Creating Body Shapes From Verbal Descriptions by Linking Similarity Spaces

Matthew Q. Hill¹, Stephan Streuber², Carina A. Hahn¹,
Michael J. Black², and Alice J. O'Toole¹

¹School of Behavioral and Brain Sciences, The University of Texas at Dallas, and

²Max Planck Institute for Intelligent Systems, Tübingen, Germany

Psychological Science
2016, Vol. 27(11) 1486–1497

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797616663878

pss.sagepub.com

Abstract

Brief verbal descriptions of people's bodies (e.g., “curvy,” “long-legged”) can elicit vivid mental images. The ease with which these mental images are created belies the complexity of three-dimensional body shapes. We explored the relationship between body shapes and body descriptions and showed that a small number of words can be used to generate categorically accurate representations of three-dimensional bodies. The dimensions of body-shape variation that emerged in a language-based similarity space were related to major dimensions of variation computed directly from three-dimensional laser scans of 2,094 bodies. This relationship allowed us to generate three-dimensional models of people in the shape space using only their coordinates on analogous dimensions in the language-based description space. Human descriptions of photographed bodies and their corresponding models matched closely. The natural mapping between the spaces illustrates the role of language as a concise code for body shape that captures perceptually salient global and local body features.

Keywords

visual perception, human body, face perception, open data

Received 1/6/16; Revision accepted 7/20/16

The human body plays a central role in social and emotional communication. People consider a person's body when assessing his or her emotional state (Aviezer, Trope, & Todorov, 2012; de Gelder, de Borst, & Watson, 2015; de Gelder & Hadjikhani, 2006), physical health (Carr & Friedman, 2005; Puhl & Heuer, 2009), attractiveness (Currie & Little, 2009; Peters, Rhodes, & Simmons, 2008), and potential as a mate (Peters et al., 2008). Social attitudes and interpersonal interactions are influenced by body-based characteristics (Currie & Little, 2009; Peters et al., 2008). The human body also provides cues to a person's identity (Hahn, O'Toole, & Phillips, 2016; Rice, Phillips, Natu, An, & O'Toole, 2013; Rice, Phillips, & O'Toole, 2013; Robbins & Coltheart, 2012). The role of the body in recognition and in the assessment of emotional states can escape conscious awareness, but it is detectable with physiological measures of eye movements and body responses (Aviezer et al., 2012; Rice, Phillips, Natu, et al., 2013).

Given the importance of the body for recognition and social interaction, it is not surprising that language

provides a rich corpus of colorful body-descriptor terms that allow people to communicate information about body shape to others. These terms can be combined in ways that give rise to vivid mental images. The “shapely, hourglass woman”; the “muscular, athletic young man”; and the “stout, portly gentleman” are imagined easily as complete human body shapes. The images conjured by these descriptions come from combinations of features that reference global shapes (e.g., pear-shaped) and local parts (e.g., long legs), as well as biologically relevant body structure (e.g., masculine) and mobility or strength potential (e.g., athletic, fit). These linguistic descriptions can provide a flexible framework for representing qualitatively diverse visual features about bodies that can exist

Corresponding Author:

Matthew Q. Hill, The University of Texas at Dallas, School of Behavioral and Brain Sciences, GR4.1, 800 W. Campbell Road, Richardson, TX 75080-3021

E-mail: matthew.hill@utdallas.edu

at different spatial scales and across multiple levels of abstraction.

How does one translate linguistic descriptions of people into visual images and visual images into descriptions? We explored the relationship between the physical dimensions of body shape and the words commonly used to describe bodies. We approached this question by constructing a multidimensional similarity-space representation of bodies using people's word-based descriptions of full-body photographs of clothed women. The proximity of two bodies in this *language space* represents the similarity of the descriptions applied to them. To link the language space to the physical variability of human body shapes, we collected verbal descriptions of three-dimensional graphics models of bodies (Allen, Curless, & Popović, 2003) and projected these descriptions into the language space, which provided physical-shape anchors to connect the linguistic descriptions to the three-dimensional shapes.

This approach was facilitated by an unexpected observation: The first five axes of the language space seemed to capture body-description "features" that could be used to label the axes of variability from a geometric *shape space* derived from three-dimensional laser scans of 2,094 bodies (cf. Angelov et al., 2005). The order of the axes in the two spaces differed, but the potentially analogous structure of these multidimensional spaces offered a direct route for testing the relationship between body shapes and descriptions. Specifically, if the axes from the language and shape spaces coarsely correspond, it should be possible to generate a three-dimensional physical body shape of an individual from the verbal description of a photograph of that person.

In the context of the similarity space, a person's description takes the form of a point in the multidimensional language space. This point is specified by its coordinates in the space, which indicate where the person's description stands relative to the descriptions of other people on each axis in the space. Thus, it was possible to select individual points from the language space (i.e., based on descriptions of photographs), and after transposing axes to account for rank-order differences between the two spaces, place the points into the body-shape space. We could then create three-dimensional graphics models of bodies at these locations in the shape space. We hypothesized that if sparse language captures the complex physical variability of bodies, descriptions of the three-dimensional bodies in the shape space should correspond to—or approximately "match"—the descriptions from which those bodies were generated.

To summarize our methodology and findings, we first created a mapping from semantics (i.e., word-based body descriptions) to three-dimensional body shape by linking five multivariate axes across two independently derived

similarity spaces: One space was based on body descriptions, and the other space was based on body shapes. Second, we projected semantic descriptions of pictures into the body-shape space and generated three-dimensional models. Third, we showed that participants' descriptions of the bodies in the original pictures closely matched the descriptions of their corresponding three-dimensional models. These results indicate that language can act as a surprisingly effective code for body shape, capturing perceptually salient global and local body features.

Method

Participants

Participants were recruited from the subject pool at The University of Texas at Dallas (UTD) through an online sign-up system and were compensated with one research credit for a psychology course. The UTD Institutional Review Board approved all experimental procedures. Twelve people (all female) participated in a pilot study to validate the description terms. In the main experiment, 60 volunteers (30 female, 30 male) rated the body photographs, and another 60 volunteers (30 female, 30 male) rated the three-dimensional body reconstructions. There were no overlapping participants in the three groups. Because the study did not employ a traditional experimental design with manipulated variables, the goal was to include enough participants to achieve stable ratings of the bodies for the multivariate analysis. We conducted the multivariate analysis when we had tested 15 participants per block (45 total) and repeated the analysis with 20 per block (60 total). The interpreted axes remained stable with the additional participants. The data analysis reported includes the full set of participants.

Stimuli

Screen captures of 164 women were taken from videos in the Human ID Database (O'Toole et al., 2005) that show people's full bodies as they walked toward a camera. Two frames were selected from each video: a standing image and a midstride image. Standing images were captured from the first frame of the video (~13.6 m from the camera) or from the earliest frame showing the person at rest with hands at her sides. Midstride images were captured from the last full-stride frame in which the entire body was visible. Standing images were somewhat blurrier than midstride images because of the difference in the distance from the camera.

The image frames were cropped to remove excess background and resized to a uniform height of 900 pixels, with each image's original aspect ratio preserved. The Adobe Photoshop CS5 (Version 12.0) sponge filter was

applied to the whole image to obscure facial detail. This filter also preserves and sharpens body contours. Figure 1a shows an example of the processed body photographs.

Descriptor terms

Descriptor terms were sourced first from online dating profiles and clothing store fit recommendations to produce an initial list. Next, a pilot study was conducted to refine this list. Participants in this pilot study freely described the bodies in the images by typing a short description of each person's body type in a small text box. They were told to ignore the face, clothing, hair, and race. The initial list was augmented using some of the words that appeared commonly in the pilot study. The final list of terms captured global shape features, such as "round" and "rectangular"; local features, such as "long legs" and "short torso"; gender-related terms, such as "curvy" and "masculine"; and health-related terms, such as "muscular" and "fit." Table 1 lists the 27 body descriptor terms used.

Procedure

On each trial, participants simultaneously viewed a person's standing and midstride image, along with the list of 27 descriptor terms (see Fig. 1). For each body, participants clicked one of three radio buttons to indicate whether that descriptor did not apply, applied somewhat, or applied perfectly to the body. At the outset of each trial, the radio buttons were all set to "does not apply." This made the task less tedious because participants usually selected only a small number of terms that applied somewhat or perfectly to the person. Trials were self-paced. Data ratings were carried out on 165 females and 60 males.¹ Because of the length of the task, photos were counterbalanced such that each participant saw a different subset of 75 of the total 225 people. Across all participants, a full set of ratings was obtained from 20 participants for each body.

Language-based body-description space. The language space was constructed using correspondence analysis (Benzécri, 1976; Greenacre, 2007)—a multivariate analysis method for categorical data, similar in form to principal component analysis. Correspondence analysis was applied to ratings of 164 of the 165 female body photographs using only ratings of "applies perfectly."² One female, who was perceived consistently as a male,³ was omitted from the analysis. The input to the correspondence analysis was an $I \times J$ matrix, X , of counts tallied across raters, where I was the number of bodies, and J was the number of descriptors. X_{ij} contained the number of participants who rated the j th descriptor as applying perfectly to the i th body. The categorical nature of the data, expressed in a χ^2 contingency table, supports a

biplot visualization of the rows (bodies) and columns (descriptors) in the same space. Individual axes were interpreted using the contribution scores of the descriptor terms. These indicate the importance of a term for establishing a component and are defined formally as the ratio of the squared factor score to the axis eigenvalue (Abdi & Williams, 2010). Contributing descriptors were selected using a rule of thumb from Abdi and Williams (2010) that assigns importance to contribution scores greater than $1/n$, where n is the number of descriptors.

As applied here, the correspondence analysis produced a multidimensional representation that enabled visualization of the bodies and descriptor terms in a common space. The distance between bodies in this space is a measure of the similarity of the linguistic descriptions applied to them. Because the terms and bodies coexist in the same space, each axis was interpreted by finding the descriptor terms with the highest axis-contribution scores. Interpretations were made by comparing terms with large contribution scores that projected to opposite (i.e., positive and negative) sides of an axis.

Figure 2 illustrates the first four axes of the space and provides a qualitative interpretation of each. The first two axes roughly correspond to weight and height. The next two axes are related to different aspects of feminine appearance: Axis 3 spans classically feminine shapes (pear-shaped and curvy) versus women with various "other," not classically feminine shapes, and Axis 4 spans masculine women versus curvy women. The fifth axis (not plotted in Figs. 2a and 2b) was interpretable as waist-height or torso-to-leg-length ratio (see Figs. 2c and 2d).

Body-shape space. The geometric shape space was an extended version of the Shape Completion and Animation of People (SCAPE) model of body pose and shape variation applied to data from laser scans of people (Anguelov et al., 2005). In the SCAPE model, body shape is represented in terms of 3×3 deformation matrices (Sumner & Popović, 2004) consisting of transformations of triangles in a template mesh into triangles in an instance mesh (see Anguelov et al., 2005, for full details). A template mesh with 86,200 triangles was aligned (registered) to 2,094 laser scans of women from the Civilian American and European Anthropometry Resource Project (CAESAR) data set (cf. Piryankova et al., 2014). CAESAR contains full-body laser scans of American and European volunteers between the ages of 18 and 65 years; all are wearing bicycle-style shorts, and women are also wearing a sports bra. The alignment process (Hirshberg, Loper, Rachlin, & Black, 2012) puts all the shapes into correspondence, which enables statistical analysis. Further, the SCAPE representation separates pose and body shape, which allows analysis of only body shape.

Fig. 1. Example rating screen for (a) photographed bodies and (b) three-dimensional body models. For each stimulus, participants viewed one standing image and one midstride image simultaneously, and they clicked one of three radio buttons to indicate whether each of 27 descriptor terms applied perfectly, applied somewhat, or did not apply to that particular stimulus. Photographed bodies were shown with a filter applied to the whole image to obscure facial detail.

Table 1. Body Descriptor Terms Used in the Experiment

proportioned	short legs	pear-shaped
rectangular	long legs	long
stocky	long torso	curvy
big	short torso	masculine
muscular	small	tall
average	lean	sturdy
skinny	round (apple)	built
fit	petite	heavysset
short	broad shoulders	feminine

A low-dimensional shape space was created by applying principal component analysis to the “shapes” of these bodies, defined as the triangle deformations of the 2,094 aligned

template bodies. This gave a three-dimensional-morphable representation of bodies that allowed for smooth transitions of body shape across arbitrary multivariate trajectories in the space.

Although the principal component axes are purely geometric, they have approximate linguistic interpretations (Fig. 3a). Axes of the shape space were interpreted by visual inspection of bodies produced by adding and subtracting three standard deviations to and from each principal component (Fig. 3a). We tentatively interpreted the shape space axes as representing (a) weight, (b) masculine versus curvy, (c) height, (d) waist height, and (e) classically feminine shapes versus other shapes. These labels were chosen from the labels we applied to the first five axes of the language space, but with the rank order

Fig. 2. Axes of the language space describing female body types. Plots of (a) Axes 1 and 2 and (b) Axes 3 and 4 illustrate the relationship between body images (gray circles) and descriptor words (blue circles). The size of the blue circles reflects the relative magnitude of their contribution scores. The percentage of explained variance is indicated next to all axis labels. The descriptor terms with the highest contribution scores for each of the five axes are shown in (c). Terms with negative projection scores appear on the left, and terms with positive scores appear on the right. The descriptor words with higher contribution scores appear in darker fonts. Semantic interpretations of the axes (d) are based on the descriptors with the highest contribution scores at the opposite ends of each axis.

Fig. 3. Shape-space bodies and their relationship to photographed bodies. The renderings in (a) illustrate the first five axes of the shape space. For each component, a body 3 standard deviations above the origin and 3 standard deviations below the origin is shown. The schematic (b) shows the process whereby three-dimensional body models of people in photographs were generated from the descriptions of the photographs. These descriptions were used to generate a five-dimensional language subspace that was aligned with the five-dimensional shape subspace on the basis of analogously interpreted axes. Corresponding points in the two subspaces were then used to generate the body models. Example pairs of photographed bodies and their corresponding approximated three-dimensional bodies (c) are also shown.

of the axes shifted between the two spaces. Specifically, we hypothesized parity between the following axes: (a) the first axis in both the language and shape space (weight), (b) Axis 2 in the language space and Axis 3 in the shape space (height), (c) Axis 3 in the language space and Axis 5 in the shape space (classically feminine shapes vs. other), (d) Axis 4 in the language space and Axis 2 in the shape space (masculine vs. curvy), and (e) Axis 5 in the language space and Axis 4 in the shape space (waist height).

Generating three-dimensional body models from descriptions. Given the potentially analogous interpretation of the first five axes of the language and shape spaces, we aligned the two spaces by reordering the axes in the shape space to match the analogously interpreted axes in the language space. Next, we approximated the shapes of the 164 photographed individuals by synthesizing their bodies in the shape space at positions specified by their locations (coordinates) in the language space. Figure 3b shows a schematic of this process. We began

with the coordinates of each person on the first five (interpreted) dimensions of the language space and constructed a corresponding three-dimensional body using the first five dimensions of the shape space. We did this for all of the 164 photographed bodies. For each body, synthesis in the shape space was accomplished by a linear combination of the principal components, where the weight applied to each principal component was the projection coefficient or “coordinate” specified by its position in the language space on the analogous dimension.

More concretely, the principal component shape-space model allows for a low-dimensional representation of body shape in the subspace U , defined by the first five “interpreted” dimensions. An individual shape S_j is represented by a set of five linear coefficients, β_j , that represent a body’s coordinates with respect to the principal components. Thus, a body at a position in this five-dimensional subspace is approximated as $S_j = U_j\beta_j + u$, where u is the average body (cf. Piryankova et al., 2014). To create body models of the 164 photographs using their coordinates in the language space, we standardized their factor scores on the first five axes. Next, we reordered the coordinates of the photographed bodies in the language space to match the analogously interpreted axes in the shape space. These new standardized coordinates were used to synthesize bodies at these positions in the shape space. Thus, the modeled body was created as a weighted sum of the first five principal components, with the coordinate vectors serving as the weights. This produced 164 synthetic geometric body models, which we rendered in two poses (standing and midstride) to match the photographic bodies. Figure 3c shows example photographs and their corresponding synthesized three-dimensional bodies. A casual visual inspection shows that these appear quite similar to each other.

Human descriptions of three-dimensional bodies. To formally test the resemblance between the two-dimensional photographs and their three-dimensional synthesized body approximations, a new set of participants rated the synthesized bodies using a procedure identical to that used for the body photographs but with the rendered models replacing the photographs (see Fig. 1b). For the analysis, we represented these data as a 27-element description vector that contained the frequency with which these descriptors were judged by participants to apply perfectly to the body, as was done for the photographs.

Results

To test the perceptual similarity between the photographed bodies and the three-dimensionally rendered bodies, we projected the description vectors for the three-dimensional

body models as supplementary points onto axes of the language space. If the models resemble the photographs, we would expect the description vector projections (points) to be close in the multivariate language space to the descriptions (points) of the photographed bodies used to create them. For brevity, we use the term *true-match pair* to refer to a pair of points in the language space representing, respectively, a three-dimensional body’s description point and the description point of the photographed body from whose coefficients it was generated. *Nonmatched pairs* refer to a three-dimensional body’s description point and the description point of an unrelated photographed body. Figure 4 shows a schematic of the method used for determining the similarity of descriptions for the photographed and three-dimensional body models.

A bootstrap hypothesis test was used to determine whether the descriptions of true-match pairs were closer in the language space than descriptions of randomly sampled nonmatched pairs. For 1,000 iterations, we selected random samples of 164 nonmatched pairs and computed the mean Euclidian distance between pair descriptions in the five-dimensional language space. This yielded a distribution of nonmatched sample means that we compared with the average true-match distance. For inferential purposes, we selected a two-tailed cutoff value of $p < .05$. In the top left graph in Figure 5, the matched mean is shown along with the histogram of the 1,000 nonmatched samples. There is a wide separation between the matched mean and the nonmatched bootstrap distribution, with no overlap between the mean of the true-match pairs ($M = 0.5626$) and the distribution of 1,000 nonmatched sample means ($M = 1.1449$, 95% confidence interval, or CI = [1.1428, 1.1469]). This indicates that the descriptions of the true matches were more similar to each other than were descriptions of random nonmatched pairs. Therefore, the language-based descriptions of photographed bodies were sufficient to synthesize three-dimensional reconstructions of body shapes that matched these descriptions. This synthesis was accomplished by linking a language space, derived from a handful of descriptor terms, to a shape space derived from a large and independent sample of human bodies.

Next, we asked whether the resemblance was based on the pattern of variation captured by the combination of all five dimensions in the language subspace or by one or two perceptually salient dimensions. To dissect the role of individual axes of variation in resemblance, the distances between the projected and original points were recomputed along single dimensions in the language space corresponding to (a) weight, (b) height, (c) classically feminine versus other, (d) masculine versus curvy, and (e) waist height. Figure 5 shows the results of a bootstrap test of the sample means of the nonmatched pairs

Fig. 4. Schematic illustration of the method for measuring resemblance in the language space between photographed bodies (blue circles) and three-dimensional body models (red circles). As illustrated, each three-dimensional body model is projected as a point into the language space using its description vector. The dotted purple lines indicate true-match pairs (i.e., a photo and its corresponding model body). An example of a true-match pair is circled at the upper left of the language space. The oval at the bottom highlights a nonmatched pair, illustrating the distance between descriptions of a photographed body and an unrelated three-dimensional body model.

along the individual axes. There was no overlap ($p < .001$) between the mean of the true-match pairs and the bootstrap histogram for the axes corresponding to weight (Axis 1; true-match: $M = 0.1710$; nonmatched: $M = 0.6961$, 95% CI = [0.6935, 0.6988]), height (Axis 2; true-match: $M = 0.1890$; nonmatched: $M = 0.4214$, 95% CI = [0.4199, 0.4229]), and classically feminine shapes versus other shapes (Axis 3; true-match: $M = 0.2187$; nonmatched: $M = 0.3750$, 95% CI = [0.3737, 0.3763]). For the waist-height axis, the descriptions of the true-match pairs were significantly more similar to each other than the bootstrapped nonmatched pairs (Axis 5; true-match: $M = 0.1976$; nonmatched: $M = 0.2329$, 95% CI = [0.2320, 0.2338], $p = .013$). For the masculine-versus-curvy axis (Axis 4), the true pairs were marginally more similar than the descriptions made to random pairs (true-match: $M = 0.2693$; nonmatched: $M = 0.3018$, 95% CI = [0.3007, 0.3029], $p = .067$). These results indicate that four of five dimensions

contributed significantly to the overall resemblance between the photographs and three-dimensional models. The remaining dimension was related marginally ($p = .067$) to the shape variation. Given that Axis 4 was not significant, we recomputed the bootstrap on a four-dimensional subspace, eliminating this axis. As was the case for the full five-dimensional subspace, there was no overlap between the mean of the matched identity distances and the histogram of means for sampled nonmatched identities (subspace of Axes 1, 2, 3, and 5; true-match: $M = 0.3963$; nonmatched: $M = 1.0734$, 95% CI = [1.0712, 1.0756]).

A closer look at the data also suggested that this resemblance was sufficient to support body-shape categorization but not identification as a unique individual. This finding is consistent with humans' preferential reliance on faces for identification (Burton, Wilson, Cowan, & Bruce, 1999; Hahn et al., 2016; Robbins & Coltheart,

Fig. 5. Results of the bootstrap test of pair distances in the language space. Each panel shows the placement of the mean of distances between members of true-match pairs, along with the histogram of mean distances of nonmatched identity pairs. Results are shown separately for the five-dimensional language subspace and each individual axis within that subspace.

2012). Because categorical information supports social and affective judgments that rely on coarse shape (e.g., femininity, athleticism, obesity), particular body descriptions apply accurately to many people. Therefore, the projection of a description of a three-dimensional body reconstruction should be close, not only to the body whose description was used to create it but also to other categorically similar bodies. Evidence for this categorical role can be seen by looking at the rank of the proximity of the true match with respect to all other nonmatched pairs (Fig. 6). The description of the three-dimensional body was the closest point (i.e., body description) to its matched photograph's description in only 8% of the 164

cases, but it was among the 10 closest points in 40% of cases and among the 30 closest points in 80% of the cases.

Discussion

Language can act as a concise code for complex body shapes, which thereby provides a support structure for the social and interpersonal judgments people make about others on the basis of their bodies. Linguistic body descriptors are compact but can operate as powerful perceptual discriminators. Here, the primary axes of variation that emerged from short, language-based descriptions

Fig. 6. Cumulative percentage of matched pairs as a function of ranked proximity in the five-dimensional language subspace. The dashed lines serve as a visual aid to show the percentage of matched identity pairs at ranks 1, 10, and 30.

of a modest number of people mirrored those extracted from high-resolution three-dimensional shape measurements captured from a far larger number of people. This provided a simple and direct method for translating between verbal descriptions of people and complete human body shapes. It is worth noting that an easily visible parity of axis interpretation (e.g., weight, height, waist-height) for the language and shape space is not necessary to establish a computationally valid mapping between descriptions and body shapes. This is because standard machine-learning techniques can be applied to learn these mappings using a sample of descriptions and their associated body shapes (cf. Streuber et al., 2016). The direct machine-learning approach bypasses the underlying psychology of the language description space that ultimately makes it possible to generate bodies using the capacity of language to encapsulate a diverse set of visual features. These complex features underlie such concepts as “stout, portly gentlemen” and “shapely, hourglass women.”

Language is fundamentally a tool for conveying meaning, with “meaning” in this case referring to body shape. A fundamental task of the visual system is to convert retinal images into representations of meaningful objects. Early visual processing extracts generic image-based features by filtering retinotopic data. High-level visual cortex, however, is organized around categories of objects (Grill-Spector & Weiner, 2014). Indeed, body-selective cortical areas in the ventral visual stream (Downing, Jiang, Shuman, & Kanwisher, 2001; Peelen & Downing, 2005) are anatomically distinct, even from presumably “related” face-selective areas (Schwarzlose, Baker, & Kanwisher, 2005).

Remarkably little is known about the visual features that operate within category-specific cortex. Using bodies as a microcosm, the present results suggest that high-level

visual features might be better modeled with linguistic constructs than with multipurpose shape primitives. By contrast to features in early visual processing, these high-level features would be category specific and would provide a flexible framework for representing qualitatively diverse, category-specific visual information (local, global, gender-specific, strength, and movement potential information) about bodies. This information can exist at different spatial scales and across multiple levels of abstraction. The perceptual relevance of these kinds of features provides psychological grounding for computational vision work that has harnessed human-sourced language-based descriptions of images to train machine-learning algorithms. This approach has made impressive progress on long-standing challenges in classifying objects (Farhadi, Endres, Hoiem, & Forsyth, 2009; Lampert, Nickisch, & Harmeling, 2014; Yumer, Chaudhuri, Hodgins, & Kara, 2015), faces (Kumar, Berg, Belhumeur, & Nayar, 2011), human activities (Ziaeefer & Bergevin, 2015), and scenes (Li & Li, 2007).

In applied terms, eyewitnesses describe suspects’ bodies as commonly as their faces (Reid, Nixon, & Stevenage, 2014), which indicates a pressing need for mechanisms that translate between body descriptions and shapes. Comparative judgments of even single-word body descriptors (i.e., taller, heavier) can be used in computational applications to improve body descriptors for identification (Reid et al., 2014). Our approach goes beyond comparative descriptions but is constrained by demographics, language, and culture. Language and body-shape mappings will vary across ethnicity and demographics (e.g., age, sex), and particular mappings will apply only among people sharing the same language and cultural views of body shape. Therefore, tuning these mappings to subpopulations is needed to optimize them for forensic applications. Notwithstanding, it is possible to engineer the general principles uncovered here with a classical machine-learning approach to make flexible and potentially customizable systems for mapping between descriptions and body shapes (see Streuber et al., 2016). Although customized mappings that work for particular languages and bodies from specific ethnic and racial groups remain a topic for future research, these tools could be useful for studying cultural and health-related aspects of person perception.

From a broader perspective, the effective use of language to guide body-shape analysis highlights the need for high-level visual representations to stand at the interface between understanding complex shapes and making this understanding available to others through language.

Action Editor

D. Stephen Lindsay served as action editor for this article.

Author Contributions

M. Q. Hill, C. A. Hahn, and A. J. O'Toole designed the experiment. M. Q. Hill programmed the experiment and collected and analyzed the data. S. Streuber and M. J. Black provided geometric body stimuli and assisted in creating the sampling methodology. C. A. Hahn assisted in data analysis. All authors wrote the manuscript.

Acknowledgments

We thank Jake Swindle and Linda Nguyen for assisting in data collection.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Open Practices

All data have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/32tsq/>. The complete Open Practices Disclosure for this article can be found at <http://pss.sagepub.com/content/by/supplemental-data>. This article has received the badge for Open Data. More information about the Open Practices badges can be found at <https://osf.io/tyvxyz/wiki/1.%20View%20the%20Badges/> and <http://pss.sagepub.com/content/25/1/3.full>.

Notes

1. We report only the analysis of female bodies to focus more on the language-vision relationship than on the actual emergent features for a particular demographic. We assume that features will differ for sex, race, and age groups. We chose to report the analysis for females rather than males, because the data set contained fewer males ($n = 60$) than females ($n = 165$). See Streuber et al. (2016) for a related machine-learning approach applied to male bodies.
2. The “applies perfectly” term was used for the analysis to declutter the correspondence-analysis visualization. The interpretation of axes was the same when all response choices were input to the correspondence analysis.
3. In a space with male and female bodies, this body clustered with males on Axis 1, which otherwise separated bodies by sex. We agreed that this body appeared male.

References

- Abdi, H., & Williams, L. J. (2010). Correspondence analysis. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (Vol. 1, pp. 267–278). Thousand Oaks, CA: Sage. doi:10.4135/9781412961288.n83
- Allen, B., Curless, B., & Popović, Z. (2003). The space of human body shapes: Reconstruction and parameterization from range scans. *ACM Transactions on Graphics*, 22, 587–594. doi:10.1145/882262.882311
- Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., & Davis, J. (2005). SCAPE: Shape completion and animation of people. *ACM Transactions on Graphics*, 24, 408–416. doi:10.1145/1073204.1073207
- Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338, 1225–1229. doi:10.1126/science.1224313
- Benzécri, J.-P. (1976). *L'analyse des données* [Data analysis]. Paris, France: Dunod.
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10, 243–248.
- Carr, D., & Friedman, M. A. (2005). Is obesity stigmatizing? Body weight, perceived discrimination, and psychological well-being in the United States. *Journal of Health and Social Behavior*, 46, 244–259. doi:10.1177/002214650504600303
- Currie, T. E., & Little, A. C. (2009). The relative importance of the face and body in judgments of human physical attractiveness. *Evolution & Human Behavior*, 30, 409–416. doi:10.1016/j.evolhumbehav.2009.06.005
- de Gelder, B., de Borst, A. W., & Watson, R. (2015). The perception of emotion in body expressions. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 149–158. doi:10.1002/wcs.1335
- de Gelder, B., & Hadjikhani, N. (2006). Non-conscious recognition of emotional body language. *NeuroReport*, 17, 583–586. doi:10.1097/00001756-200604240-00006
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470–2473. doi:10.1126/science.1063414
- Farhadi, A., Endres, I., Hoiem, D., & Forsyth, D. (2009). Describing objects by their attributes. In *Proceedings of the 2009 IEEE International Conference on Computer Vision and Pattern Recognition* (pp. 1778–1785). Los Alamitos, CA: IEEE. doi:10.1109/CVPR.2009.5206772
- Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15, 536–548. doi:10.1038/nrn3747
- Hahn, C. A., O'Toole, A. J., & Phillips, P. J. (2016). Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology*, 107, 117–134. doi:10.1111/bjop.12125
- Hirshberg, D. A., Loper, M., Rachlin, E., & Black, M. J. (2012). Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *Lecture Notes in Computer Science* (Vol. 7577, Part 6, pp. 242–255). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-642-33783-3_18
- Kumar, N., Berg, A., Belhumeur, P. N., & Nayar, S. (2011). Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 1962–1977. doi:10.1109/TPAMI.2011.48
- Lampert, C. H., Nickisch, H., & Harmeling, S. (2014). Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 453–465.

- Li, L.-J., & Li, F.-F. (2007). What, where and who? Classifying events by scene and object recognition. In *Proceedings of the 2007 IEEE International Conference on Computer Vision* (pp. 1–8). Los Alamitos, CA: IEEE. doi:10.1109/ICCV.2007.4408872
- O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., & Abdi, H. (2005). A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, 812–816. doi:10.1109/TPAMI.2005.90
- Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, 93, 603–608. doi:10.1152/jn.00513.2004
- Peters, M., Rhodes, G., & Simmons, L. W. (2008). Does attractiveness in men provide clues to semen quality? *Journal of Evolutionary Biology*, 21, 572–579. doi:10.1111/j.1420-9101.2007.01477.x
- Piryankova, I. V., Stefanucci, J. K., Romero, J., de la Rosa, S., Black, M. J., & Mohler, B. J. (2014). Can I recognize my body's weight? The influence of shape and texture on the perception of self. *ACM Transactions on Applied Perception*, 11, Article 13. doi:10.1145/2641568
- Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity*, 17, 941–964. doi:10.1038/oby.2008.636
- Reid, D. A., Nixon, M. S., & Stevenage, S. V. (2014). Soft biometrics: Human identification using comparative descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 1216–1228. doi:10.1109/TPAMI.2013.219
- Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science*, 24, 2235–2243. doi:10.1177/0956797613492986
- Rice, A., Phillips, P. J., & O'Toole, A. (2013). The role of the face and body in unfamiliar person identification. *Applied Cognitive Psychology*, 27, 761–768. doi:10.1002/acp.2969
- Robbins, R. A., & Coltheart, M. (2012). The effects of inversion and familiarity on face versus body cues to person recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 1098–1104. doi:10.1037/a0028584
- Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience*, 25, 11055–11059. doi:10.1523/JNEUROSCI.2621-05.2005
- Streuber, S., Quiros-Ramirez, M. A., Hill, M. Q., Hahn, C. A., Zuffi, S., O'Toole, A., & Black, M. J. (2016). Body talk: Crowdshaping realistic 3D avatars with words. *ACM Transactions on Graphics*, 35, Article 54. doi:10.1145/2897824.2925981
- Sumner, R. W., & Popović, J. (2004). Deformation transfer for triangle meshes. *ACM Transactions on Graphics*, 23, 399–405. doi:10.1145/1015706.1015736
- Yumer, M. E., Chaudhuri, S., Hodgins, J. K., & Kara, L. B. (2015). Semantic shape editing using deformation handles. *ACM Transactions on Graphics*, 34, Article 86. doi:10.1145/2766908
- Ziaeefer, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review. *Pattern Recognition*, 48, 2329–2345. doi:10.1016/j.patcog.2015.03.006