale of the EDE-Q. Interestingly, we also observed significant reductions in the shape and weight concern sub-scales of the EDE-Q even for controls, by day 14. This may have been a statistical artefact produced by regression to the mean, although it is difficult to explain why such an effect should be restricted to only two of the four sub-scales.

Table SM1. Raw means and SD for the BSQ, the EDE-Q and its four subscales on days 1, 4, and 14, reported separately for each group.

Condition	Task	Day 1		Day 4		Day 14		Day 1 v 4 Difference	<i>p/adj p</i>	Hedges' g	Day 1 v 14 Difference	<i>p/adj p</i>	Hedges' g
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>						
Control	BSQ	67.15	12.46	63.35	12.19	63.20	14.46	3.80	.10/.44	0.31	3.95	.09/.40	0.29
	EDE-Q global	3.32	0.79	3.26	0.97	3.04	0.98	0.06	.66/.99	0.07	0.28	.06/.27	0.31
	EDE-Q res	2.84	1.20	2.83	1.32	2.53	1.59	0.01	.96/1.0	0.01	0.31	.12/.51	0.22
	EDE-Q eat	1.96	1.17	2.10	1.32	2.08	1.28	-0.14	.48/.97	-0.11	-0.12	.55/.99	-0.10
	EDE-Q sc	4.46	0.74	4.41	0.92	3.94	1.17	0.05	.78/.99	0.06	0.52	.007/.04	0.53
	EDE-Q wc	4.03	0.72	3.69	1.11	3.61	1.12	0.34	.06/.30	0.36	0.42	.02/.02	0.45
	RSE	14.35	3.80	15.20	4.18	14.40	4.73	-0.85	.19/.67	-0.21	-0.05	.94/1.0	-0.01
	BDI	22.20	9.33	19.25	8.94	18.45	11.27	2.95	.07/.32	0.32	3.75	.02/.11	0.36
Original Intervention	BSQ	70.70	10.78	65.00	10.60	60.25	14.06	5.70	.01/.08	0.53	10.45	<.0001/.0002	0.83
	EDE-Q global	3.71	0.63	3.49	0.79	3.23	0.84	0.22	.12/.52	0.31	0.48	.001/.007	0.65
	EDE-Q res	2.87	1.19	2.74	1.31	2.74	1.32	0.13	.51/.98	0.10	0.13	.51/.98	0.10
	EDE-Q eat	2.78	1.08	2.65	1.03	2.27	1.04	0.13	.51/.98	0.12	0.51	.01/.06	0.48
	EDE-Q sc	4.93	0.57	4.45	0.93	4.20	1.02	0.48	.01/0.7	0.62	0.73	.0002/.001	0.88
	EDE-Q wc	4.28	0.74	4.11	0.84	3.70	1.01	0.17	.35/.90	0.21	0.58	.002/.009	0.66
	RSE	13.25	4.58	15.00	5.11	15.70	6.17	-1.75	.007/.04	-0.36	-2.45	.0002/.001	-0.45
	BDI	27.10	10.90	22.05	10.78	19.25	12.97	5.05	.002/.01	0.47	7.85	<.0001/<.0001	0.66
New Intervention	BSQ	66.20	10.73	58.65	13.52	56.10	16.64	7.55	.001/.007	0.62	10.10	<.0001/.0002	0.72
	EDE-Q global	3.14	1.14	2.86	1.09	2.54	1.28	0.28	.06/.29	0.25	0.60	<.0001/.0006	0.50
	EDE-Q res	2.75	1.42	2.63	1.21	2.45	1.50	0.12	.55/.99	0.09	0.30	.13/.54	0.21
	EDE-Q eat	2.20	1.38	1.92	1.30	1.68	1.27	0.28	.16/.61	0.21	0.52	.009/.05	0.39
	EDE-Q sc	4.12	1.35	3.70	1.30	3.22	1.64	0.42	.03/.15	0.32	0.90	<.0001/<.0001	0.60
	EDE-Q wc	3.49	1.36	3.20	1.22	2.83	1.41	0.29	.11/.47	0.22	0.66	.0004/.0025	0.48
	RSE	15.45	4.80	16.20	4.99	17.15	5.52	-0.75	.24/.78	-0.06	-1.7	.008/.05	-0.13
	BDI	18.05	10.69	14.85	11.07	12.40	10.83	3.2	.04/.24	0.27	5.65	.0006/.002	0.49