Supplementary material

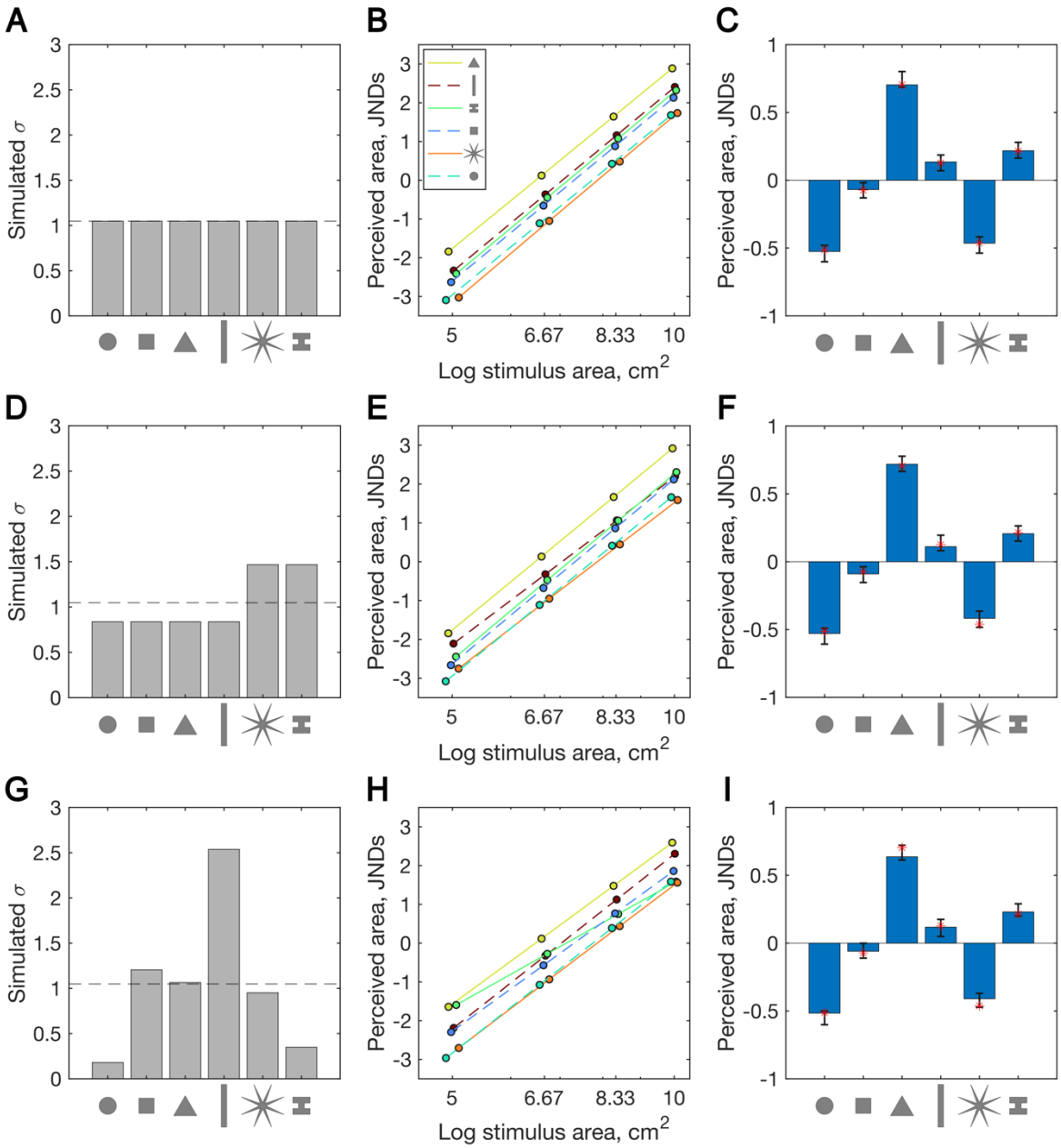
S1. Simulations

Our Thurstonian scaling analysis method assumes that perception of stimulus area is perturbed by noise and that the spread (σ) of the Gaussian noise distribution is the same for every stimulus. Here we investigate the effect on estimated parameters (i.e. perceived relative area) when this assumption is violated, i.e., if shapes differ in noise/uncertainty with respect to their perceived area. To this end, we simulated data for three different variants of observers performing Experiment 1.

The first variant (V1) corresponds to an idealised observer completing Experiment 1, for whom all shape-related biases are perfectly scale invariant, and all stimuli are associated with the same noise ($\sigma = 1.05$, see Figure S1 A). The mean shape-related simulated biases (JNDs, averaged across stimulus size) were fixed to match those measured in Experiment 1. Discriminability within shapes was also matched to the average of those measured in Experiment 1.

Figure S1

Data From Simulations Showing the Effects of Different Noise Distributions



Note. Each row corresponds to one of the three observer variants. (A, D, G) Simulated noise parameters (σ) for each shape. (B, E, H) Perceived area in JNDs for each shape as a function of stimulus size for one observer completing 1000 trials per condition. (C, F, I) Perceived area in JNDs, averaged across stimulus size. Bars show data for the same single observer. Error bars give 95% CIs for the mean biases estimated from 35 simulated observers performing 2 trials per condition, from bootstrapping. Red stars give the mean simulated biases for each shape.

For the second and third observer variants (V2, V3) the simulated means of perceived area (i.e., the simulation JNDs) were unchanged, but we adjusted the noise parameters, violating the

equal-noise assumption of our analysis. In V2, two shapes (rectangle, 8-pointed star) are associated with more uncertainty than the others (see Figure S1 D). Finally, in V3, the noise parameter for each shape was randomly assigned by sampling from a Gaussian distribution (Figure S1 G), resulting in huge variation across shapes.

For each variant, we first simulated a single observer performing the experiment, with 1000 trials for each stimulus pairing featured in the original experiment. This revealed the effects of different simulation parameters, with minimal sampling noise. The results are shown in Figure S1 B, E, H (all stimulus sizes) and also by the bars in Figure S1 C, F, I (biases averaged across size). For comparison, the simulated biases are shown by red stars.

For variant one, as expected, the simulated data reflects the simulation parameters (alignment of bars and stars in Figure S1 C). Under V2, the mismatched noise parameters result in small, spurious deviations from scale-dependent biases: the lines representing the rectangle and star in Figure S1 E have smaller slopes. For these shapes, perceived area is slightly overestimated for smaller stimuli and underestimated for larger stimuli (compare Figure S1 D with the actual data from Experiment 1 in **Error! Reference source not found.** B). However, examination of Figure S1 F reveals that, after averaging across stimulus size, the resultant errors in the estimates of perceived relative area for each shape are minimal. Similarly, even the extreme range of values of σ simulated in V3 produce very small errors in estimated bias, when averaged across stimulus size Figure S1 I.

Next, we investigated the reliability of estimates of perceptual biases for the three simulated observer variants, given the true number of observers and trials in Experiment 1. To this end, for each variant, we simulated 1000 independent observers, each of whom performed the same number of stimulus comparisons as in the real experiment. Then we simulated experimental data collection by randomly selecting 35 of the 1000 candidate observers, and recording the mean shape-related biases, averaged across observers and stimulus size. This was repeated 1000 times, to derive 95% confidence intervals for the mean estimated area biases; these CIs are shown by the error bars in Figure S1 C, F, and I. It is apparent that large deviations from the equal noise assumption are well

tolerated by our method; such deviations do not introduce meaningful systematic errors in estimates of perceptual bias, nor do they alter the reliability of those estimates. As noted above, deviations from equal noise do introduce small systematic (and opposite) errors in estimation of perceived area for small and large stimuli (the different slopes in Figure S1 E and H). In other words, our method can produce spurious scale dependencies in area biases. For this reason, our quantification and modelling of shape-related biases ignores the apparent small deviations from scale invariance.

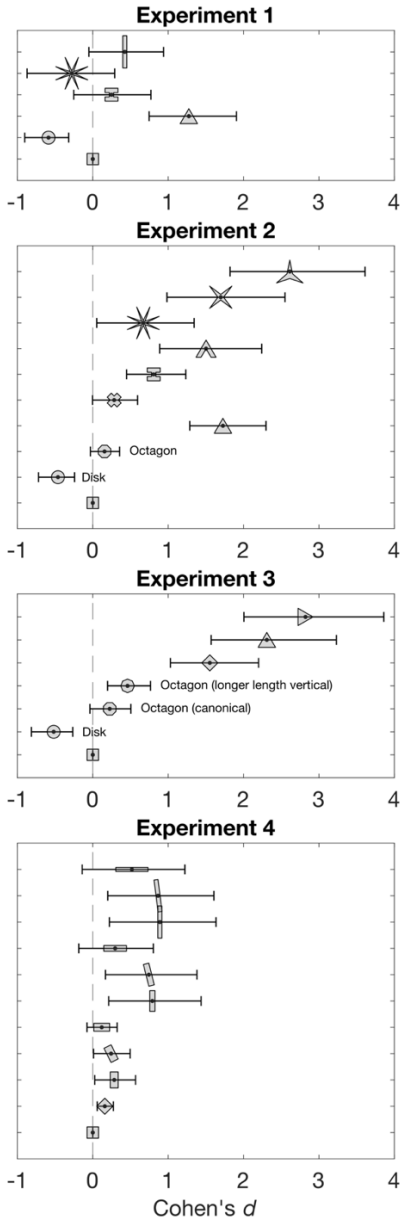
BIASES IN THE PERCEIVED AREA OF SHAPES: SUPPLEMENTARY MATERIAL

S2 Effect sizes

For each experiment, we calculated effect sizes for biases for each condition (shape / orientation combination, averaged across sizes 1–3 for each participant) compared against the reference condition (square, canonical orientation), i.e. perceived $area_{condition}$ - perceived $area_{reference}$. Figure S2 shows Cohen's *d* for each comparison.

Figure S2

Effect Sizes for All Experiments



Note. Marker shapes show the condition being compared to the reference stimulus (square, canonical orientation; vertical dashed line). Error bars show 95% CIs for paired samples (meanEffectSize, MATLAB).

BIASES IN THE PERCEIVED AREA OF SHAPES: SUPPLEMENTARY MATERIAL

S3: Regression analyses for each experiment

For Experiment 1, we ran a regression to predict biases in perceived area as a function of stimulus shape. Each shape was dummy coded relative to the square (common to all experiments). Regression coefficients are shown in Table S1.

Table S1

Regression Coefficients for Experiment 1

Effect	Estimate	*SE*	*p*
(Shape: square)	-0.70	0.14	.60
Shape: disk	-0.44	0.20	.03
Shape: triangle	0.78	0.20	< .001
Shape: h	0.20	0.20	.31
Shape: 8:1 rectangle	0.29	0.20	.14
Shape: 8-point star	-0.39	0.20	.05

For Experiment 2, perceived area bias was regressed on convex hull shape (square, octagonal, triangular), compactness, and their interaction (Table S2). Compactness (3 levels) was modelled quantitatively as the average compactness (area / area of circumscribing disk) for stimuli of high (regular polygons), moderate and low compactness (stars). (Note that the full regression analyses across all shapes (see section S3 below) uses compactness as a continuous variable calculated for each shape). The disk was not included in the current analysis as it was the only shape with a circular convex hull.

Table S2

Regression Coefficients for Experiment 3

Effect	Estimate	*SE*	*p*
(Convex hull: square)	0.40	0.15	.008
Convex hull: octagon	-0.04	0.26	.88
Convex hull: triangle	0.81	0.23	.001
Compactness	-2.83	0.53	< .001
Convex hull: octagon × Compactness	0.70	0.78	.37
Convex hull: triangle × Compactness	1.41	0.76	.06

For Experiment 3, the bias in perceived area was modelled as a function of the stimulus shape (square, octagon, triangle), orientation (canonical vs. longest length vertical) and their interaction (Table S3). The disk was excluded as it is orientation-invariant.

Table S3

Regression Coefficients for Experiment 3

Effect	Estimate	*SE*	*p*
(Shape: square; Orientation: canonical)	-0.80	0.11	< .001
Shape: octagon	0.26	0.16	.11
Shape: triangle	1.84	0.16	< .001
Orientation: longest vertical	1.11	0.16	< .001
Shape: octagon x Orientation: longest length vertical	-0.98	0.16	< .001
Shape: triangle x Orientation: longest length vertical	-0.64	0.16	.004

For experiment 4, biases in perceived area were regressed on stimulus orientation (longest length vertical vs. longest edge horizontal), elongation (aspect ratio) and their interaction (Table S4). We excluded the 'longest edge vertical' orientation as it is equivalent to the first for square shape.

Table S4

Regression Coefficients for Experiment 4

Effect	Estimate	*SE*	*p*
(Orientation: longest length vertical)	-0.58	0.17	.001
Orientation: longest edge horizontal	-0.10	0.24	.69
Elongation	0.18	0.04	< .001
Orientation: longest edge horizontal x Elongation	-0.08	0.05	.13

S3 Model selection

We compared regression models using leave-one-out cross validation over shapes. The full list of candidate predictors is reported in Table S5. For each N (the number of predictors), the model with the lowest cross-validation error is considered as the best model. For each model, cross-validation was performed by fitting it to 21 shapes (leaving out 1) and then using the fitted parameters to predict the left out shape, and noting the error. This process was repeated 22 times, leaving out a different shape each time, summing the squared errors. Thus, a model's cross-validation error reflects its ability to generalise from the fitted stimulus set to novel shapes, with a smaller error indicating better performance.

Table S5

List of Geometric Features Evaluated as Predictors in the Linear Regression Models Presented in the Main Paper

	Measure	Description
Surface measures (in cm^2)		
1	Area of base (1)	Area of the bottom 1/8 portion of the shape
2	Area of base (2)	Area of the bottom 1/4 portion of the shape
3	*Area of bounding box*	Area of the smallest rectangle, with horizontal and vertical edges, that encloses the shape
4	Area of circumscribing disk	Area of the smallest disk enclosing the shape
5	Area of circumscribing ellipse	Area of the smallest ellipse enclosing the shape; equals to the area of circumscribing disk for shapes of aspect ratio of 1 and triangular shapes
6	Area of indisk	Area of the largest disk enclosed by the shape
7	Convex area	Area of the shape's convex hull
8	Convex area of base (1)	Area of the bottom 1/8 portion of the shape's convex hull
9	Convex area of base (2)	Area of the bottom 1/4 portion of the shape's convex hull
Linear measures (in cm)		
10	'Additive area'	Sum of the shape's height and width (Yousif & Keil, 2019)
11	*Compactness: circumradius*	Radius of the smallest disk enclosing the shape
12	Compactness: inradius	Radius of the largest disk enclosed by the shape
13	Convex perimeter	Perimeter of the shape's convex hull
14	Height (maximum vertical distance)	Longest distance between two boundary points on the y axis (Warren & Pinneau, 1955)
15	*Height of centroid*	Distance between the shape's base and centroid on the y axis
16	Width (maximum horizontal distance)	Longest distance between two boundary points on the x axis

	Measure	Description
17	Width of base (1)	Average width calculated at the bottom 1/8 portion of the shape
18	*Width of base (2)*	Average width calculated at the bottom 1/8 portion of the shape convex hull
19	Perimeter	(Anastasi, 1936)
Ratios		
20	Compactness (1): Area-to-area of circumdisk ratio	Ratio of the shape area to that of the smallest circumscribing disk
21	Compactness (2): Circularity	Ratio of the shape area to that of a circle of the same convex perimeter (perimeter of convex hull); equals IPQ (below) for convex shapes
22	Compactness (3): Convex area-to-area of circumdisk ratio	Ratio of the shape convex area (area of convex hull) to that of the smallest circumscribing disk (circumdisk)
23	Compactness (4): Isoperimetric quotient (IPQ)	Ratio of the shape area to that of a circle of same perimeter
24	Compactness (5): P2A	Squared perimeter-to-area ratio (Smets, 1970)
25	Compactness (6): Roundness	Indisk-to-circumdisk ratio: Ratio of the area of the maximum inscribed disk (indisk) to the circumdisk of the shape
26	Convexity	Ratio of the shape perimeter to the perimeter of the shape's convex hull
27	*Elongation*	Aspect ratio of the ellipse enclosing the shape; orientation-invariant
28	*Elongation × sin(orientation)*	Elongation × sin(θ); θ is the angle from the x axis to the major axis of the ellipse circumscribing the shape
29	Elongation × cos(orientation)	Elongation × cos(θ)
30	Extent	Ratio of the shape's area to the area of the shape's bounding box
31	Height of centroid-to-height ratio	Ratio of the shape's height of centroid (see above) to the shapes' height
32	Height-to-width ratio	Ratio of the maximum vertical distance (height) to the maximum horizontal distance (width); orientation-dependent (Holmberg & Holmberg, 1969; as cited in Krider et al., 2001)
33	Solidity	Ratio of the shape area to the area of the shape's convex hull
34	Width of base-to-height ratio (1)	Ratio of the average width calculated at the bottom 1/8 portion of the shape to the shapes' height
35	Width of base-to-height ratio (2)	Ratio of the average width calculated at the bottom 1/8 portion of the shape's convex hull to the shapes' height

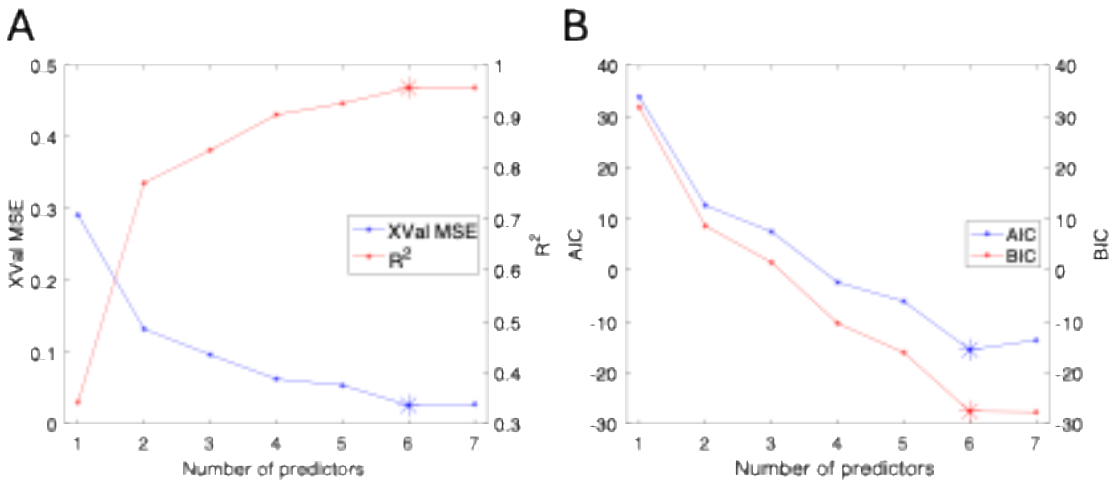
Note. The predictors featuring in the best model are listed in italics. Features within each subset (surface measures, linear measures, ratios) are presented in alphabetical order. The same features were used to construct principal component and partial least squares components to characterise the stimulus space in alternative models (see below).

BIASES IN THE PERCEIVED AREA OF SHAPES: SUPPLEMENTARY MATERIAL

Figure S3 A shows the goodness-of-fit (R^2) and cross-validation error (MSE) for the best one- to seven-predictor models. The best model (i.e., the model with the lowest cross-validation error overall), described in the main text, includes six predictors.

Figure S3

Best Perceived Area Bias Regression Models: Model Comparisons



Note. (A) Cross validation error (MSE, left-hand y axis, blue lines and markers) and goodness of fit (R2, right-hand y axis, red lines and markers) for the best perceived area bias regression models with 1–7 predictors. Models were compared on their cross-validation error. Values for the best model overall are indicated by the large asterisks. (B) AIC (left-hand y axis, blue lines and markers) and BIC (right-hand y axis, red lines and markers) for the best perceived area bias regression models with 1–7 predictors. Values for the best model overall (according to the cross-validation error) are indicated by the large asterisks.

We compared our cross-validation approach to two other methods for model selection: AIC and BIC, to support model comparisons for readers unfamiliar with cross-validation. Figure S3 B shows the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for the best one- to seven-predictor models. AIC supports the same conclusions as LOO XVAL (lowest AIC: 6 predictors), whilst BIC supports a more complex model (7+ predictors). Predictors for the best 1- to 6-predictor models are reported in

Table **S6**.

Table S6

Predictors For the Best N-predictor Models Identified Via Cross-Validation

N	Predictors
1	Compactness (1): Area-to-area of circumdisk ratio
2	Convex area
	Area of circumscribing ellipse
3	Height
	Convex area
	Area of circumscribing ellipse
4	Height
	Perimeter
	Compactness (1): Area-to-area of circumdisk ratio
	Elongation × cos(orientation)
5	Convex area
	Circumradius
	Elongation
	Elongation × cos(orientation)
	Width of base (1)
6	Circumradius
	Elongation
	Area of bounding box
	Width of base (1)
	Height of centroid
	Elongation × sin(orientation)

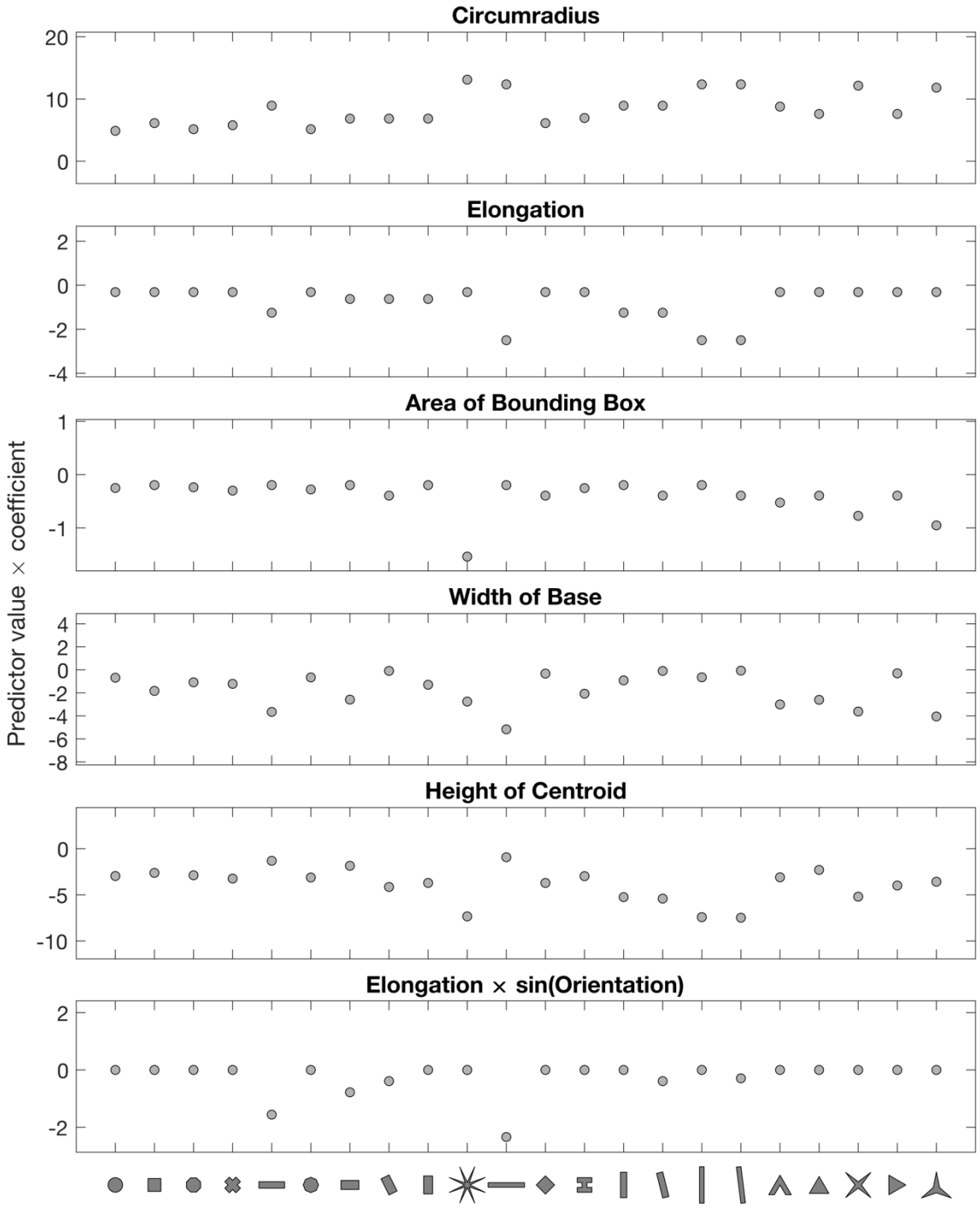
Note. Best N-predictor models were identified via cross-validation. See Table S1 for descriptions and the full list of predictors.

Figure S4 shows the influence of each predictor in in each condition (i.e., shape / orientation combination) for the best model.

BIASES IN THE PERCEIVED AREA OF SHAPES: SUPPLEMENTARY MATERIAL

Figure S4

Best Model: Influence Of Each Predictor in Each Condition



Note. Each condition is a shape / orientation combination. Conditions are reported in order of perceived size (smaller to larger).

S4 Alternative models

The regression model presented in the main paper provides a very good fit to biases in perceived area, using 6 predictors, and was selected via cross-validation, as summarised in Figure S3. However, it is a little difficult to interpret, due to partially correlated predictors and suppressor effects. In addition, candidate predictors were arbitrarily chosen from literature, with further candidate predictors added to capture the observed pattern of biases.

Here we explore whether it is possible to develop a model that is more parsimonious and easily interpretable. PCA (Principal Component Analysis) and PLS (Partial Least Squares) are two methods used to address the issues of correlated predictors and high dimensionality. Both methods entail the construction of new components (latent variables) that more efficiently capture variation across the stimulus space (and, with PLS, the response space). These new components serve as predictor variables in linear regression models.

Using PCA, independent (orthogonal) components can be identified that efficiently capture shape variations across our 22 stimulus shapes. However, this method is blind to the dependent variable (area bias). Thus, whilst the first few principal components will efficiently capture shape variability, these dimensions might be poor correlates of biases in perceived area. In contrast, PLS regression considers variation in both the stimulus space and response space. Components are identified in the multidimensional stimulus (shape) space that explain the maximum variance in the responses (area biases).

Using PCA and PLS, we explored two different characterisations of our stimulus space. Our 22 stimuli were either represented by each shape's geometric feature values (i.e., compactness, elongation etc., as in the original regression model, see Table S5) or by the x-y coordinates of 32 points on the shapes' perimeters. These points were equidistant along each shape's perimeter, with the first point directly above the shape's geometric centre (preserving orientation). In both cases, these input variables were z-scored. Resultant regression models using the PCA (PCR, Principal Component Regression) or PLS (PLSR, Partial Least Squares Regression) components were compared

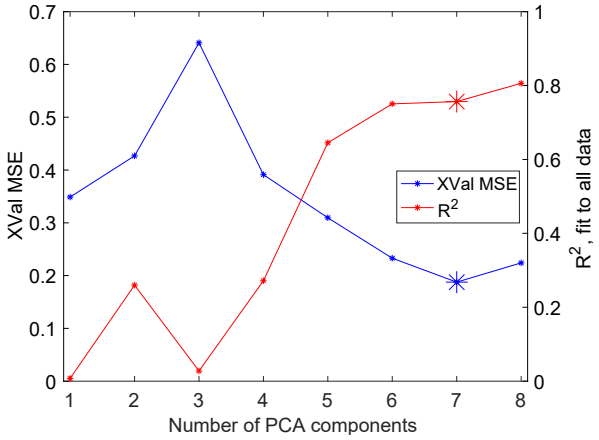
using leave-one-out cross validation over shapes. To preview the key results: PCA produces poor models of biases in perceived area, either using geometric features or perimeter points. Whilst the stimulus shapes are efficiently described by a subset of components, these are poor predictors of area biases. PLS is somewhat better in capturing area biases but is not more parsimonious or interpretable than the regression model presented in the main text. However, PLS components from perimeter points (see below) provides a useful visualisation of shape variations that correspond to variation in perceived area.

PCA, PCR on geometric features

The first seven PCA components of the geometric feature data approximate our shape stimuli very well (98% variance explained). However, the best PCR model (i.e. regression model using the principal components as predictors of area bias, Figure S5) is poor: it includes seven predictors (R^2 = 0.7567, XVal MSE = 0.1876), providing a much worse fit than the original multiple linear regression model on geometric features, compare with Figure S3 A.

Figure S5

Best PCR on Geometric Features Models: Cross-validation Error and Goodness of Fit



Note. Cross validation error and goodness of fit for the model fitted to the full dataset for the best 1- to 8-predictor PCR models. Plotting conventions as in Figure S3 A. Predictors are PCA components of the shape's geometric features. Models were compared on their cross-validation mean squared error (lower = better).

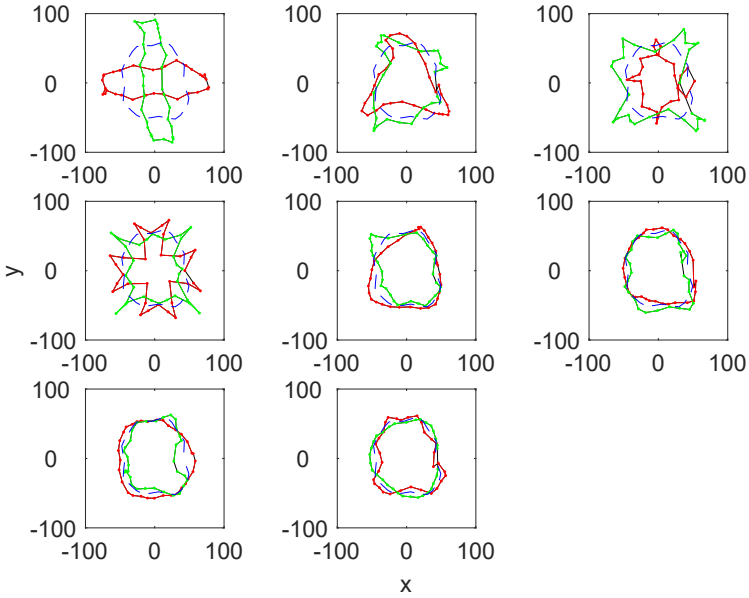
BIASES IN THE PERCEIVED AREA OF SHAPES: SUPPLEMENTARY MATERIAL

PCA, PCR on perimeter locations

Using the 32 equidistant points on each shape's perimeter, 8 principal components provide a good approximation to our shapes (98% variance in shape perimeter locations). Figure S6 shows these components (dashed blue line = mean shape, red and green = + / - 3 SDs of component). The first component appears to represent how tall vs. wide a shape is, whereas the second appears to roughly correspond to triangularity (long side vertical vs. horizontal).

Figure S6

Representation of the First Eight Principal Components of Shape Perimeter Points.

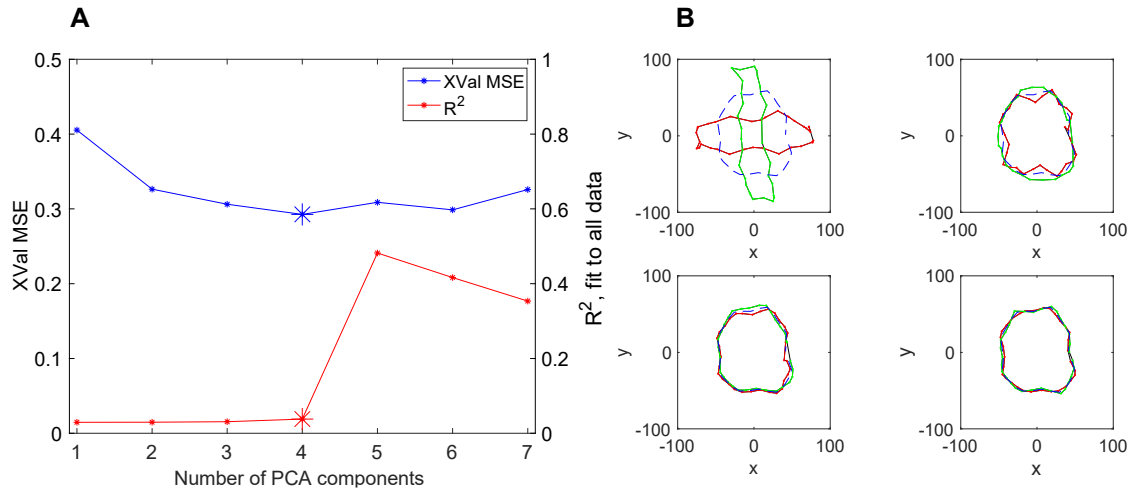


Note. The dashed blue line shows the mean shape. Red and green lines indicate +, - 3 SDs respectively.

Unfortunately, PCR models using perimeter points provide a very poor account of area biases. The best model includes four components (n. 1, 9, 11, 13, Figure S7 B), but the fit is extremely poor (R^2 = 0.0381, XVal MSE = 0.2926, Figure S7 A).

Figure S7

Best PCR on Perimeter Locations Models



Note. (A) Cross validation error and goodness of fit for the model fitted to the full dataset for the best 1- to 7-predictor PCR models. Plotting conventions as in Figure S3 A. Predictors are PCA components of the shape's perimeter points. Models were compared on their cross-validation error (lower = better). Values for the best model overall (i.e., with the lowest cross-validation error) are indicated by the large asterisks. (B) Representation of the four principal components featuring in the best model. Plotting conventions as in Figure S6.

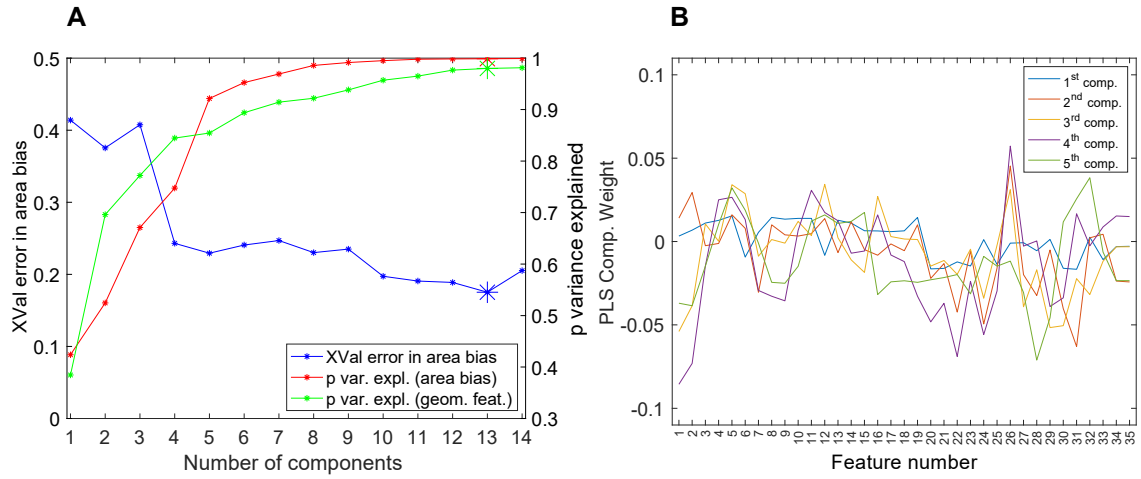
Figure S8 A shows cross-validation error and proportion of variance explained in area biases, as a function of the number of PLS components in a regression model. The best model has 13 predictors. This provides a better fit than the original regression model with geometric features (R^2 0.9799 vs. 0.9556; Figure S3 A), but the cross-validation error is larger (XVal MSE 0.175 vs. 0.026), and the model includes more than double the predictors (13 vs. 6 predictors). A PLSR model with six

predictors gives an R^2 similar to that for the original regression model, but the cross-validation error is larger (R^2 0.9524, XVal MSE 0.241; see

Table **S7** for model comparisons).

Figure S8

Best PLSR on Geometric Features Models



(A) Cross validation error (MSE, left-hand y axis, blue line and markers) and proportion of variance explained in y (perceived area bias; right-hand y axis, red lines and markers) and x (shape geometric features; right-hand y axis, green lines and markers) for the model fitted to the full dataset for the best 1- to 14-predictor PLSR models. Predictors are PLS components of the shape's geometric features. Models were compared on their cross-validation error. Values for the best model overall (i.e., with the lowest cross-validation error) are indicated by the large asterisks. (B) Composition of the first five PLS components, shown as the weight assigned (y axis) to each geometric feature (x axis) within each component (coloured lines). The features corresponding to each number are listed in Table S5).

A limitation of PLS is that the components can be hard to interpret. Figure S8 B shows the composition of the first 5 components (one line each), i.e. the weight given to each geometric feature (see Table S5). The first component assigns a positive weight (larger than an arbitrary threshold of |0.01|) to features that are negatively correlated with compactness (features 3–5, 8–11, height (14), and perimeter (19)) and a negative weight to features that are positively correlated to compactness (compactness measures 20–23 and features 30–31), and negligible weights to features that capture orientation / elongation (e.g. features 26–29, height-to-width ratio (32)). Thus, the first component seems to roughly capture how spread out (in all directions) the shape is. Vertical elongation appears to be captured by the fifth component, which includes positive weights for

features that are positively correlated with a shape's vertical elongation (e.g. centroid-to-height ratio (feature 31), height-to-width ratio (32)), and a negative weight to features that are positively correlated with horizontal elongation (e.g. elongation × orientation (features 28, 29), area of base (1, 2)). However, this component also includes a (smaller) negative weight for orientation-invariant elongation (27). Other components are harder to characterise succinctly.

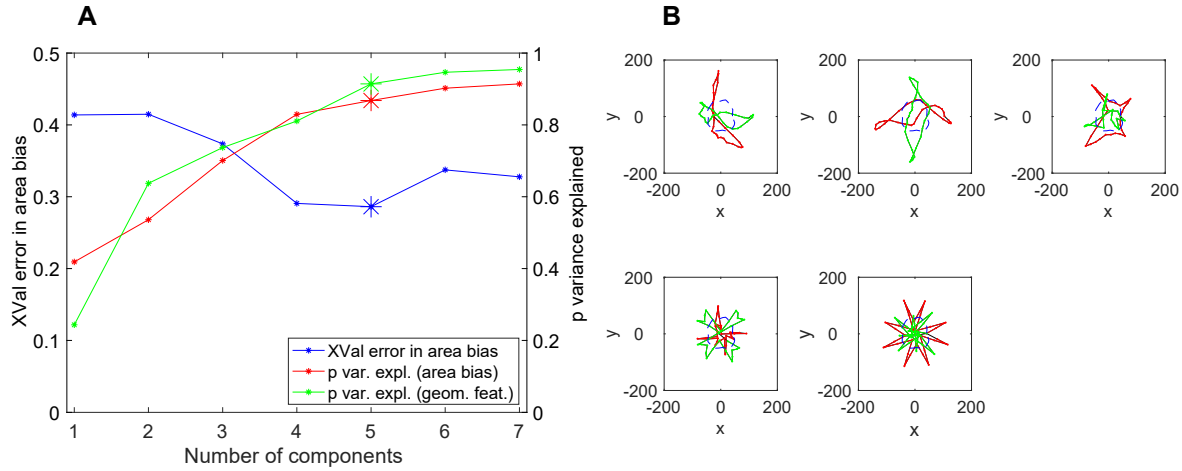
PLS / PLSR on perimeter locations

Figure S9 A gives the cross-validation error and variance explained in biases in perceived area, as a function of the number of PLS components in the regression model. The best model has 5 predictors, shown in Figure S9 B. This model has fewer predictors (5 vs. 6) but has a worse fit and larger cross-validation error (R^2 0.8679 vs. 0.9556, XVal MSE = 0.286 vs 0.026; see

Table **S7** for model comparisons) than the original regression model. Is this PLSR model more parsimonious than the original model? Although it has one fewer predictor, each of the PLS predictors is constructed from the complete stimulus (64 perimeter coordinates) and response variables. Thus, it does not provide a superior or more interpretable account of area biases.

Figure S9

Best PLSR on Perimeter Locations Models



Note. (A) Cross validation error and proportion of variance explained in geometric features and perceived area bias and for the model fitted to the full dataset for the best 1- to 7-predictor PLSR models. Plotting conventions as in Figure S5. Predictors are PLS components of the shape's perimeter locations. (B) Representation of the five components featuring in the best model. The dashed blue line shows the mean shape. Green and red surfaces show + / - 1 PLS component, respectively.

Nonetheless, the PLS components (Figure S9 B) can provide some insight into the shape variations that predict the most variability in perceived area. The first two components (top row, left and middle) capture how vertically vs. horizontally elongated a shape is, but also how 'periphery heavy' (component 1) vs 'centre heavy' (component 2) it is. The third component is hard to interpret. The fourth (bottom row, left) appears to relate to whether a shape is 'cross-like', with points in either cardinal or diagonal orientation. The last component may be primarily capturing the identity of our 8-pointed star stimulus. How do these relate to the geometric features / predictors in the original regression model? This included measures of compactness (circumradius, area of bounding box), elongation (aspect ratio, scale invariant) and its relationship with orientation (elongation × orientation), and 'bottom-heaviness' (width of base, height of centroid). All five components capture compactness, as large positive or negative loadings correspond to lower compactness (circumradius, area of bounding box). The first and second components capture variation in elongation (orientation invariant), and concurrent variations in vertical elongation. The first component also captures the shape bottom-heaviness (width of base, height of centroid). All five components variously capture the height of centroid (with either positive or negative weights).

Model comparisons

We compared the original regression model ('R6', see Figure S3 A) with PLRS models using t-tests on the cross-validation residuals (one residual for each left-out shape, for each model).

Table **S7** reports model comparisons with the best PLSR models (PLS on geometric features: 13 components; PLS on perimeter points: 5 components) and the six-predictor PLSR model with components from geometric features.

Table S7

Model Comparisons

Compared models	**t-test *p***
$R_6 < PLSR_{PP\ 5}$	0.042 *
$R_6 < PLSR_{GF\ 13}$	0.003 *

Compared models	**t-test *p***
R_6 < $PLSR_{GF\ 6}$	0.035 *

Note. Models were compared using t-tests on the cross-validation residuals (one residual for each left-out shape, for each model). The greater-, less-than, equal signs refer to residual magnitudes (i.e., $M_x < M_y$ indicates that M_x residuals are smaller (better) for at least half + 1 of the left-out shapes); R, PCR, PLSR denote regression, principal component regression, partial least squares regression models; GF, PP denote components (predictors) from geometric features, perimeter points. Subscript numbers indicate the number of predictors in each model. Stars show statistical significance ($p < .05$) as indicated by the corresponding t-test p values.