The Self in the Mind's Eye: Revealing how we truly see ourselves through reverse correlation

Maister, Lara; De Beukelaer, Sophie; Longo, Matthew; Tsakiris, Manos

Psychological Science

Accepted/In press: 16/04/2021

Peer reviewed version

The Self in the Mind's Eye: Revealing how we truly see ourselves through reverse correlation

L. Maister[1], S. De Beukelaer[2], M. R. Longo[3] & M. Tsakiris[2,4]

[1]School of Psychology, Bangor University, Wales, UK

[2]The Warburg Institute, School of Advanced Study, University of London, UK

[3]Department of Psychological Sciences, Birkbeck University of London, UK

[4]Department of Psychology, Royal Holloway University of London, UK

Corresponding Author: Dr. Lara Maister, School of Psychology, Bangor University, Brigantia Building, Penrallt Road, Bangor, Gwynedd, UK. LL57 2AS. Tel: 01248 383657.

Email: l.maister@bangor.ac.uk

Open science: The data and analysis scripts reported in this paper can be found at https://osf.io/9jrpu/?view_only=8a9569decd9245eb833b83c5fa6ccd74**.**

Abstract

Is there a way to visually depict, for all to see, how people "see" themselves with their minds' eyes? And if so, what can these mental images tell us about ourselves? We use a computational reverse-correlation technique to explore individuals' mental 'self-portraits', of their faces and body shapes, in an unbiased, data-driven way (total $N = 116$). Self-portraits were similar to individuals' real faces, but importantly, also contained 'clues' to each person's self-reported personality traits, which were reliably detected by external observers. Furthermore, people with higher social self-esteem produced more true-to-life self-portraits. Unlike face-portraits, body-portraits had negligible relationships with individuals' actual body shape, but as with faces, they were influenced by people's beliefs and emotions. We show how psychological beliefs and attitudes about oneself bias the perceptual representation of one's appearance, and provide a unique window into the internal mental self-representation, with important implications for mental health and visual culture.

Key words: *self-representation, body, appearance, reverse correlation, personality, self-face*

Statement of Relevance

Do we really know what we look like? Given the number and sophistication of procedures for digital or physical manipulation of our appearance, and the increasing prevalence of body-image related disorders, the study of physical self-representation appears more relevant than ever. Yet, the way in which we picture ourselves "in our mind's eye" remains poorly understood. Here, we succeed in directly visualising individuals' mental 'self-portraits' of their faces and bodies in an unbiased, data-driven way. We find individual differences in their accuracy, which are linked to social self-esteem. Furthermore, we reveal how individuals 'imprint' their psychological traits on these visualisations, leading to biased and exaggerated mental self-images to match their beliefs about themselves. Our findings show the close interaction between different aspects of self-representation, and raises intriguing possibilities for understanding body-image disorders and our cultural practices of portraying the self.

How we represent and experience our self is a long-standing topic of intense interest for psychological sciences, and a recurring theme in the history of culture, demonstrating humanity's fascination with depicting selfhood. The creation of self-portraits has long been understood to be not only a representation of the actual physical appearance of the artist, but also an exploration of the artist's identity, emotions, and beliefs (Hall, 2014). This dual nature of self-representation maps onto a long-standing distinction between physical and psychological self-representations (Hu et al., 2016; Northoff et al., 2006). The physical self contains sensory information, pertaining to both the representation and perception of the body (Carruthers, 2008), and is distinct from the psychological self, which contains semantic, propositional, and affective information such as self-knowledge, beliefs, and attitudes (Hu et al., 2016).

An important, yet understudied, constituent of the physical self is the mental representation of our body's perceptual appearance (Pitron, Alsmith, & de Vignemont, 2018), including our size, shape and facial characteristics (Carruthers, 2008). These are likely to be stored and retrieved in a pictorial, depictive format (Chang, Nemrodov, Lee, & Nestor, 2017), essentially a mental picture of the self. How we picture ourselves in our mind's eye has fundamental socio-economical and clinical implications. Our perception of our own physical qualities is tightly related to our self-esteem (Feingold, 1992), and also affects a spectrum of social behaviours ranging from choice of romantic partners (Feingold, 1988), to use of appearance-modification practices such as plastic surgery (Crerand, Franklin, & Sarwer, 2006). Holding distorted self-representations can be distressing, and is linked to serious clinical disorders, such as body dysmorphia and anorexia (Kaplan, Rossell, Enticott, & Castle, 2013).

The theory that our mental representation of our physical appearance may give us clues into the more psychological aspects of the self is not a new one (e.g. see Blanke, 2007).

Although this question has not yet been directly empirically tested with regards to the self, evidence suggests that we spontaneously use the physical appearance of *others* to make physiognomic inferences regarding their psychological attributes, such as personality traits, and social group membership (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Therefore, according to external observers, the body's physical appearance does not merely reflect the physical, but also the psychological attributes of an individual. Here, we investigated if and how the representation of the *self's* physical appearance is related to the psychological self, in a similar way.

In a unique approach to this problem, we developed a novel implementation of a reverse correlation task (Mangini & Biederman, 2004), which allows us to directly visualise the rich mental representation of one's physical appearance (herein referred to as 'self-portraits'), and assess its accuracy and underlying mechanisms (cf. Moon, Kim, Kim, Kim, & Ko, 2020). Reverse correlation has already provided a revealing window into internal mental representations of others' faces (Dotsch & Todorov, 2012), body shapes (Lick, Carpinella, Preciado, Spunt, & Johnson, 2013), and most recently one's own face (Moon et al., 2020). A strength of this technique is that it provides a depictive representation of the physical self, as a direct pictorial image, which matches the native format the representation is likely to be stored in and retrieved (Kosslyn, 2005). It also enables us to measure the representation with a qualitatively different level of fidelity than previous methods have achieved – a level which preserves holistic perceptual information and may support direct identity recognition. Finally, it is primarily unconstrained and data-driven, and therefore provides an unbiased reflection of the physical self 'in the mind's eye'. This allows us to avoid a key limitation of traditional self-recognition paradigms (Epley & Whitchurch, 2008; Verosky & Todorov, 2010) in which the use of true, or only mildly distorted images of the participant's real face as stimuli may unintentionally correct participants' stored mental self-face representations during measurement to be closer to reality. This limitation is also characteristic of studies exploring traditional self-portraiture (e.g. Blanke, 2007); not only are these studies restricted to artist populations and confounded by artistic skill and style, the majority of artists create self-portraits from a physical reference, e.g. from a photograph of themselves or whilst viewing themselves in a mirror, again preventing the direct assessment of an internal stored representation.

We therefore aimed to elucidate whether and how physical self-representations of one's face (**Experiment 1**) and one's body (**Experiment 2**) interact with more psychological self-representations, such as beliefs and attitudes towards ourselves by directly measuring the accuracy of representations of our appearance, and furthermore, to qualitatively and quantitatively assess the nature of systematic distortions. By comparing these internal representations with participants' real facial and bodily characteristics , we were able to objectively measure the accuracy of their mental self-portraits. We predicted that these physical self-representations would contain accurate identity information, due to the high familiarity and frequent exposure to one's own face and body, as well as the widely-reported enhancements in visual memory for self-related stimuli (Sui & Humphreys, 2015). However, we also expected that they would contain some incorrect information reflecting biases or error, due to the reconstructive nature of visual memory (Kosslyn, 2005). Crucially, we predicted that individual patterns of error in the physical self-representation would be significantly related to psychological aspects of the self, such as beliefs about one's personality traits or attitudes.

Experiment 1

Materials and Methods

Design. In the primary phase, we obtained a self-portrait from each participant, using a reverse-correlation task. We also obtained their self-reported ratings of various psychological aspects of self-representation (their beliefs about their own personality traits, and their state self-esteem). In the secondary phase of data-collection, a new sample of independent participants were asked to rate the self-portraits and photographs of the participants' real faces on the same personality traits.

Participants. For the primary data collection, a convenience sample of 77 White Caucasian adult participants (34 males; M: 24.3 years, SD: 3.9) were recruited. Ethnicity was not specifically selected for, but due to the analysis of facial appearance in this experiment, homogenous samples were required. At the end of the recruitment phase, there was not a sufficient number of participants of any other single ethnic origin to create a full sample. This sample size, reflecting the number we successfully managed to recruit across a fixed-duration recruitment period of two months, provided high power (>99.9%, 95% CI [99.6, 100.0]) to detect an estimated medium-sized effect for the fixed effect of self-reported personality traits within the linear mixed-effects model. This test was chosen for the power analysis as it directly assesses the central hypothesis, namely that beliefs about oneself (in this case, beliefs about one's personality traits) would be related to corresponding visual features of the self-portrait. Power calculations were based on Monte Carlo simulations using the simr package in R (Green & Macleod, 2016). Participants gave written informed consent, and the experiment was approved by the ethics committee of Bangor University's School of Psychology. Participants attended a laboratory-based testing session, and first completed the reverse correlation task, then personality and self-rating measures, and finally had a passport-style photograph taken of their face. For secondary data collection phase, 112 participants (35 male; *M*: 34.8 years, *SD*: 11.0) were recruited online using the participant recruitment platform Prolific (https://www.prolific.co/).

Measures.

Reverse correlation task. For the reverse correlation task (Dotsch & Todorov, 2012), stimuli were generated using the rcicr R package (Dotsch, 2016), which randomly generates patterns of sinusoidal noise superimposed over a 'base face', resulting in a different-looking face with each random noise pattern. The base face was an average composite image, either male or female depending on the gender of the participant, obtained from an existing database (DeBruine & Jones, 2017). Five hundred random noise patterns, and their corresponding inverted patterns, were generated, creating 500 perceptually opposing pairs of facial images. Each stimulus pair was presented side-by-side to participants on a computer monitor, one pair per trial (see Figure 1, and SOM-R for details). Images resulting from each participants' performance on the reverse correlation task were generated with the rcicr package in R (Dotsch, 2016). All selected face images were averaged to produce a final image for each participant, which visualised the perceptual information used to make a 'self' judgement. The videos found here https://osf.io/9jrpu/ show the progressive creation of the self-portrait across 500 trials, for two example participants.

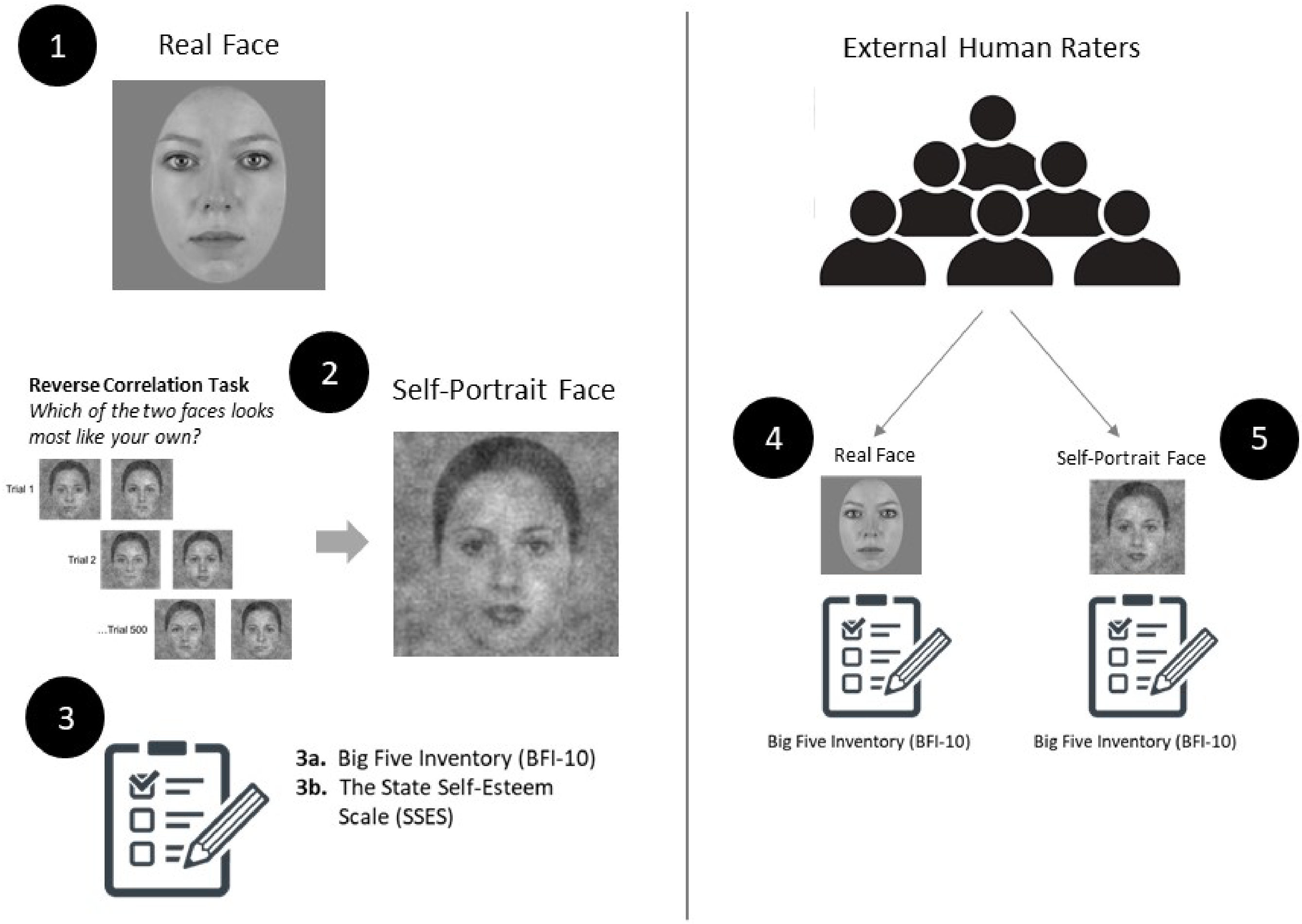

Figure 1. Experiment 1 consisted of two data-collection phases. In the primary phase, we obtained a self-portrait for each participant, using a reverse-correlation task. We also obtained their self-reported ratings of their own personality traits, and their state self-esteem. In the secondary phase of data-collection, 112 independent participants were asked to rate the self-portraits and photographs of the participants' real faces on the same personality traits. We answered four central research questions. Q1: Do self-portraits look like the participant? To test, each participant's real face (1) was compared to their self-portrait (2), using similarity scores and classification accuracy from both a face-recognition algorithm and human raters. Q2: Can external observers reliably infer personality traits from self-portraits? Inter-rater reliability scores were calculated for personality traits rated by external raters for both the self-portraits and real face photographs (4 and 5). Q3: Are self-portraits influenced by the psychological self? To test, we analysed the relationship between perceived personality features of the self-portraits (4) and self-reported personality traits (3b), whilst controlling for personality features present in the participants' real faces (5). Q4: Investigating individual differences in self-portrait accuracy. We assessed the

relationship between each participant's self-similarity score (1 vs. 2) and their self-reported personality traits and self-esteem (3a and 3b).

Questionnaires. A small battery of questionnaires was used to assess self-rated personality traits, self-esteem and facial attributes. To assess personality traits, a short 10-item form of the widely-employed Big Five Inventory (BFI10) was used (Rammstedt & John, 2007), providing a sub-score for each of the five personality traits, whereby the higher the score, the more strongly the participant believed they held that specific personality trait (in the case of the self-ratings) or the more strongly the external raters perceived that trait in a face's features (in the case of the external 'other' ratings of the real faces and self-portraits). To assess self-esteem, the 20-item State Self-Esteem Scale (SSES) was used (Heatherton & Polivy, 1991). It produces three correlated factors; performance, social, and appearance self-esteem.

Photograph. A facial photograph was taken at the end of the session. This was taken in passport-style, with a neutral facial expression, direct gaze and frontal positioning. The faces were subsequently cropped round the hairline to remove extraneous features. See SOM-R for further details of post-processing.

Secondary data collection. Ratings from a third-person perspective were obtained for both the real faces and the self-portraits obtained from the entire sample of 77 participants. Each rater saw two images from each of a subgroup of 18-20 participants (M= 19.3, SD=0.83), in order to reduce rater workload and fatigue. These images were randomly allocated, with the restriction that the same external raters rated both the self-portrait and the real face of the same primary participants. In total, each image received scores from a mean of 28.08 raters (SD=2.00). In separate presentations, raters completed the BFI10 for each image. This was presented in the same format as was used for the primary participants, but instead of items beginning with the words “I see myself as someone who…”, they saw the words “This person looks like they…”. Faces and questions were fully randomised.

Results

Do self-portraits look like the participant? Accuracy of each participant’s resulting self-portrait was assessed objectively using a face-recognition algorithm (Openface; Amos, Ludwiczuk, & Satyanarayanan, 2016), which provides a self-specific dissimilarity score between each individual’s self-portrait and a photograph of their real face (please see Supplementary Material for further details). We also performed cross-individual comparisons between each participant’s self-portrait and all the other participants’ real faces in the sample to produce non-self dissimilarity scores. The self-dissimilarity scores were significantly lower, at the group level, than cross-individual non-self dissimilarity scores; paired t-test; M_{SELF}= 1.43 (SD = 0.35), $M_{\text{NON-SELF}}$ =1.77 (SD = 0.16), 95% $CI_{\text{DIFFERENCE}}$ [-0.41, -0.26], $t(76) = -8.69$, $p<.001$, Cohen’s $d = 0.99$. This confirmed that participants’ self-portraits contained self-identifying facial information.

To assess to what extent inter-individual differences in real facial structure could explain the inter-individual differences in facial features of the portraits across our sample, we constructed two Representational Dissimilarity Matrices (RDMs), by calculating all pairwise dissimilarity scores between (i) each participant’s self-portrait with every other participant’s self-portrait; and (ii) each participant’s real face with every other participant’s real face. These were created from same-gender comparisons only (N = 2928 comparisons), to remove the potential confounding effect of same vs. different genders on dissimilarity scores. Using a linear regression analysis, the real-face RDM was shown to significantly predict the portrait RDM, $\beta = 0.06$, 95% CI [0.03, 0.09], $t(2926) = 3.63$, $p < .001$, demonstrating that the physical similarity structure of the real faces of the sample was represented in the self-portraits. Although highly significant, this effect was small, $r^2 = .004$.

This indicates that, although self-portraits contained accurate self-specific facial information, there remains substantial variance not accounted for by individuals' real facial features.

To validate, we tested whether human raters could correctly identify facial identity from the self-portraits, in an independent sample of 40 individuals who completed a two-alternative forced choice classification task (Experiment 1b, see SOM-R for further details). A one-sample t-test confirmed that the mean accuracy score across raters for each portrait was significantly higher than chance level (0.5); $M = 0.57$ ($SD = 0.16$), $t(76) = 3.93$, 95% CI [0.53, 0.61], $p < .001$, Cohen's $d = 0.45$. For comparison, classification accuracy was also derived for the Openface algorithm using a simulated experiment identical to that which the humans completed. Accuracy was numerically higher than the human accuracy scores, $M = 0.62$ ($SD = 0.31$), and again significantly higher than chance performance, $t(76) = 3.59$, 95% CI [0.56, 0.69], $p < .001$, Cohen's $d = 0.41$. A bootstrapped hypothesis test across 10,000 samples showed that the difference in accuracy between the algorithm and the human participants was not significant, estimated $p = .076$.

Can external observers reliably infer personality traits from self-portraits? On the ratings obtained from the secondary data collection phase, inter-rater reliability was calculated using average intra-class correlation coefficients (ICC) on the ratings of each personality trait, assessing consistency in ratings across each group of external raters. For each personality trait score averaged across external raters, the ICC ranged from fair to excellent (Cicchetti, 1994); for the self-portraits (averaged across personality traits), $M_{ICC} = 0.68$ ($SD = 0.11$), for the real faces $M_{ICC} = 0.76$ ($SD = 0.07$), see Table S1 for details. This confirmed that the personality scores obtained by averaging across external raters were sufficiently reliable for further analysis, and that the self-portraits contained visual information that reliably supported personality judgements. Thus, self-portraits contain self-specifying information related to individuals' real facial characteristics, but it is also clear that there remains substantial variance in self-portraits' facial features that deviated from individuals' real faces.

Are self-portraits influenced by the psychological self? To test whether one source of this variance could be associated with individuals' beliefs about their personality traits, we assessed, with a linear mixed-effects analysis (Baayen, Davidson, & Bates, 2008), whether the personality traits evident in self-portraits (as measured by the external personality ratings, Ratings $_{PORTRAIT}$) were predicted by participant's self-reported personality traits (Self $_{TRAITS}$, as measured using the Big Five Inventory (Rammstedt & John, 2007)). Critically, this analysis controlled for the external ratings of the personality traits inferred from participants' *real* faces (Ratings $_{REAL}$). This was necessary, to allow us to disentangle a true effect of self-reported personality traits on self-portrait ratings from a situation where participants were merely producing accurate, unbiased self-portraits but possessed real facial features that matched their self-reported personalities. See SOM-R for full details of this analysis and conceptual replication.

We first derived an optimal H_0 model, containing explanatory and control variables predicting external ratings of self-portraits, including external personality ratings of the real faces (AIC(H_0) =194.4). Using a systematic model comparison procedure, we demonstrated that a H_1 model that additionally included self-ratings of the five personality traits (Self $_{TRAITS}$) explained significantly more variance in Ratings $_{PORTRAIT}$ than the H_0 model, AIC(H_0)=194.4, AIC(H_1)=192.17, $\chi^2(1)=4.23$, p=.040. In this winning model, Self $_{TRAITS}$ had a positive parameter estimate of 0.03 (*SE*=0.02), $t(359.6)=2.04$, $F(1,359.6) = 4.17$, p=.042 (see Figure 2A), indicating that the higher participants rated themselves on a certain personality trait, the more facial features associated with that trait were present in their self-portrait, even when controlling for the *actual* presence of those features in participants' real faces (Table S2). A control model, in which self-ratings on the five personality traits were randomly shuffled within each participant, performed poorly, AIC = 196.4, $\chi^2 < .001$, $p >$ .999, and the parameter estimate of the randomly-shuffled Self $_{\text{TRAITS}}$ variable was non-significant, β= <-0.001, *t*(358.9)=-0.06, *p* = .95. This suggests that individual personality traits were indeed meaningfully linked with specific configurations of facial features in the self-portraits.

Finally, we investigated individual differences in overall portrait accuracy in relation to self-rated character traits, by investigating whether the accuracy of self-portraits relates to self-reported personality traits or self-esteem. An exploratory analysis was run using a hierarchical multiple linear regression on the self-dissimilarity scores, as calculated from the face-recognition algorithm. An important consideration at this point was to ensure that we were only investigating the accuracy of the *self-specific* information contained in the self-portraits. Each self-portrait contained 'generic' facial features, common to many faces, as well as self-specific content. By controlling for the similarity between each participant's self-portrait and all the other real faces in the sample, we adjusted the self-dissimilarity scores of the self-portraits to reflect accuracy of self-specific content, ensuring that the averageness of the self-portrait did not lead to biases in the self-dissimilarity scores.

Therefore, at the first step, the mean cross-individual dissimilarity scores between each participant's self-portrait and all other same-gender real faces was entered, β= 0.50, 95% *CI* [0.07, 0.93], *t*(75)= 2.30, *p*= .024, to ensure that we were analysing self-specific accuracy as our dependent variable. At the second step, individual difference variables of interest were added (the five personality self-ratings, to test whether self-beliefs regarding personality were associated with self-face representation, and the three self-esteem subscales, to assess whether more attitudinal aspects of self-concept were associated with self-representation). The winning model from the stepwise procedure included social self-esteem as a significant negative predictor of self-dissimilarity, β= -0.13, 95% *CI* [- 0.23, -0.04], *t*(74)

=2.68, p=.009, which survived Bonferroni correction for family-wise multiple comparisons. The higher the participant's self-esteem with regards to social interactions, the more accurate, i.e. 'true to life' their self-portraits were (see Figure 2B). No other predictor variables were included in the winning model.

However, this result could have been influenced by the attractiveness of participants' real faces. If participants tend to select the more attractive faces when performing the reverse-correlation task, by default those with more attractive real faces will generate self-portraits that gain a lower self-dissimilarity score than those who have less attractive real faces. Given that more attractive individuals may have a higher self-esteem, this could explain the reported relationship between self-esteem and self-portrait accuracy. In order to test this alternative explanation, two further analyses were conducted. First, a correlational analysis between social self-esteem and real-face attractiveness revealed that these two variables were not significantly correlated, $r(75)=.178$, $p = .121$. Second, when controlling for real facial attractiveness in the first step of the original hierarchical linear regression, the significance of social self-esteem as a predictor of self-portrait accuracy remained unchanged, $\beta= -0.13$, 95% CI [- 0.23, -0.03], $t(73)$ =2.55, p=.013. Therefore, it is unlikely that the existing findings can be explained by a confounding effect of real facial attractiveness.

Another alternative explanation involves the averageness of participants' real faces. For participants with highly average real facial features, the reverse-correlation task could have generated portraits that were highly similar to their real face by chance, giving artificially low self-dissimilarity scores with the self-portrait. This could lead to a potential confound, as facial averageness may be directly linked with self-rated character traits such as self-esteem. To ensure that this was not the case, the key result was retested whilst controlling for real-face averageness, as calculated by the mean cross-individual dissimilarity scores between the participants' real faces and all other same-gender real faces in the sample.

This confirmed that the relationship between social self-esteem and self-dissimilarity remained significant even when additionally controlling for real-face averageness, β= -0.14, 95% *CI* [-0.23, -0.04], $t(73)$ =2.75, p=.007. Real-face averageness was not significantly related to self-dissimilarity in this analysis, β= -0.38, 95% *CI* [-0.84,0.08], $t(74)$ =-1.63, p=.107. Furthermore, a separate analysis demonstrated that real-face averageness was not significantly related to social self-esteem; β=-0.16, 95% *CI* [-1.20, 0.89], $t(75)$ = -0.30, p = .763.

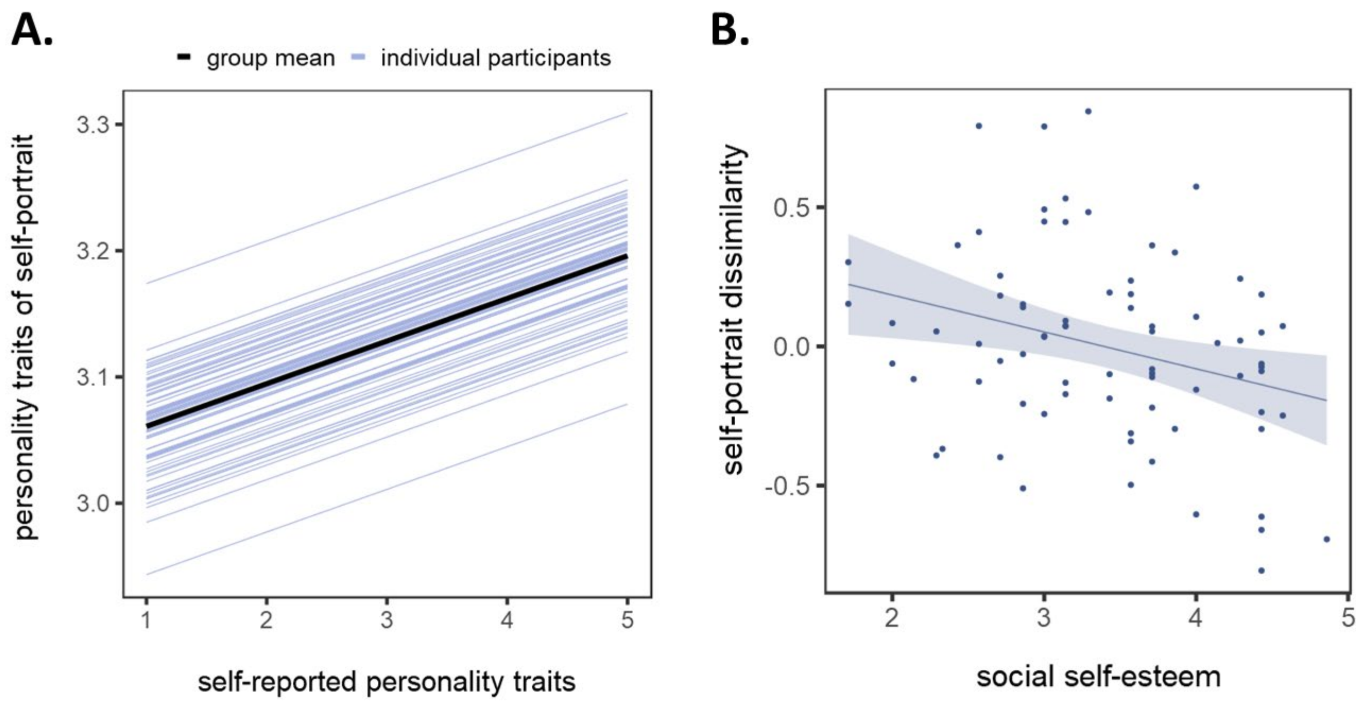

Figure 2. Key results from Experiment 1. A: Results from the linear mixed models analysis; the black line indicates the population-level fixed effect of self-reported personality traits (as rated by participants themselves) on the intensity of the corresponding personality traits perceived in the facial features of the self-portraits (as reported by external raters). The blue lines indicate the marginal effects for each individual participant (N=77), allowing for random variation of intercepts as dictated by the best-fitting linear mixed model. B: Scatter plot illustrating the relationship between individual differences in self-portrait dissimilarity (statistically controlled for the effect of non-self same-gender dissimilarity) and social self-esteem. The higher the participant's self-esteem with regards to their social interactions, the

more accurate their self-portrait, as determined by Openface face-recognition algorithms. Shaded region reflects 95% confidence interval. Individual data points represent raw data (N = 77).

Taken together, the results show that, at the group-level, self-portraits were accurate enough to support recognition. Importantly, the self-portraits also contained visual 'clues' to each person's self-reported personality traits, which were reliably detected by external observers. Finally, the higher the participants' self-esteem with regards to social interactions, the more accurate their self-portraits were.

Experiment 2

Materials and Methods

Design. We used the same reverse-correlation procedure as in Experiment 1 but replaced the face stimuli with body silhouettes (as in Lick et al., 2013), and a self-reported body self-esteem questionnaire measure, which reflects emotional attitudes towards the body and therefore provides us with an estimate of a relevant aspect of the psychological self. One further addition was made to Experiment 2; not only did we obtain a bodily 'self-portrait' from the reverse-correlation procedure, we also repeated the task in order to generate each participant's perceptual representation of a body shape that was 'typical' or 'normal' for an individual of their age and gender. This allowed us to investigate whether affective representations of the self were related solely to perceptions of one's own appearance, or whether they were related also to the way one's personal norms were perceived, and whether these effects were similar in terms of direction and magnitude (Figure 4).

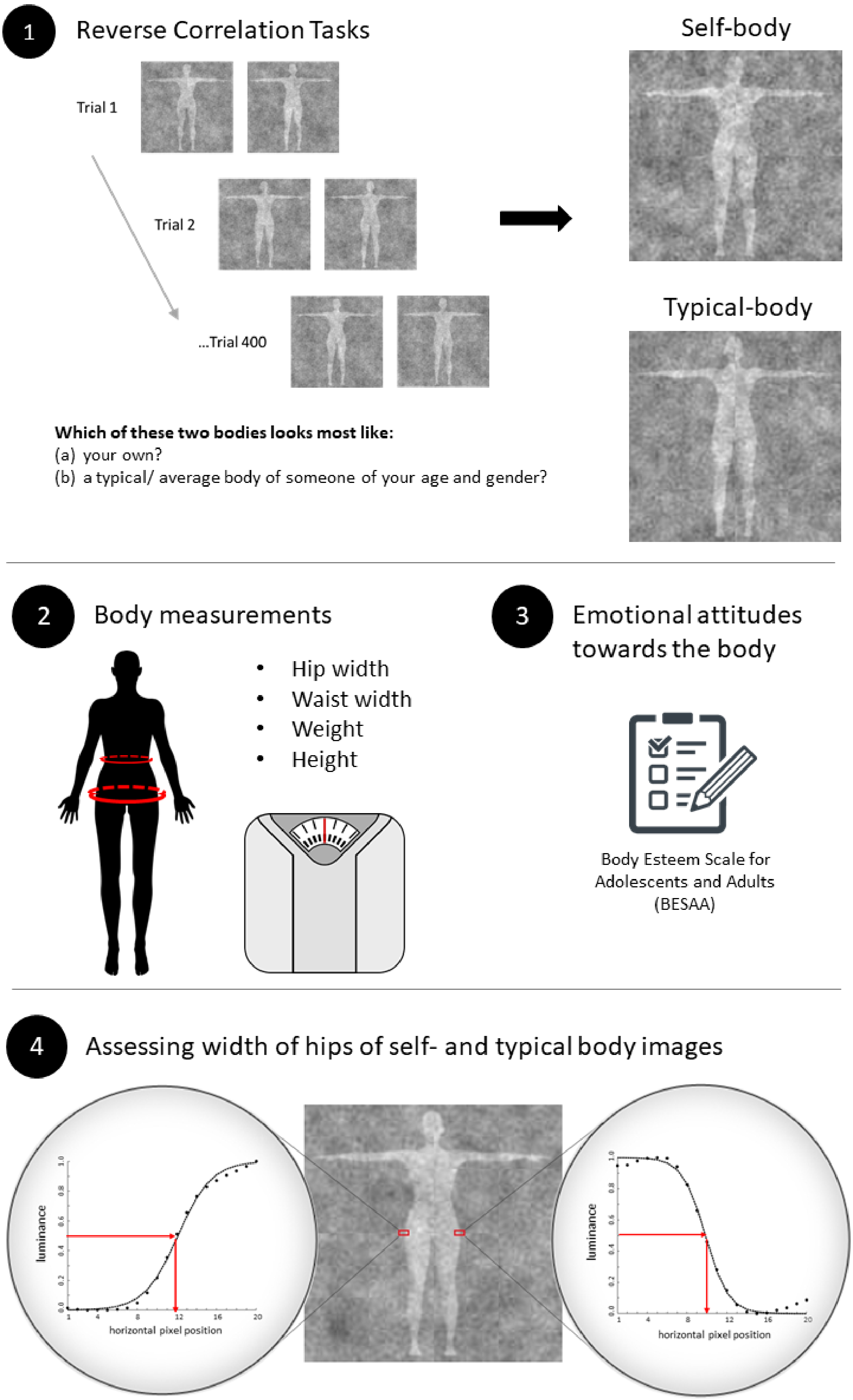
1
Reverse Correlation Tasks
Trial 1
Trial 2
...Trial 400
Self-body
Typical-body
Which of these two bodies looks most like:
(a) your own?
(b) a typical/ average body of someone of your age and gender?
2
Body measurements
Hip width
Waist width
Weight
Height
3
Emotional attitudes towards the body
Body Esteem Scale for Adolescents and Adults (BESAA)
4
Assessing width of hips of self- and typical body images
luminance
horizontal pixel position
luminance
horizontal pixel position

Figure 3. The design of Experiment 2. (1) Participants completed two reverse correlation tasks, answering with regards to either (a) their own body or (b) a typical body. (2) Several body measurements were taken, to assess the participants' real body dimensions. (3) Participants completed a 23-item questionnaire assessing their affective attitudes towards their bodies, the BESAA. (4) Illustration of the curve-fitting procedure used to estimate location of body boundaries in the classification images for self- and typical-body reverse-correlated portraits. Two hip ROIs were selected (20 x 10 pixels, indicated by red rectangles), and a logistic function was fitted to the luminance change of the pixels in each ROI. The point of subjective equality (PSE; reflecting which position on the horizontal axis whereby the average luminance of the pixels was at the mid-point of the scale) was ascertained for each curve as an estimate of edge location of each hip, indicated by the red arrows. The PSE value for the left hip was inverted, so that lower values indicated narrower hip for both left and right hips. The two PSE values were then averaged to produce an estimate of perceived hip width for each classification image. Graphs present sample data from one participant.

Participants. Forty participants were recruited, with a mean age of 23.9 years (SD = 4.1). They were from a mixture of ethnic origins. Recruitment was restricted to young (aged 18-35 years) females for this study, due to the high incidence of body image concerns in this demographic (Tiggemann & Lynch, 2001), and the differences in the stereotypical 'desirable' vs. 'undesirable' body shapes between males and females (Cohn & Adler, 1992). This sample size provided adequate power (81.4%, 95% CI [78.9, 83.8]) to detect an estimated medium-sized effect (0.35 standardised slope coefficient, Acock, 2014) for the fixed main effect of body self-esteem within the linear mixed-effects model. This test was chosen for the power analysis as it directly assesses the central hypothesis, namely that attitudes towards oneself (body self-esteem, in this case) would be related to visual features of the bodily self-portrait. Participants completed the two reverse correlation tasks, then the Body Esteem Scale for Adolescents and Adults (BESAA Mendelson, Mendelson, & White, 2001). Their body dimensions were then measured, before being debriefed and paid. One participant scored >2 standard deviations from the mean when the hip size was estimated from the reverse-correlated portrait, and was excluded from the final sample as an outlier. This left 39 participants in this experiment.

Method.

Reverse correlation task. The reverse correlation task closely followed that in Experiment 1, but with body silhouette images (see SOM-R and Figure 3 for details and examples of stimuli). Participants completed two reverse-correlation tasks (consisting of a SELF task and a TYPICAL task) using these noise-distorted body silhouettes. In the SELF task, participants were required to select the image that looked most similar to their own actual body shape. In each trial of the TYPICAL task, they were asked instead to select the image that looked most similar to the actual body shape of a "typical or average person of your age and gender". In total, participants completed 400 trials of the SELF task and 400 trials of the TYPICAL task, split across four blocks of 200 trials each in an A-B-B-A pattern which was counterbalanced across participants.

The resulting data from each task was pre-processed separately as in Experiment 1, to generate two images per participant; one reflecting their perceptual representation of their own body shape, and one reflecting their perceptual representation of what was a typical or normal body shape for someone of their age and gender.

Body Esteem Scale for Adolescents and Adults (BESAA). This 23-item questionnaire provided a measure of participants' affective attitudes towards their bodies (Mendelson et al., 2001). Each item loaded onto one of three subscales; appearance (measuring general feelings about one's appearance), weight (measuring satisfaction with one's body weight) and attribution (evaluations attributed to others about one's body and appearance), with higher scores reflecting more positive body-attitudes.

Real body measurement. Participants were weighed on a digital scale, and their height was measured. Several key body-part measurements were also taken, specifically the waist width and the hip width. As the study focussed on two-dimensional visualisation of the body, viewed from the front (as participants would see themselves in the mirror), we measured width from frontal view using callipers, rather than circumference, although it is reasonable to suppose that these two measurements are closely correlated. Body measurements were taken at the end of the testing session, after all other tasks had been completed.

Results

We first asked whether body-portraits look like the participant. As there are many body dimensions that could have been quantified, we first defined a ‘region of interest’ (ROI) around the hip area to focus our analysis (an area particularly associated with body image dissatisfaction in young women; Monteath & McCabe, 1997). A psychometric curve-fitting procedure allowed us to ascertain hip width for each participant’s reverse-correlated body-shape portraits (see Figure 3).

Simple correlations were first calculated between self-perceived hip-width from the self-portraits and the participants’ real hip measurements, which revealed no significant relationship, $r(37)= 0.05$, $p= .759$. Neither were participants’ real hip widths related to the *difference* between the self-portrait and typical portrait (self-portrait minus typical portrait hip width), $r(37) = 0.16$, $p = .341$, suggesting that unlike the facial self-portraits, the body-shape portraits had negligible direct relationships with individuals’ actual body shapes (also see SOM-R for a Bayesian analysis supporting no relationship).

We next asked whether body-portraits are influenced by attitudes towards the self. Linear mixed-effects models were employed where the dependent variable was the hip width of the self- and typical-body images generated by the reverse correlation procedure, referred to as Hip $_{PORTRAIT}$. We first derived a H_0 model (AIC_{NULL} = 249.4), containing three predictor terms; (i) participants' real hip measurements, Hip $_{REAL}$, (ii) whether they were judging their own or a typical body (Image-Type), and (iii) their interaction. Although these terms were not significant predictors of Hip $_{PORTRAIT}$, they were included to provide the strongest test for our hypothesis.

A H_1 model that including an interaction between Image-Type and Self-Esteem significantly improved model fit; AIC = 236.9, χ^2 = 16.54, p = .0003. In the most parsimonious winning model, including Self-Esteem, Image-Type, and their interaction, Self-Esteem significantly predicted Hip $_{PORTRAIT}$ positively for the typical-body, β = 0.27 (*SE* = 0.08), $t(71.0) = 3.59$, $p = .0006$, but negatively for the self-body, β = -0.14 (*SE* = 0.08), $t(71.0) = -1.91$, $p = .060$. The interaction term was strongly significant, β = 0.41 (*SE* = 0.09), $t(37.0) = 4.37$, $p < .0001$ (see Figure 4, Table S4), suggesting that participants with negative attitudes towards their own bodies produced self-portraits with larger hips, and produced "typical" portraits with slimmer hips, than participants with positive attitudes (see SOM-R for full details).

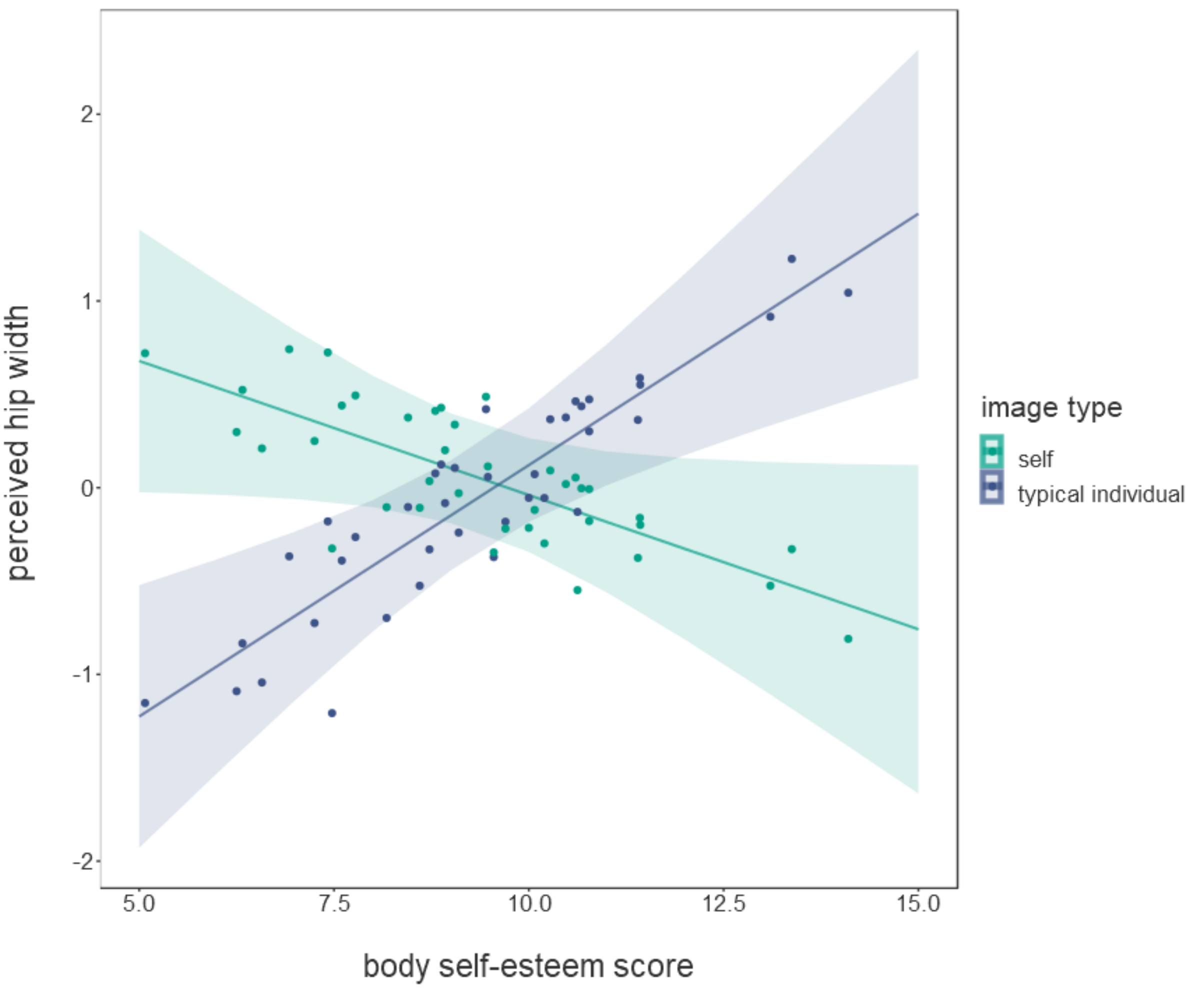

Figure 4. Results from the linear mixed models analysis of Experiment 2, showing the relationship between perceived hip width and self-esteem, for both the self and for a typical other. Perceived hip width is derived from the images resulting from the reverse correlation paradigm, giving horizontal pixel position of hip boundaries. Body self-esteem score reflects the total score achieved on the BESAA questionnaire, whereby higher scores reflect higher self-esteem. Individual points reflect predicted values from the fitted model. Shaded region represents 95% pointwise confidence intervals drawn around the estimated effect. N = 39.

Experiment 2 shows that attitudes towards one's own body, i.e. body self-esteem, did indeed shape the physical bodily self-representation. Individuals who were unhappy with their body's appearance visually represented their hips as wider, even when controlling for real body shape. In addition, when testing for the influence of body satisfaction on participants' visual representations of what 'typical' bodies looked like we found the opposite relationship; the more unhappy an individual is with their own body, the slimmer they visualise a ‘normal’ body in their mind’s eye.

Discussion

We investigated how we see ourselves in our mind’s eye, by visualizing individual participants’ representations of both their faces and their body shapes in a data-driven, unconstrained way, minimising participant biases and experimenter assumptions. This technique produced rich, holistic, and multidimensional visualisations of the face and body, which we found carried not only accurate information about physical appearance, but also provided novel insights into the way in which participants’ thoughts and feelings about themselves can ‘colour’ their self-image.

We observed clear interactions between the physical and psychological aspects of the self, whereby self-portraits of both the face and the body were significantly related to higher-level, more abstract self-beliefs and attitudes. In Experiment 1, representations of one’s facial appearance were influenced by beliefs regarding one’s personality traits; for example, if a participant believed that they were highly extraverted, they also held an internal representation of their face which had exaggerated stereotypically ‘extraverted’ facial features as compared to their true appearance. In Experiment 2, we demonstrated similar results for perceptual representations of body shape, where participants with negative attitudes towards their bodies also held visual representations of their body’s physical appearance as wider, and typical peers as slimmer, than participants with more positive attitudes.

Until now, there has been little investigation of the interaction between physical and psychological selves, with most consideration given to the bottom-up effects of multisensory and sensorimotor contingencies, on higher-level psychological self-representations (Preston

& Ehrsson, 2014). Our findings uniquely focus on self-representations stored in long-term memory, to point to a close, interactive relationship between physical and psychological representations of the self, consistent with an interactive hierarchical model of self-representation (as proposed by Sugiura, 2013). Higher-level self-beliefs and attitudes may influence the perceptual quality of the self-portraits (via a top-down modulation during the reconstruction of these images, see Kosslyn, 2005), but conversely, the perceptual features of the physical self-representation might also lead to congruent inferences about one's self-beliefs and attitudes. Indeed, evidence from studies on social perception supports a bidirectional causal relationship for our representations of others (Dotsch, Wigboldus, Langner, & Van Knippenberg, 2008; Todorov et al., 2015), and therefore a similar bidirectional relationship with regards to self-representations may also be likely.

Although the results with regards to the relationship between physical and psychological self-representations were similar for faces and bodies, there were interesting differences. Participants' representations of their *facial* appearance were clearly related to their real facial characteristics, showing a significant level of self-specificity. Classification studies, both using human participants and simulated using a face-recognition algorithm, confirmed that identity could be correctly classified from the self-portraits at well-above-chance levels. In contrast, participants' perceptual representations of their *bodies* were less related to real body characteristics (e.g. actual body size), and were more strongly influenced by affective attitudes towards the self. This is consistent with previous evidence using single-dimension measures of body parts (Ben-Tovim, Walker, Murray, & Chin, 1990), and brings into question the wide literature attempting to characterise perceptual body representations in eating disorders in terms of over- or under-estimation biases (see Mölbert et al., 2017 for review). However, it will be important to replicate our findings using larger samples of more diverse participants, increasing generalisability, as the young adult females used in

Experiment 2 may have been relatively homogenous with regards to body size as compared to the wider population.

Interestingly, individual differences in objective accuracy of the facial self-portraits were correlated with self-esteem, specifically with regards to social confidence. The higher an individual's social self-esteem, the more objectively accurate their self-portrait was. This raises interesting considerations regarding the causal role of social interaction in the development and maintenance of self-representations. Social interactions are an important source of information about our appearance, via feedback on our appearance and via social comparisons (Cash, Cash, & Butters, 1983). Therefore, individuals with higher social self-esteem may have engaged in more frequent, close social interactions, and thus received more social input about their appearance, leading to more accurate self-perception. Alternatively, individuals with more accurate perception of their appearance may also have smoother, more reciprocal and more predictable social relationships, leading to greater social confidence. For example, having an accurate perception of one's own attractiveness may lead to more successful romantic interactions, with a lower chance of being 'rebuffed' by someone poorly matched (see Le Lec et al., 2017) leading to a higher social self-esteem. Both these potential explanations appeal to a long-term relationship between self-esteem and the development of an accurate self-face representation. However, it is important to note that in our study, *state* self-esteem was assessed, rather than trait self-esteem. Although it is likely that state and trait self-esteem measures are highly correlated (see e.g. Heatherton & Polivy, 1991), future research may explore whether this finding holds for more stable aspects of self-esteem.

Our results are consistent with the findings of a very recent study, which has also used the reverse correlation technique to visualise self-face representations (Moon et al., 2020). In this study, links were found between the valence of the self-face representations generated, as rated by external observers, and various self-reported traits. Self-esteem, explicit selfevaluation and extraversion were found to be linked to more positive or pleasant-appearing self-portraits, and social anxiety was related to more negative or unpleasant-appearing self-portraits. The authors concluded that the valence of self-face representations visualised in this manner were able to reflect the attitude toward self. In the present study, in agreement with Moon et al., we also find a significant association between self-reported psychological traits and the physical features of the self-face representation. However, our results further refine our understanding of this relationship, by demonstrating that self-reported personality traits were not merely linked with the perceptual valence of self-face representations, as in Moon et al., but that individual personality traits were linked to specific facial configurations in the self-portraits that were recognisable as such by independent raters.

Our study further extends existing knowledge in several key ways. First, although Moon et al. measured participants' perceptions of self-similarity with their own self-portraits, no work has yet been done to explore the actual accuracy of self-representations, or to provide a well-controlled, unbiased assessment of their links to self-beliefs and attitudes. Here, we confirm the validity of the reverse correlation method in self-face representation research, demonstrating that the resulting images contain enough visual information to support recognition using both subjective ratings from an independent sample of raters as well as objectively using simulated experiments implementing a face-recognition algorithm. Furthermore, when exploring whether these self-face representations are influenced by higher-level self-processing, we control for real facial features, which is crucial to avoid confounds and to provide a valid, strict test of our hypothesis. Finally, we extend our investigation not just to consider face representations, but to consider body shapes, which enriches and generalises our findings to lend support to a broader mechanism whereby beliefs and attitudes influence perceptual body representations.

In this study, we used a combination of objective, algorithm-based techniques, and subjective personality ratings from human observers in order to analyse both the self-portraits and real photographs. It is possible that the human ratings of the real photographs may have been informed by superficial features of the faces, such as make-up, facial hair and grooming habits, despite the participants providing the ratings being instructed to ignore such features. However, it is important to note that the effects of this potential source of information could not explain the key results reported here. Such effects would only serve to increase the correlation found between the personality ratings of participants' real faces and their self-reported personalities. Importantly, it could not alter the relationship between the personality ratings of the *self-portraits* and the self-reported personality ratings, which is key for our hypothesis, because superficial features such as facial hair and make-up were not represented in the reverse correlation images. This issue further reiterates the importance of carefully controlling for participants' real facial ratings, which we ensured was done in each key analysis.

Both the approach we used to produce the self-portraits and our findings are highly relevant to our understanding of clinical disorders of body-image, such as anorexia nervosa and body dysmorphia. Previous studies into these disorders have normally focussed on *online* perception of the body, or have used distorted images of the patients' own bodies as stimuli which did not allow for unbiased measurement (Smeets, Ingleby, Hoek, & Panhuysen, 1999). Our approach could be used as a unique, direct method of assessing distortions in visual memory in these patients, allowing us to reveal whether they stem from higher-level self-beliefs and attitudes, or even a disorder in the link between these attitudes and the physical self-representation. This approach will also allow us to compare the effects of different treatments, e.g. those targeting perceptual distortions vs. emotional or cognitive aspects of the disorder, as well as assessing the effects of treatment across time.

In conclusion, we present a novel way to visually depict, for all to see, how people see themselves in their mind's eye, and in doing so, revealed visual clues as to people's deeply-held self-beliefs and attitudes. Our mental images of our own appearance are fundamental to our understanding of some of the most severe mental disorders that are clustered under the term of body-image disorders. In addition, at a time when our culture is powered by images at an unprecedented level, and our obsession with our own image is evidenced in our social media use (Storr, 2018), our approach and novel insights presented here pave the way for future explorations, in a data-driven, unconstrained and richly detailed way, of how we mentally see ourselves.

References

Acock, A. C. (2014). *A Gentle Introduction to Stata* (4th Ed.). College Station, TX: Stata Press.

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. CMU School of Computer Science.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Ben-Tovim, D. I., Walker, M. K., Murray, H., & Chin, G. (1990). Body size estimates: Body image or body attitude measures? *International Journal of Eating Disorders*, *9*(1), 57–67. https://doi.org/10.1002/1098-108X(199001)9:1<57::AID-EAT2260090107>3.0.CO;2-S

Blanke, O. (2007). I and Me: Self-Portraiture in Brain Damage. In J. Bogousslavsky & M. G. Hennerici (Eds.), *Neurological Disorders in Famous Artists - Part 2* (pp. 14–29). Basel: Karger.

Carruthers, G. (2008). Types of Body Representation and the Sense of Embodiment. *Consciousness and Cognition*, *17*(4), 1302–1316.

Cash, T. F., Cash, D. W., & Butters, J. (1983). “Mirror, Mirror, on the Wall...?”: Contrast Effects and Self-Evaluations of Physical Attractiveness. *Personality and Social Psychology Bulletin*, *9*(3), 351–358. https://doi.org/10.1177/0146167283093004

Chang, C. H., Nemrodov, D., Lee, A. C. H., & Nestor, A. (2017). Memory and Perception-based Facial Image Reconstruction. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-06585-2

Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohn, L. D., & Adler, N. E. (1992). Female and Male Perceptions of Ideal Body Shapes: Distorted Views Among Caucasian College Students. *Psychology of Women Quarterly*, *16*(1), 69–79. https://doi.org/10.1111/j.1471-6402.1992.tb00240.x

Crerand, C. E., Franklin, M. E., & Sarwer, D. B. (2006). Body dysmorphic disorder and cosmetic surgery. *Plastic and Reconstructive Surgery*, *118*(7), 167–180.

DeBruine, L. M., & Jones, B. C. (2017). Young Adult White Faces with Manipulated Versions. https://doi.org/https://doi.org/10.6084/m9.figshare.4220517.v1

Dotsch, R. (2016). Reverse-correlation image-classification toolbox.

Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., Langner, O., & Van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Epley, N., & Whitchurch, E. (2008). Mirror, mirror on the wall: Enhancement in self-recognition. *Personality and Social Psychology Bulletin*, *34*(9), 1159–1170. https://doi.org/10.1177/0146167208318601

Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends. *Psychological Bulletin*, *104*(2), 226–235.

Feingold, A. (1992). Good-Looking People Are Not What We Think: Conceptualization and

Measurement of Attractiveness. *Psychological Bulletin*, *Ill*(2), 304–341.

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Hall, J. (2014). *The Self-Portrait: a Cultural History*. London: Thames & Hudson.

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

Hu, C., Di, X., Eickhoff, S. B., Zhang, M., Peng, K., Guo, H., & Sui, J. (2016). Distinct and common aspects of physical and psychological self-representation in the brain: A meta-analysis of self-bias in facial and self-referential judgements. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2015.12.003

Kaplan, R. A., Rossell, S. L., Enticott, P. G., & Castle, D. J. (2013). Own-body perception in body dysmorphic disorder. *Cognitive Neuropsychiatry*, *18*(6), 594–614. https://doi.org/10.1080/13546805.2012.758878

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643290442000130

Le Lec, F., Alexopoulos, T., Boulu-Reshef, B., Fayant, M.-P., Zenasni, F., Lubart, T., & Jacquemet, N. (2017). The Out-of-my-league effect. *Behavioral and Brain Sciences*, *40*, 32. https://doi.org/10.1017/S0140525X16000534ï

Lick, D. J., Carpinella, C. M., Preciado, M. A., Spunt, R. P., & Johnson, K. L. (2013). Reverse-correlating mental representations of sex-typed bodies: The effect of number of trials on image quality. *Frontiers in Psychology*, *4*(JUL), 476. https://doi.org/10.3389/fpsyg.2013.00476

Mangini, M. C., & Biederman, I. (2004). Making the Ineffable Explicit : Estimating Representations for Face Classifications . Making the Ineffable Explicit : Estimating Representations for Face Classifications . *Cognitive Science*, *6102*(213), 209–226. https://doi.org/10.1207/s15516709cog2802_4

Mendelson, B. K., Mendelson, M. J., & White, D. R. (2001). Body-Esteem Scale for Adolescents and Adults. *Journal of Personality Assessment*, *76*(1), 90–106. https://doi.org/10.1207/S15327752JPA7601_6

Mölbert, S. C., Klein, L., Thaler, A., Mohler, B. J., Brozzo, C., Martus, P., … Giel, K. E. (2017, November 1). Depictive and metric body size estimation in anorexia nervosa and bulimia nervosa: A systematic review and meta-analysis. *Clinical Psychology Review*. Elsevier Inc. https://doi.org/10.1016/j.cpr.2017.08.005

Monteath, S. A., & McCabe, M. P. (1997). The influence of societal factors on female body image. *Journal of Social Psychology*, *137*(6), 708–727. https://doi.org/10.1080/00224549709595493

Moon, K., Kim, S. J., Kim, J., Kim, H., & Ko, Y. G. (2020). The Mirror of Mind: Visualizing Mental Representations of Self Through Reverse Correlation. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01149

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain-A meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Pitron, V., Alsmith, A., & de Vignemont, F. (2018). How do the body schema and the body image interact? *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2018.08.007

Preston, C., & Ehrsson, H. H. (2014). Illusory changes in body size modulate body satisfaction in a way that is related to non-clinical eating disorder psychopathology. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0085773

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Smeets, M. A. M., Ingleby, J. D., Hoek, H. W., & Panhuysen, G. E. M. (1999). Body size perception in anorexia nervosa: A signal detection approach. *Journal of Psychosomatic Research*, *46*(5), 465–477. https://doi.org/10.1016/S0022-3999(99)00005-7

Storr, W. (2018). *Selfie: How We Became So Self-Obsessed and What It's Doing to Us*. London: Picador.

Sugiura, M. (2013). Associative Account of Self-Cognition: Extended Forward Model and Multi-Layer Structure. *Frontiers in Human Neuroscience*, *7*, 535. https://doi.org/10.3389/fnhum.2013.00535

Sui, J., & Humphreys, G. W. (2015). The Integrative Self: How Self-Reference Integrates Perception and Memory. *Trends in Cognitive Sciences*.

Tiggemann, M., & Lynch, J. E. (2001). Body image across the life span in adult women: the role of self-objectification. *Developmental Psychology*, *37*(2), 243–253. https://doi.org/10.1037/0012-1649.37.2.243

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, *66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Verosky, S. C., & Todorov, A. (2010). Differential neural responses to faces physically similar to the self as a function of their valence. *NeuroImage*, *49*(2), 1690–1698.

https://doi.org/10.1016/j.neuroimage.2009.10.017