The Self in the Mind's Eye: Revealing How We Truly See Ourselves Through Reverse Correlation

Lara Maister[1] , Sophie De Beukelaer[2], Matthew R. Longo[3], and Manos Tsakiris[2,4,5]

[1]School of Psychology, Bangor University; [2]The Warburg Institute, School of Advanced Study, University of London; [3]Department of Psychological Sciences, Birkbeck, University of London; [4]Department of Psychology, Royal Holloway, University of London; and [5]Department of Behavioural and Cognitive Sciences, Faculty of Humanities, Education and Social Sciences, University of Luxembourg

Abstract

Is there a way to visually depict the image people "see" of themselves in their minds' eyes? And if so, what can these mental images tell us about ourselves? We used a computational reverse-correlation technique to explore individuals' mental "self-portraits" of their faces and body shapes in an unbiased, data-driven way (total $N = 116$ adults). 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—findings that have important implications for mental health and visual culture.

How we represent and experience our self is a long-standing topic of intense interest for the 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 et al., 2018), including our size, shape, and facial characteristics (Carruthers, 2008). These are likely to be stored and retrieved in a pictorial, depictive format (Chang et al., 2017), essentially a mental picture of the self. How we picture ourselves in our mind's eye has fundamental socioeconomical 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 behaviors ranging from choice of romantic partners (Feingold, 1988) to use of appearance-modification

Corresponding Author:
Lara Maister, Bangor University, School of Psychology
Email: l.maister@bangor.ac.uk

practices such as plastic surgery (Crerand et al., 2006). Holding distorted self-representations can be distressing and is linked to serious clinical disorders, such as body dysmorphia and anorexia (Kaplan et al., 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 regard 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 et al., 2015). Therefore, according to external observers, the body's physical appearance reflects not merely the physical but also the psychological attributes of an individual. Here, we investigated whether 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 allowed us to visually represent the rich mental representation of one's physical appearance (referred to hereafter as a "self-portrait") and assess its accuracy and underlying mechanisms (cf. Moon et al., 2020). Reverse correlation has already provided a revealing window into internal mental representations of other people's faces (Dotsch & Todorov, 2012), other people's body shapes (Lick et al., 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 that matches the native format in which the representation is likely to be stored and retrieved (Kosslyn, 2005). It also enabled us to measure the representation with a qualitatively different level of fidelity than previous methods have achieved—a level that 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 allowed 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, but also the majority of artists create self-portraits from a physical reference, for example, from a photograph

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 succeeded in creating a viewable visual image of individuals' mental "self-portraits" of their faces and bodies in an unbiased, data-driven way. We found that individual differences in the accuracy of these portraits were linked to social self-esteem. Furthermore, we reveal how individuals imprint their psychological traits on these internal images, 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 raise intriguing possibilities for understanding body-image disorders and our cultural practices of portraying the self.

or while 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 toward 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 body 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 because of 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 because of 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

Method

Design. In the primary phase, we used a reverse-correlation task to obtain a self-portrait from each participant. 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, we asked a new sample of independent participants to rate the self-portraits and photographs of the participants' real faces on the same personality traits.

Participants. For the primary data collection, 77 White adult university students (34 male; age: $M = 24.3$ years, $SD = 3.9$) were recruited through volunteer and opportunity sampling. Ethnicity was not specifically selected for, but because of the analysis of facial appearance in this experiment, homogeneous 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 of participants we successfully managed to recruit across a fixed-duration recruitment period of 2 months, provided high power (> 99.9%, 95% confidence interval [CI] = [99.6, 100.0]) to detect an estimated medium-size effect for the fixed effect of self-reported personality traits in the linear mixed-effects model. This test was chosen for the power analysis because it directly assessed 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 (Version 1.0.5; Green & MacLeod, 2016) in the R programming environment. 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; they first completed the reverse-correlation task, then completed personality and self-rating measures, and finally had a passport-style photograph taken of their face. For the secondary data-collection phase, 112 participants (35 male; age: $M = 34.8$ 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* package (Version 0.3.4.1; Dotsch, 2016) in R; *rcicr* 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 Fig. 1; for details, see the supplementary material at https://osf.io/sh8qg/). Images resulting from each participant's performance on the reverse-correlation task were generated with the *rcicr* package. For each participant, all selected face images were averaged to produce a final image, which provided a visual representation of the perceptual information used to make a "self" judgment. The videos on this project's OSF page (https://osf.io/9jrpu/) show the progressive creation of the self-portrait across 500 trials for two example participants.

Questionnaires. A small battery of questionnaires was used to assess self-rated personality traits, self-esteem, and facial attributes. To assess personality traits, we used a short 10-item form of the widely employed Big Five Inventory (Rammstedt & John, 2007), providing a subscore for each of the five personality traits; the higher the score, the more strongly participants 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, we used the 20-item State Self-Esteem Scale (Heatherton & Polivy, 1991). It produces three correlated factors: performance self-esteem, social self-esteem, and appearance self-esteem.

Photograph. At the end of the session, we took a passport-style facial photograph of each participant with a neutral facial expression, direct gaze, and frontal positioning. The faces were subsequently cropped around the hairline to remove extraneous features. For further details of postprocessing, see https://osf.io/sh8qg/.

Secondary data collection. Ratings from a third-person perspective were obtained for each participant's real face and their self-portraits. Each rater saw two images from each of a subgroup of 18 to 20 participants ($M = 19.3$, $SD = 0.83$) 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. Each image received scores from a mean of 28.08 raters ($SD = 2.00$). In separate presentations, raters completed the 10-item Big Five Inventory for each image. This measure was presented in the same format as was used for
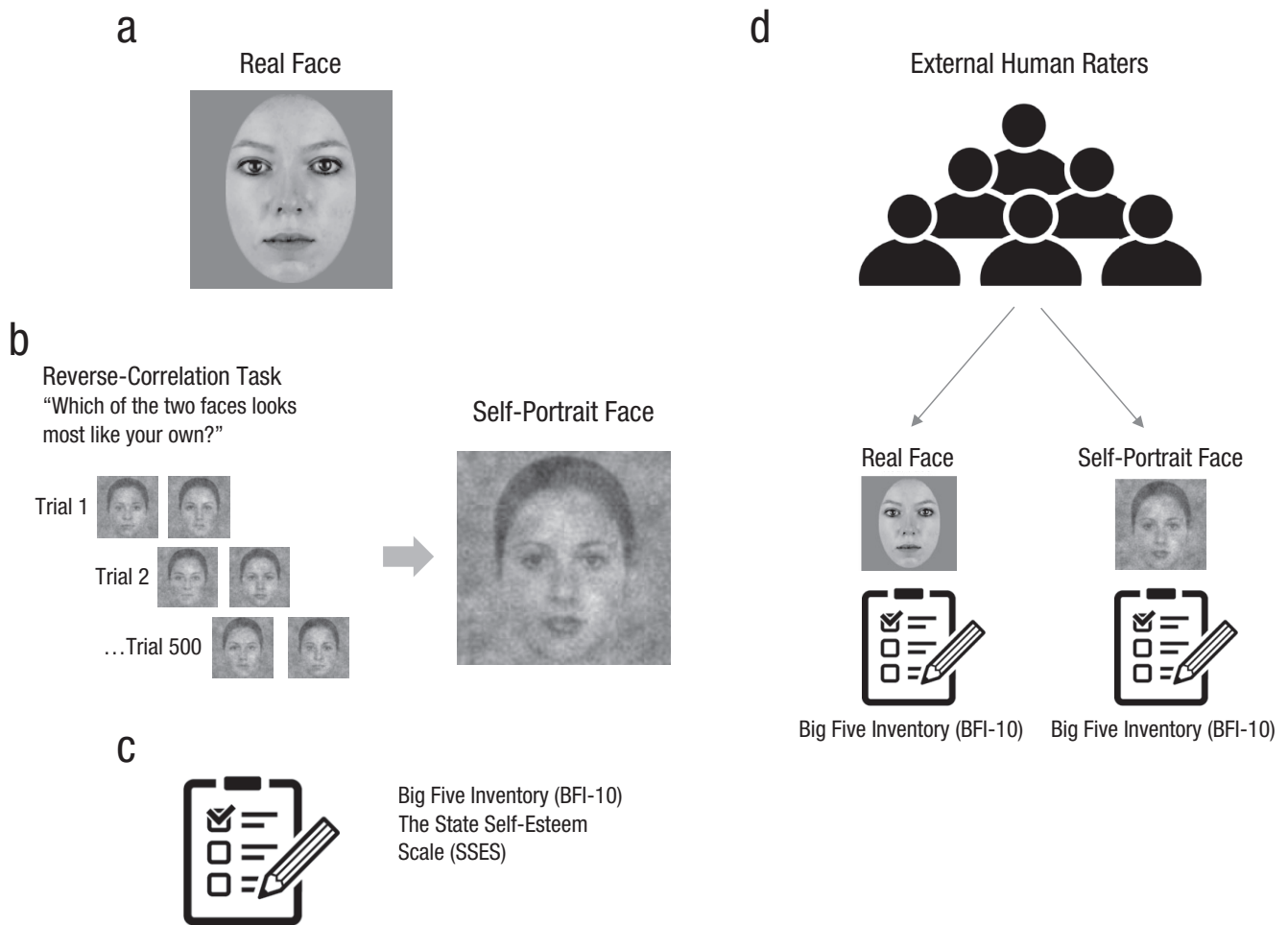
Fig. 1. Design of Experiment 1. In the primary phase of data collection, passport-style photos were taken of each participant's face (a) and then cropped around the hairline to remove extraneous features. The participant then completed a reverse-correlation task (b), in which they chose between pairs of randomly generated faces to create a "self-portrait" face that they felt looked like their own. Afterward, they filled out questionnaires (c) measuring their personality traits (BFI-10) and state self-esteem (SSES). In the second phase of data collection, 112 independent raters were shown the participant's real face and self-portrait face (d), and they used the BFI-10 to rate how strongly they perceived each personality trait in both of the faces.

the primary participants, but instead of items beginning with the words "I see myself as someone who...," raters saw the words "This person looks like they..." Faces and questions were presented in a fully randomized order.

Results

We answered four central research questions. First, does the self-portrait look like the participant? To test this question, we compared each participant's real face with their self-portrait using similarity scores and classification accuracy from both a face-recognition algorithm and human raters. Second, can external observers reliably infer personality traits from self-portraits? Interrater reliability scores were calculated for personality traits rated by external raters for both the self-portraits and real face photographs. Third, are self-portraits influenced by the psychological self? To test this, we analyzed the relationship between perceived personality features of the self-portraits and self-reported personality traits while controlling for personality features present in the participants' real faces. Fourth, which individual traits might be related to differences between participants in self-portrait accuracy? We assessed the relationship between each participant's self-similarity score and their self-reported personality traits and self-esteem.

Does the self-portrait look like the participant?

Accuracy of each participant's resulting self-portrait was assessed objectively using a face-recognition algorithm (*OpenFace*; Version 2.0; Amos et al., 2016), which provides a self-specific dissimilarity score between each individual's self-portrait and a photograph of their real

face (for further details, see https://osf.io/sh8qg/). 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 (self: $M = 1.43$, $SD = 0.35$; non-self: $M = 1.77$; $SD = 0.16$, 95% CI for the mean difference = [–0.41, –0.26]), paired-samples $t(76) = -8.69$, $p < .001$, Cohen's $d = 0.99$. This confirmed that participants' self-portraits contained self-identifying facial information.

To assess to what extent interindividual differences in real facial structure could explain the interindividual differences in facial features of the portraits across our sample, we constructed two representational-dissimilarity matrices (RDMs) by calculating all pairwise dissimilarity scores between (a) each participant's self-portrait with every other participant's self-portrait and (b) each participant's real face with every other participant's real face. These were created from same-gender comparisons only ($N = 2,928$ comparisons) to remove the potential confounding effect of same versus different genders on dissimilarity scores. Using a linear regression analysis, we found that the real-face RDM significantly predicted 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, substantial variance was not accounted for by individuals' real facial features.

To validate the findings from the face-recognition algorithm, 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 (for further details, see Experiment 1b at https://osf.io/sh8qg/). A one-sample t test confirmed that the mean accuracy score across raters for each portrait ($M = .57$, $SD = .16$, 95% CI = [.53, .61]) was significantly higher than chance level (.50), $t(76) = 3.93$, $p < .001$, 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 human participants completed. Accuracy was numerically higher than the human accuracy scores ($M = .62$, $SD = .31$, 95% CI = [.56, .69]) and again significantly higher than chance performance, $t(76) = 3.59$, $p < .001$, 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? Interrater reliability was calculated using average intraclass correlation coefficients (ICCs) on the ratings of each personality trait obtained from the secondary data-collection phase. Consistency in ratings was assessed 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): $M = .68$, $SD = .11$; for the real faces: $M = .76$, $SD = .07$ (for details, see Table S1 in the supplementary material at https://osf.io/sh8qg/). 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 judgments. Thus, self-portraits contain self-specifying information related to individuals' real facial characteristics, but it is also clear that there is 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 used a linear mixed-effects analysis (Baayen et al., 2008) to assess whether the personality traits evident in self-portraits (as measured by the external personality ratings) were predicted by participants' self-reported personality 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. This was necessary to allow us to disentangle a true effect of self-reported personality traits on self-portrait ratings from a situation in which participants were merely producing accurate, unbiased self-portraits but possessed real facial features that matched their self-reported personalities. For full details of this analysis and conceptual replication, see https://osf.io/sh8qg/.

We first derived an optimal null-hypothesis model containing explanatory and control variables predicting external ratings of self-portraits, including external personality ratings of the real faces (Akaike information criterion [AIC] = 194.4). Using a systematic model-comparison procedure, we demonstrated that an alternative-hypothesis model that additionally included self-ratings of the five personality traits explained significantly more variance in external personality ratings derived from participants' self-portraits than the null-hypothesis model did (null hypothesis: AIC = 194.4, alternative hypothesis: AIC = 192.17), $\chi^2(1) = 4.23$, $p = .040$. In this winning model, the variable indexing participants'
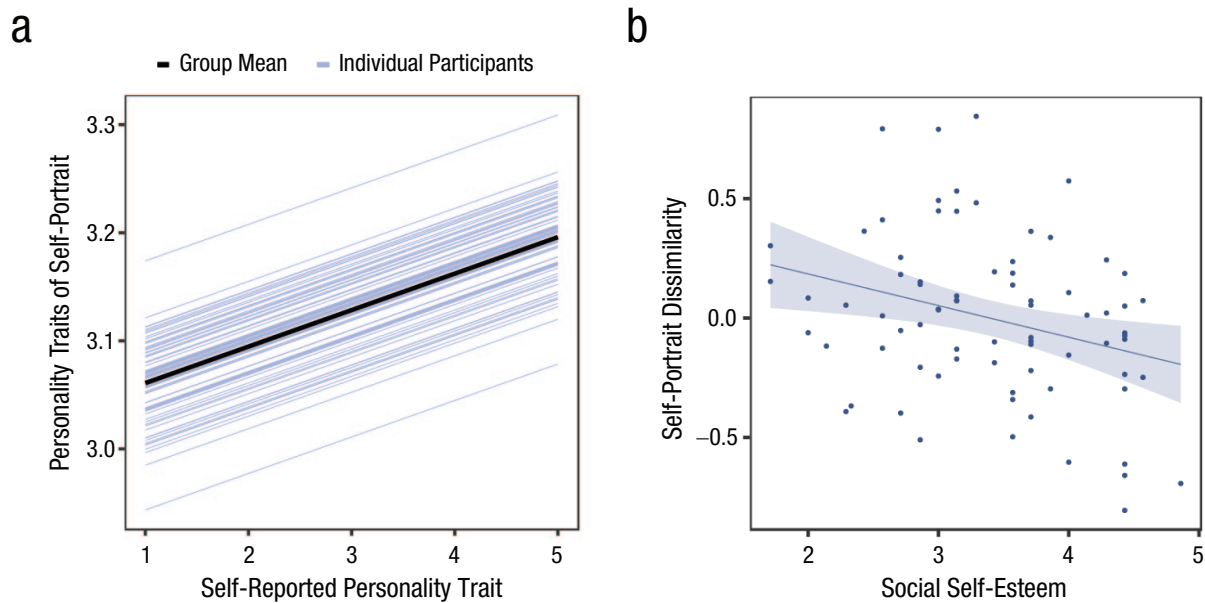
a

b

Fig. 2. Key results from Experiment 1 (*N* = 77). Results from the linear mixed-effects models analysis (a) show the 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 black line indicates the population-level mean. The blue lines indicate the marginal effects for each individual participant, allowing for random variation of intercepts as dictated by the best-fitting linear mixed-effects model. The scatterplot (b) illustrates the relationship between individual differences in self-portrait dissimilarity (statistically controlling for the effect of non-self same-gender dissimilarity) and social self-esteem. The higher the participant's self-esteem with regard to their social interactions, the more accurate their self-portrait. The solid line shows the best-fitting regression, and the shaded region reflects the 95% confidence interval.

self-reported personality traits had a positive parameter estimate, $b = 0.03$ ($SE = 0.02$), $t(359.6) = 2.04$, $F(1, 359.6) = 4.17$, $p = .042$ (see Fig. 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 the model controlled for the actual presence of those features in participants' real faces (see Table S2 in the supplementary material at https://osf.io/sh8qg/). 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 variable assessing participants' self-reported personality traits was nonsignificant, $\beta \leq -0.001$, $t(358.9) = -0.06$, $p = .95$. 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 investigating the accuracy of only 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 were entered, $\beta = 0.50$, 95% CI = [0.07, 0.93], $t(75) = 2.30$, $p = .024$, to ensure that we were analyzing 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, $\beta = -0.13$, 95% CI = [-0.23, -0.04], $t(74) = 2.68$, $p = .009$, which survived Bonferroni

correction for familywise multiple comparisons. The higher the participant's self-esteem with regard to social interactions, the more accurate (i.e., true to life) their self-portraits were (see Fig. 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 higher self-esteem, this could explain the reported relationship between self-esteem and self-portrait accuracy. To test this alternative explanation, we conducted two further analyses. 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 we controlled 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 because facial averageness may be directly linked with self-rated character traits such as self-esteem. To ensure that this was not the case, we retested the key result while 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 we additionally controlled for real-face averageness, $\beta = -0.14$, 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, $\beta = -0.38$, 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, $\beta = -0.16$, 95% CI = $[-1.20, 0.89]$, $t(75) = -0.30$, $p = .763$.

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 regard to social interactions, the more accurate their self-portraits were.

Experiment 2

Method

Design. We used the same reverse-correlation procedure as in Experiment 1 but replaced the face stimuli with body silhouettes (similar to the procedure of Lick et al., 2013) and a self-reported body self-esteem questionnaire, which reflected emotional attitudes toward the body and therefore provided 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 body self-portrait from the reverse-correlation procedure, but also we repeated the task to generate each participant's perceptual representation of a body shape that was typical for an individual of the participant's 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 as well as whether these effects were similar in terms of direction and magnitude.

Participants. Forty university students (age: $M = 23.9$ years, $SD = 4.1$, range = 18–35 years) were recruited through volunteer and opportunity sampling. They were from a mixture of ethnic origins. Recruitment was restricted to young women because of the high incidence of body-image concerns in this demographic (Tiggemann & Lynch, 2001) and the differences between the stereotypical desirable and undesirable body shapes for men and women (Cohn & Adler, 1992). This sample size provided adequate power (81.4%, 95% CI = [78.9, 83.8]) to detect an estimated medium-size effect (slope: $\beta = 0.35$; Acock, 2014) for the fixed main effect of body self-esteem in the linear mixed-effects model. This test was chosen for the power analysis because it directly assessed the central hypothesis, namely, that attitudes toward oneself (body self-esteem, in this case) would be related to visual features of the body self-portrait. Participants completed the two reverse-correlation tasks and then the Body Esteem Scale for Adolescents and Adults (Mendelson et al., 2001). Their body dimensions were then measured, and they were debriefed and paid. One participant scored greater than 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.
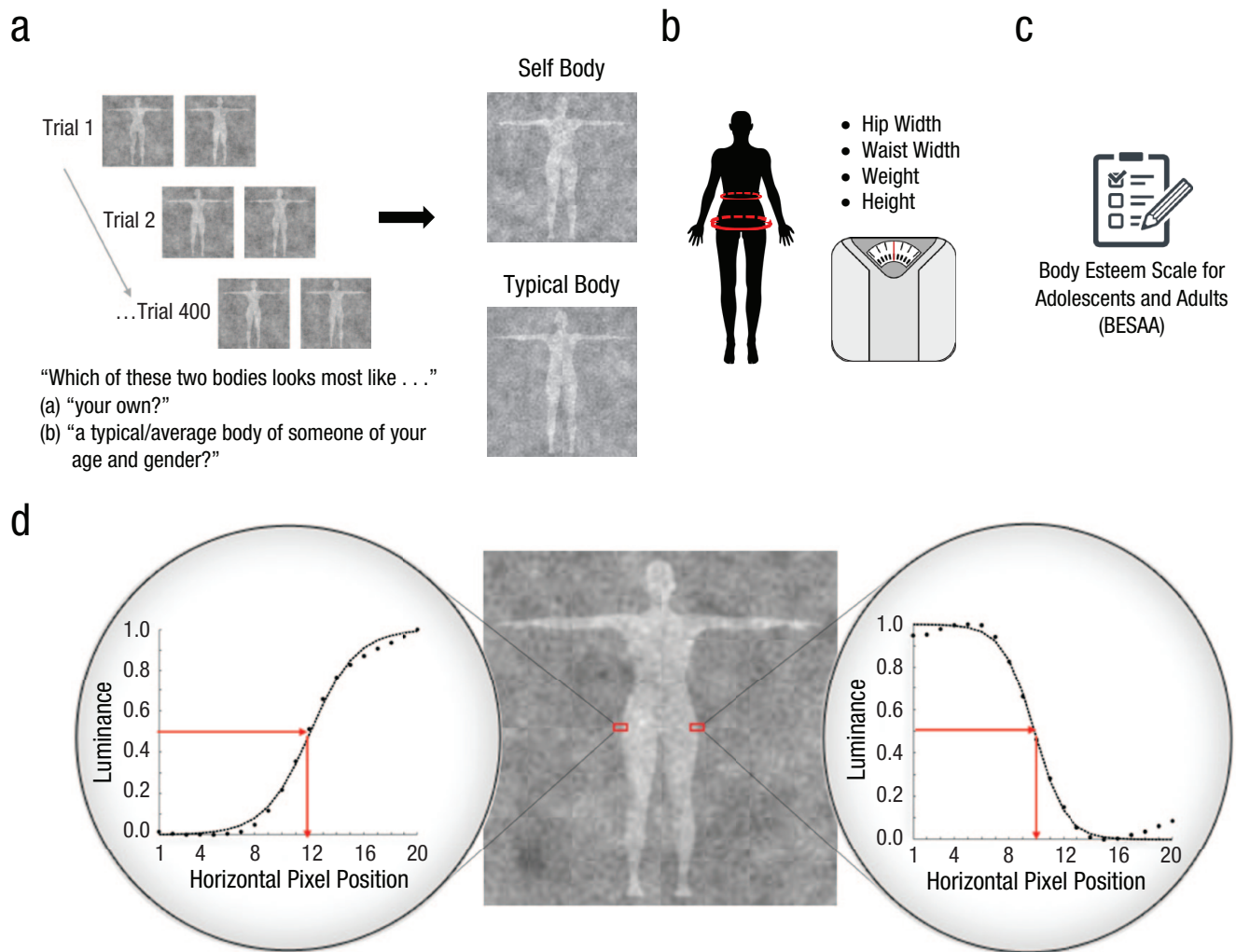
Fig. 3. Design of Experiment 2. Participants completed two reverse-correlation tasks (a), in each of which they were shown pairs of randomly generated body silhouettes and indicated whether the shape was similar to their own body or a typical body. Several body measurements (b) were taken to assess the participants' real body dimensions. Participants then completed a 23-item questionnaire (c) assessing their affective attitudes toward their bodies. The illustration (d) shows 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 regions of interest were selected (20 × 10 pixels, indicated by red rectangles), and a logistic function was fitted to the luminance change of the pixels in each region of interest. The point of subjective equality (PSE; reflecting the position on the horizontal axis where the average luminance of the pixels was at the midpoint of the scale) was ascertained for each curve as an estimate of the edge location of each hip, indicated by the red arrows. The PSE value for the left hip was inverted so that lower values indicated a 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. The graphs present sample data from one participant.

Measures.

Reverse-correlation task. The reverse-correlation task closely followed that in Experiment 1 but with body silhouette images (for details and examples of the stimuli, see https://osf.io/sh8qg/ and Fig. 3). 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 were preprocessed 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 a typical body shape for someone of their age and gender.

Body Esteem Scale for Adolescents and Adults. The 23-item Body Esteem Scale for Adolescents and Adults (Mendelson et al., 2001) questionnaire measured participants' affective attitudes toward their bodies. 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 other people about one's body and appearance). Higher scores reflect 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. Because the study focused on two-dimensional visual representations of the body viewed from the front (as participants would see themselves in the mirror), we measured width from frontal view using calipers, 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 the body portrait looks like the participant. Because there are many body dimensions that could have been quantified, we first defined a region of interest 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 Fig. 3d). The point of subjective equality (PSE; reflecting the position on the horizontal axis where the average luminance of the pixels was at the midpoint of the scale) was ascertained for each curve as an estimate of the edge location of each hip. The PSE value for the left hip was inverted so that lower values indicated a 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.

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, $r(37) = .05$, $p = .759$. Neither were participants' real hip widths related to the difference between the self-portrait and typical portrait (self-portrait – typical-portrait hip width), $r(37) = .16$, $p = .341$, suggesting that unlike the facial self-portraits, the body-shape portraits had negligible direct relationships with individuals' actual body shapes (for a Bayesian analysis supporting no relationship, see https://osf.io/sh8qg/).

We next asked whether body portraits are influenced by attitudes toward the self. Linear mixed-effects models were employed in which the dependent variable was the hip width of the self and typical body images generated by the reverse-correlation procedure. We first derived a null-hypothesis model (AIC = 249.4), containing three predictor terms: (a) participants' real hip measurements, (b) whether they were judging their own or a typical body (image type), and (c) their interaction. Although these terms were not significant predictors of hip width of the self and typical body images, they were included to provide the strongest test for our hypothesis.

An alternative-hypothesis model that included an interaction between image type and self-esteem significantly improved model fit, AIC = 236.9, $\chi^2 = 16.54$, $p = .0003$. In the most parsimonious winning model, including self-esteem, image type, and their interaction, self-esteem significantly predicted hip width of the self and typical body images, positively for the typical body ($\beta = 0.27$, $SE = 0.08$), $t(71.0) = 3.59$, $p = .0006$, but negatively for the self body ($\beta = -0.14$, $SE = 0.08$), $t(71.0) = -1.91$, $p = .060$. The interaction term was strongly significant ($\beta = 0.41$, $SE = 0.09$), $t(37.0) = 4.37$, $p < .0001$ (see Fig. 4; see also Table S4 in the supplementary material at https://osf.io/sh8qg/), suggesting that participants with negative attitudes toward their own bodies produced self-portraits with larger hips and produced typical portraits with slimmer hips, compared with participants who had positive attitudes (for full details, see https://osf.io/sh8qg/).

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

Discussion

We investigated how we see ourselves in our mind's eye by creating visual images of individual participants' mental representations of both their faces and their body shapes in a data-driven, unconstrained way, minimizing participant biases and experimenter assumptions. This technique produced rich, holistic, and multidimensional visual representations of the face and body, which we found not only carried accurate
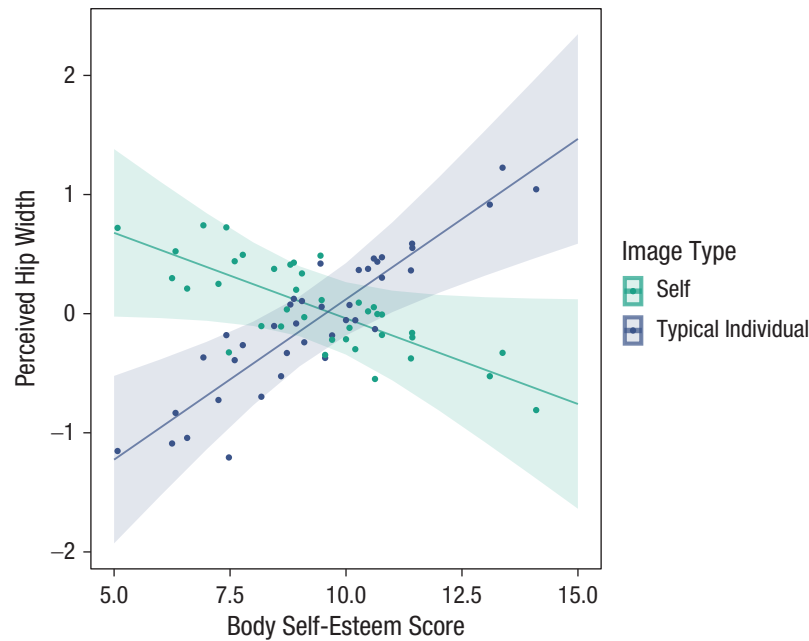
Fig. 4. Results from the linear mixed-effects models analysis of Experiment 2 (*N* = 39) showing the relationship between perceived hip width and body self-esteem, separately for the self and a typical other. Perceived hip width is derived from the images resulting from the reverse-correlation paradigm, giving the horizontal pixel position of hip boundaries. Body self-esteem score reflects the total score achieved on the Body Esteem Scale for Adolescents and Adults; higher scores reflect higher self-esteem. Individual data points reflect predicted values from the fitted model. The slopes depict the best-fitting regressions, and the shaded regions represent 95% pointwise confidence intervals drawn around the estimated effect.

information about physical appearance but also provided novel insights into the way in which participants' thoughts and feelings about themselves can color their self-image.

We observed clear interactions between the physical and psychological aspects of the self: 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 that had exaggerated stereotypically extraverted facial features compared with their true appearance. In Experiment 2, we demonstrated similar results for perceptual representations of body shape: Participants with negative attitudes toward their bodies also held visual representations of their body's physical appearance as wider and typical peers as slimmer, compared with participants who had more positive attitudes.

Until now, there has been little investigation of the interaction between physical and psychological selves, and most of the work that has been done has focused on the bottom-up effects of multisensory and sensorimotor contingencies on higher-level psychological self-representations (Preston & Ehrsson, 2014). Our work uniquely focuses 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 et al., 2008; Todorov et al., 2015); therefore, a similar bidirectional relationship with regard to self-representations may also be likely.

Although the results with regard 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 toward the self. This is consistent with previous evidence using single-dimension measures of body parts (Ben-Tovim et al., 1990) and brings into question the wide literature attempting to characterize perceptual body representations in eating disorders in terms of overestimation or underestimation biases (for a review, see Mölbert et al., 2017). However, it will be important to replicate the findings of both experiments using larger samples of more diverse participants before drawing conclusions. The generalizability of the present study may be limited. In Experiment 1, only young Caucasian adults were tested, and therefore it is necessary to follow up with studies using a wider range of ethnicities. Furthermore, in Experiment 2, only young adult women were tested, and their body size may have been relatively homogeneous compared with the general population.

Interestingly, individual differences in objective accuracy of the facial self-portraits were correlated with self-esteem, specifically with regard 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 et al., 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 and a lower chance of being rebuffed by someone poorly matched (see Le Lec et al., 2017), leading to higher social self-esteem. Both of 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, we assessed state self-esteem rather than trait self-esteem. Although it is likely that state and trait self-esteem measures are highly correlated (e.g., see 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 also used the reverse-correlation technique to create visual 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 self-evaluation, 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 created in this manner was able to reflect the attitude toward self. In the present study, consistent with Moon et al.'s findings, our results also showed 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.'s study, but that individual personality traits were linked to specific facial configurations in the self-portraits that were recognizable as such by independent raters.

Our study further extends existing knowledge in several key ways. First, although Moon et al. (2020) 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 confirmed 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 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 controlled for real facial features, which is crucial to avoid confounds and to provide a valid, strict test of our hypothesis. Finally, we extended our investigation to consider not only face representations but also body shapes, which enriched and generalized 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 to analyze 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 makeup, 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 serve only 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 makeup 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 focused 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 et al., 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 and those targeting emotional or cognitive aspects of the disorder) as well as assess the effects of treatment across time.

In conclusion, we present a novel way to visually depict how people see themselves in their mind's eye and, in doing so, revealed visual clues 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 the 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.

ORCID iDs

Lara Maister (iD) https://orcid.org/0000-0001-8308-9722
Manos Tsakiris (iD) https://orcid.org/0000-0001-7753-7576

References

Acock, A. C. (2014). *A gentle introduction to Stata* (4th ed.). Stata Press.

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). *OpenFace: A general-purpose face recognition library with mobile applications.* http://cmusatyalab.github.io/openface/

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). 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), Article 6499. 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/10.6084/m9.figshare.4220517.v1

Dotsch, R. (2016). *rcicr: Reverse-correlation image-classification toolbox*. https://github.com/rdotsch/rcicr

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*, *111*(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*. 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*, *61*, 197–207. 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*, *22*, 333–347. 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 & Brain Sciences*, *40*, Article e32. https://doi.org/10.1017/S0140525X16000534

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*, Article 476. https://doi.org/10.3389/fpsyg.2013.00476

Mangini, M. C., & Biederman, I. (2004). 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., Karnath, H.-O., Zipfel, S., & Giel, K. E. (2017). Depictive and metric body size estimation in anorexia nervosa and bulimia nervosa: A systematic review and meta-analysis. *Clinical Psychology Review*, *57*, 21–31. 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*, Article 1149. 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*, *65*, 352–358. 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), Article e85773. 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*. Picador.

Sugiura, M. (2013). Associative account of self-cognition: Extended forward model and multi-layer structure. *Frontiers in Human Neuroscience*, *7*, Article 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*, *19*(12), 719–728.

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