

The shape of novel objects contributes to shared impressions

Aaron Kurosu

Department of Psychology, Princeton University,
Princeton, NJ, USA

Alexander Todorov

Department of Psychology, Princeton University,
Princeton, NJ, USA

How do people share impressions of novel objects, and is this even possible? We tested whether the shape of novel 3-D objects can lead to similar impressions across people. To do this, we introduced a technique for manipulating highly complex shapes and measured four types of evaluative impressions (approachable, dangerous, beautiful, likable). Because relatively little is understood regarding how people form impressions of novel objects, we first sought to confirm the reliability of this behavior by examining how similar impressions are for an individual asked to re-evaluate the stimuli (i.e., impression consistency). To situate the magnitude of reliability, we compared novel objects to faces—familiar and extensively studied stimuli. Impression consistency was always present for both types of stimuli and comparable across all evaluations. Second, and more importantly, we tested how similar impressions are across people (i.e., impression consensus). Impression consensus was always present for faces, but not always for novel objects. In Study 2 we examined a greater diversity of shapes and replicated the findings of Study 1 for novel objects. The findings suggest that impression consensus for novel objects only emerges when certain types of shapes and evaluations map together. When such mapping is possible, impressions are isomorphic with the parametrized shapes.

General introduction

People constantly encounter objects they have not seen before. Despite the frequency of these events, it remains unclear how people form impressions of novel objects. Whether it be window-shopping or gallivanting through a museum without much thought, people can easily dismiss most objects in view and gravitate to just a few. One possibility is that the shape of these objects affects their nascent impression. From prior research on perceptual categorization, it is known that people are meaningfully sensitive to nuanced shape differences (e.g., Folstein, Palmeri, & Gauthier, 2013; Freedman,

Riesenhuber, Poggio, & Miller, 2003; Goldstone & Styvers, 2001). Yet few publications have directly examined the effects of shapes on impressions, and even fewer using novel and abstract objects: rectangular width-to-length ratio (e.g., McManus, 1980), radial frequency (Chen, Huang, Woods, & Spence, 2016), and curviness (Amir, Biederman, & Hayworth, 2011; Bar & Neta, 2006; Bertamini, Palumbo, Gheorghes, & Galatsidas, 2016; Gómez-Puerto, Munar, & Nadal, 2016). The limited differences in which shapes have been explored in prior research could be due to two reasons: a lack of algorithms for systematically manipulating the shape of novel objects, or difficulty defining meaningful metrics by which to manipulate shapes (Phillips, Norman, & Beers, 2010; Wilkinson, Wilson, & Habak, 1998). For example, curviness is sometimes operationalized as being the opposite of straight, whereas other times it is the opposite of sharp or angular (Gómez-Puerto et al., 2016). Cognizant of these issues, we introduce a new technique for generating stimuli and revisit the general question of whether shapes influence impressions of an object without preconceptions of what constitutes meaningful shape differences.

Studying the effects of an object's shape on impressions is difficult because the moment a person knows that an object is present, semantic information is already available (Grill-Spector & Kanwisher, 2005). To focus on the possible effects of shapes themselves, we test abstract objects, because any associations to them are presumably random across the population. Crucially, these objects are unique in that they also maintain parametric relationships across their differing shapes. Thus, if shapes have a systematic and knowable role in the formation of a novel object's impression, we should be able to observe shape differences being isomorphic with impression differences.

We test the reliability of shape's effect on impressions by analyzing impression consistency and impression consensus. Impression consistency (i.e., intrarater agreement, test–retest reliability) signifies the extent to which individuals form the same impression upon re-

Citation: Kurosu, A., & Todorov, A. (2017). The shape of novel objects contributes to shared impressions. *Journal of Vision*, 17(13):14, 1–20, doi:10.1167/17.13.14.

doi: 10.1167/17.13.14

Received March 14, 2017; published November 30, 2017

ISSN 1534-7362 Copyright 2017 The Authors



This work is licensed under a Creative Commons Attribution 4.0 International License.

evaluation. Impression consensus (i.e., cross-observer agreement, interrater reliability) signifies the extent to which impressions are similar across people. Statistically, impression consistency is a precondition for impression consensus. Conceptually, whereas impression consistency captures the reliability of the subjective experience of stimuli, impression consensus captures the reliability of the shared experience across the population. Consistency and consensus can be used to model the contributions of subjective (i.e., private taste, individual preference) and objective (i.e., shared taste, common preference) evaluations to impressions (Germiné et al., 2015; Hönekopp, 2006).

To produce our stimuli, we manipulated 3-D shapes through parametric algorithms using a method called procedural generation (for an overview, see Krispel, Schinko, & Ullrich, 2015). A single parametric algorithm generates a shape family of objects per inputted values. The resulting objects share relatable shapes. Just as genetic siblings appear relatively similar and yet different, objects belonging to the same shape family appear similar and yet different. While there have been previously published novel object families, such as “greebles” (Gauthier, Williams, Tarr, & Tanaka, 1998) and “fribbles” (Williams & Simons, 2000), the shape families we introduce here are different in that they can be defined within a continuous parametric space. These objects therefore afford a unique opportunity to test whether the impressions of complex novel objects are predictable given the parametric values defining their shape.

We also test impressions of faces to situate the reliability of object impressions. People form consistent impressions of faces, and different people largely agree (i.e., consensus) on these impressions (for a review, see Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). First impressions of faces can be codified in terms of how trustworthy (i.e., valence of evaluation) and dominant they seem (Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008), as well as how attractive they are (Sutherland et al., 2013). Evaluations of trustworthiness and dominance are related to judgments of both approachability and danger (Todorov et al., 2008; Todorov & Duchaine, 2008).

In contrast to the literature on impressions of faces, the literature on impressions of novel objects is fragmented and unclear. For example, Duckworth, Bargh, Garcia, and Chaiken (2002) have observed that people consistently react to approach or avoid novel stimuli. However, recent studies have not observed strong impression consensus among preferences for novel and abstract images (Rodway, Kirkham, Schepman, Lambert, & Locke, 2016; Vessel & Rubin, 2010). Taken together, prior literature makes it uncertain whether shapes reliably affect impressions of novel

objects. We build upon prior research by measuring four impressions that are applicable to both faces and novel objects: approachable, beautiful, dangerous, and likable. Approachable and dangerous relate to approach and avoidance behavior. Beautiful and likable relate to aesthetics and preferences.

In Study 1, we first verify the possibility that different people can independently come to share the same impressions of novel objects. We do so by probing the reliability of impressions in terms of consistency and consensus. Conceptually, this phenomenon would be similar to when random people share impressions of strangers. Therefore, pictures of faces were included in the design of Study 1 to compare the reliability of impressions from novel objects versus people. After verifying that people can share impressions of novel objects, it was possible to test whether the objects’ shapes are a contributing factor. To the extent that impression consensus forms, we expected the underlying shape parameters to be predictive of impressions. The results of Study 1 confirmed that shapes affect impressions. However, a greater variety of shapes should have been tested. Based on Study 1 alone, it could not be ascertained whether impression consensus emerges for only some evaluations, like dangerousness, or emerges when specific shapes map to specific evaluations. In Study 2, we therefore included two additional shape families and focused on impressions of beauty and dangerousness, because they yielded the lowest and highest levels of impression consensus in Study 1. We observed that impression consensus varied as a function of both the type of evaluation and the type of shape family. Together, these studies demonstrate that people can share impressions of objects they have not seen before when the shape of the object maps to a specific evaluation.

Study 1

Introduction

We started with a single family of novel objects, referred to as *family x*. First we tested whether participants would evaluate the novel stimuli consistently across time. To assess this, we asked participants to report their impression of each stimulus 10 times (the stimuli were presented in 10 different blocks). If participants are consistent, their evaluations will be positively correlated; moreover, these correlations should be detectable in the first two blocks of evaluation. We also expected that the first evaluations would predict the last evaluations. To the extent that these hypotheses are confirmed, the consistency in evaluation could be attributed to the effect of shapes on

first impressions rather than to an effect of learning the shapes over the course of the experiment. Second, we tested whether impression consensus would emerge across the four types of evaluation: approachability, beauty, dangerousness, and likability. We compared impressions of novel objects to impressions of faces. Finally, we expected that the consensus among impressions of novel objects would be isomorphic with the underlying shape parameters, which relate each object from *family x*. In other words, we sought to demonstrate that different people can come to share the same impression of a novel abstract object because of its shape.

Method

Participants

One hundred eighty-eight volunteers from Princeton University and the neighboring community participated in this study. Volunteers from the community were monetarily compensated, while students were compensated with course credit. Among the objects, 25 participants were assigned to each of four evaluations. Among the faces, the participants were assigned to different evaluations as follows: approachable = 27, beautiful = 19, dangerous = 21, likable = 21; the uneven assignment was due to a notation error in the random-assignment process. All participants had normal or corrected-to-normal vision. None of the participants had seen the stimuli prior to their participation.

Object stimuli

The set of novel objects used in this study, *family x*, were procedurally generated using an algorithm housing two parameters we refer to as *points* and *flux*. These particular shape parameters were selected because the resulting objects did not appear like anything in particular (we later verified the assumption of abstractness and novelty; see the Method section of Study 2). Within this algorithm, points dictate the number of distal projections (3–100) and flux dictates the surface acceleration around these projections (ranging from *quick*, $1/x^{1.00}$, to *slow*, $1/x^{4.00}$). Sixty-six unique objects were sampled from this family by assigning random values to its parameters; Figure 1 plots the objects along its parametric values. For further details, refer to the algorithm provided on our lab's website (www.tlab.princeton.edu/databases), where 3-D object files and 2-D renderings are also made available. The algorithm for *family x* was written based on a code called Spatial Deform by Michael Pryor (<http://www.3d-dreaming.com>). These objects were generated using Grasshopper version .9 (<http://www.grasshopper3d.com>) and then

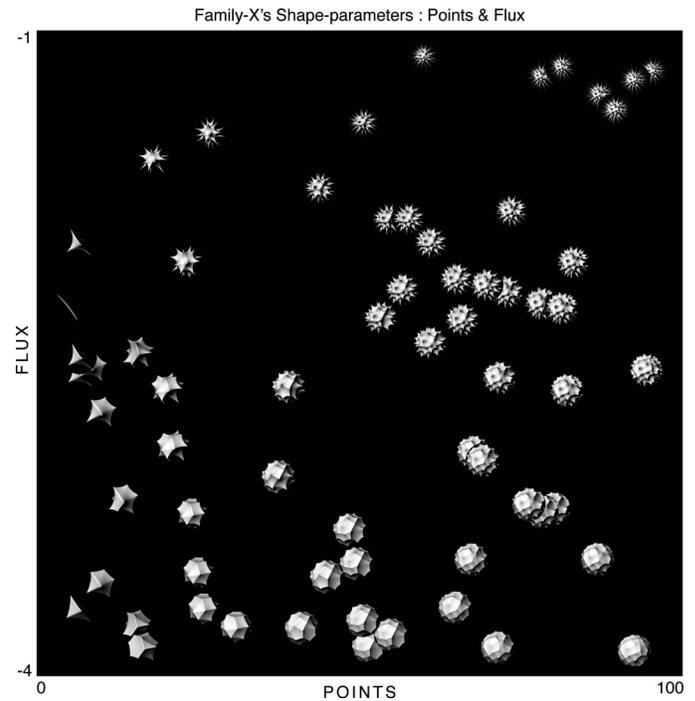


Figure 1. Objects sampled from *family x*. The stimuli consisted of 66 objects sampled from *family x*. The shape of these objects is defined by the number of distal points (points) and the surface acceleration around these points (flux).

rendered as 2-D images (800×800 pixels) using Rhino version 5 (<http://www.rhino3d.com>).

All rendering qualities were set constant across all images using the default rendering engine in Rhino. The objects appeared neutral gray on a black background. This appearance was achieved using the default settings of the program. The default shader (i.e., the specifications for simulating a material visually) was white in color with 0% gloss, 0% reflectivity, 0% transparency, and 1.0 index of refraction. The shader specified no textures. The emission and ambient color were black. Diffuse lighting was enabled; a directional light was positioned above and slightly to the left of the object emitting parallel rays toward the object.

Face stimuli

The face stimuli consisted of 66 images from the Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Öhman, 1998). These images are cropped portraits (562×762 pixels) of nonfamous White individuals photographed in front of a gray background with a neutral facial expression.

Procedure

Participants sat in a private sound-isolated cubicle with a computer. In a 2×4 between-subjects design,

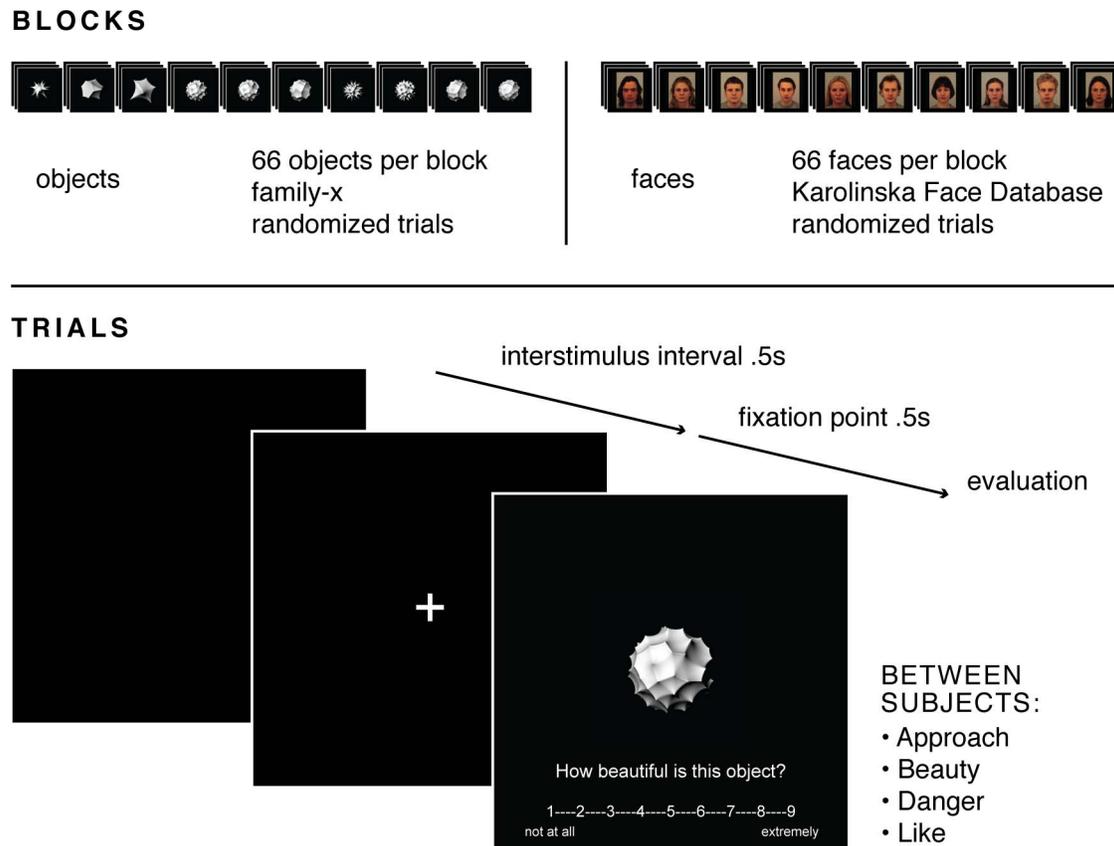


Figure 2. Study 1 experimental design. Participants evaluated 66 stimuli (objects or faces) 10 times using a scale from 1 through 9. Participants made evaluations of how approachable, beautiful, dangerous, or likable the stimuli seem.

participants evaluated either objects or faces with one of four types of impression (approachable, beautiful, dangerous, or likable). For example, one participant would answer the question “How approachable is this object?” throughout their session.

In each trial, a fixation point appeared for 500 ms in the center of the screen. This fixation point was replaced by a randomly selected object or face, the evaluation, and a scale ranging from 1 (not at all) to 9 (extremely). Participants were encouraged to go with their gut feeling when keying in their impression. There were no time constraints; objects remained visible until an evaluation was made. Afterwards, a black screen appeared for 500 ms. After each image was presented once, constituting a block, participants had an opportunity to take a break. There were 10 blocks in this study and it took approximately 30 min to complete; see Figure 2 for a visualization of this procedure. All stimuli were centered upon a black background and presented at a 20° visual angle using CRT monitors. This procedure was coded using PsychoPy version 1.82 (Peirce, 2007).

Main analyses

Consistency: Each item was presented 10 times to test how reliable impressions are across time. The impression consistency was determined by calculating an average Pearson’s r correlation coefficient per participant. For example, Participant 1’s impressions from the first block are correlated with the second block, the second block with the third block, and so on, all possible pairings being correlated together. The resulting correlation coefficients were averaged together to generate a single value representing that individual’s impression consistency. We then calculated t tests against zero to assess whether the level of consistency was significantly greater than zero. Significance testing was calculated using Fisher’s r to z transformed scores.

Because participants rated the objects multiple times, it is possible that our measure of impression consistency is driven by the effect of learning the stimuli and recalling a former impression. To assess whether first impressions are truly meaningful and impression consistency is not merely an artifact of learning, we report the correlation between first impressions (in the very first block of stimulus presentation) and every subsequent impression of the same item. Specifically,

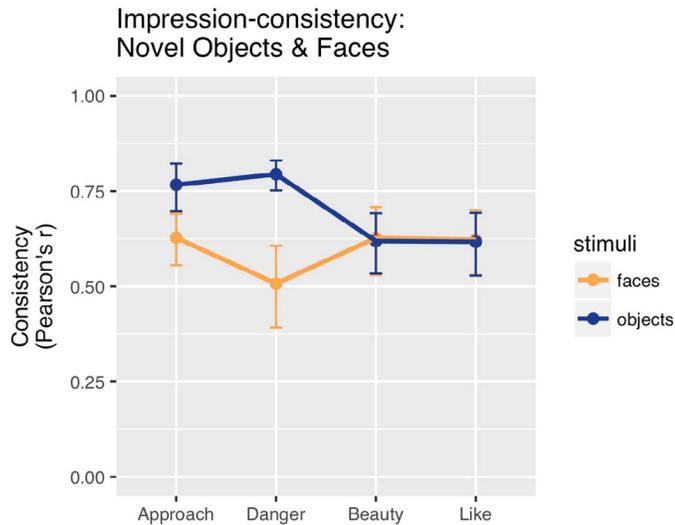


Figure 3. Impression consistency. The level of consistency among object impressions was equal to or greater than that for face impressions. The means are plotted with standard-error bars representing the 95% confidence intervals.

not only should impressions among the different blocks be highly correlated, but these correlations should be detectable between the very first two blocks, as well as between the first and the last block.

Consensus: Measures of consensus were calculated using a leave-one-out correlation with the mean statistic (e.g., Engell, Haxby, & Todorov, 2007; Germine et al., 2015). First, the averaged impression reflective of one participant is correlated with the averaged impressions representative of all other participants. Repeating this process for each participant generates correlation coefficients representing the similarity of one participant's impressions with those of the entire sample—the greater the correlation, the greater the level of consensus. We also calculated consensus using another method in which we computed pair-wise correlations among impressions formed by each participant. Since this second method did not allow for certain tests of significance, we report the results of this alternative method and the results of various significant tests afforded by the different methods in the Supplementary Appendix. Fisher's r to z transformed scores were used in t tests against zero to determine whether the consensus values observed were significant. We also calculated measures of consensus for the first block only and used the same tests, as already outlined, for significance.

Analyzing shapes: To the extent that there is impression consensus, we hypothesized that it would occur in a systematic fashion attributable to the constituent shape parameters. We therefore calculated linear mixed-effects regressions to test the extent to which shape parameters were significantly predictive of impressions. The shape parameters represent the values used to

generate an object. For example, 40 points in *family x* represents 40 distal extensions of mass in 3-D. Not all 40 points would be visible in a 2-D rendering, but since the renderer's settings are standardized across objects, a proportionate number of points are represented across renderings. These parametric values were then centralized and standardized per family before being submitted to the regression models. The shape parameters were tested as fixed effects and the participant differences were accounted for as a general random effect upon the model's main intercept. In each model, we focused on the proportion of variance, R^2 , explained by the fixed factors using the method by Nakagawa and Schielzeth (2013). A higher proportion of variance attributable to the fixed effects strengthens our evidence that shapes contribute to nascent impressions of objects.

Note regarding plots

To plot means and standard errors, the statistics were first calculated using the z scores and then transformed back into r coefficients. The purpose of doing this was to illustrate the contrasts between different variables while conveying the relative strength of each correlation. However, outside of these plots, untransformed r coefficients are reported because they are more conservative. Original and transformed values of impression consistency and impression consensus are provided in the Supplementary Appendix.

Results

Consistency

Individuals experienced a similar impression every time they were re-exposed to the same item. We first collapsed the level of consistency across the different types of evaluations and calculated separate t tests against zero for objects and faces. The average impression consistency across 10 blocked exposures was significantly correlated for both objects, $t(99) = 24.11$, $p < 0.001$, $d = 2.41$, $r_M = 0.66$, and faces, $t(87) = 20.13$, $p < 0.001$, $d = 2.15$, $r_M = 0.56$. There were significant differences, however, in the level of impression consistency across conditions; see Figure 3. A between-subjects ANOVA revealed a significant interaction between the type of evaluation and stimuli upon the level of consistency, $F(3, 180) = 7.06$, $p < 0.001$, $\eta_p^2 = 0.11$, $\eta^2 = 0.09$. Impression consistency significantly differed by evaluation type for novel objects, $F(3, 96) = 8.18$, $p < 0.001$, $\eta^2 = 0.20$, but not for faces, $F(3, 84) = 1.61$, $p = 0.19$, $\eta^2 = 0.05$. Regardless of the differences, significant levels of impression consistency were found in every condition.

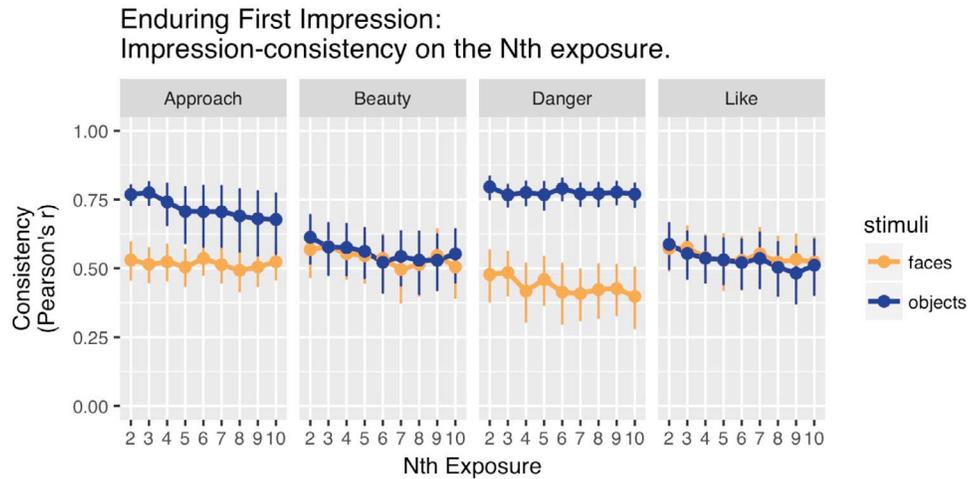


Figure 4. First impressions correlated with the impression on the *N*th exposure. The first impressions of faces and novel objects were correlated with their impressions from each subsequent re-evaluation—for example, first impressions were consistent with 10th impressions.

Were the first impressions meaningful? As shown in Figure 4, the consistency of later impressions with first impressions remained stable, although in some cases it did decrease. Examining the correlation between first and second impressions, consistency was significant for faces, $t(87) = 20.70, p < 0.001, d = 2.21, r_M = 0.51$, and objects, $t(99) = 25.19, p < 0.001, d = 2.52, r_M = 0.66$. Moreover, the correlation between the first and 10th impressions was also significant for faces, $t(99) = 17.62, p < 0.001, d = 1.88, r_M = 0.49$, and objects, $t(99) =$

$18.25, p < 0.001, d = 1.83, r_M = 0.63$. These results support our hypothesis that engaging in the evaluation of novel objects is nonarbitrary for an individual insofar as it elicits consistent impressions. Hence, to estimate the reliability of an individual’s impressions, it appears that including just one subsequent repetition is sufficient.

Consensus

Overall, people appeared to share similar impressions. Using *t* tests against zero, significant impression consensus was present for both objects, $t(99) = 10.48, p < 0.001, d = 1.05, r_M = 0.55$, and faces, $t(87) = 27.46, p < 0.001, d = 2.93, r_M = 0.63$. However, closer inspection revealed large variation in the level of impression consensus; see Figure 5. Impression consensus for novel objects was found for only two types of evaluation: approachable and dangerous. This was not the case for faces; impression consensus was found for all four types of evaluation.

Analyzing the different types of stimuli and evaluations together with ANOVA revealed a significant main effect of the type of stimuli, $F(1, 180) = 12.80, p < 0.001, \eta_p^2 = 0.07, \eta^2 = 0.03$; the type of evaluation, $F(3, 180) = 40.42, p < 0.001, \eta_p^2 = 0.41, \eta^2 = 0.29$; and the interaction between different evaluation types and stimuli, $F(3, 180) = 40.69, p < 0.001, \eta_p^2 = 0.40, \eta^2 = 0.28$. While faces elicited a reliably significant level of impression consensus across the different evaluation types, novel objects elicited highly variable levels of impression consensus.

The highest level of consensus occurred among object impressions of danger, $r_M = 0.94$, and approach, $r_M = 0.85$. The difference between these two types of

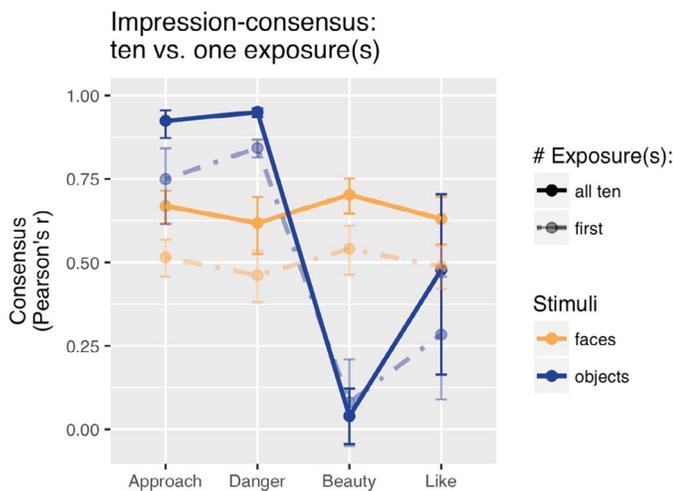


Figure 5. Impression consensus. The extent to which impressions are shared varies depending on the type of evaluation and stimuli. The mean consensus values are plotted along with standard-error bars representing the 95% confidence intervals. Averaging impressions collected across 10 exposures (solid lines) results in a higher level of impression consensus than from first impressions alone (dashed lines), yet the pattern of differences between conditions is unaffected.

evaluation was not significant, $p = 0.66$. Following suit was the magnitude of consensus among the face impressions: beauty, $r_M = 0.68$; approach, $r_M = 0.65$; likability, $r_M = 0.60$; danger, $r_M = 0.58$. These differences were not significantly different from each other, $F(3, 84) = 1.32$, $p = 0.27$, $\eta_p^2 = 0.04$, $\eta^2 = 0.04$. The next level of consensus occurred among the likability impressions of objects, $r_M = 0.38$, which was still significantly greater than zero, $t(24) = 2.87$, $p < 0.01$, $d = 0.57$. Finally, consensus was not observed for beauty impressions of novel objects, $t(24) = 0.93$, $p = 0.36$, $d = 0.19$, $r_M = 0.03$. Multiple comparisons between conditions are reported in the Supplementary Appendix.

The prior results show that impressions are consistent and can be shared, so we should expect the very first impressions to be shared too. We examined impressions made during the first exposure and found the resulting pattern of significance to be the same; see Figure 5. Using t tests against zero, we found significant consensus for object impressions of danger, $t(24) = 26.53$, $p < 0.001$, $d = 5.31$, $r_M = 0.83$; approach, $t(24) = 7.45$, $p < 0.001$, $d = 1.49$, $r_M = 0.67$; and likability, $t(24) = 2.83$, $p < 0.01$, $d = 0.57$, $r_M = 0.26$. There was also significant consensus for every type of face impression—beauty: $t(18) = 11.39$, $p < 0.001$, $d = 2.61$, $r_M = 0.52$; approach: $t(26) = 14.90$, $p < 0.001$, $d = 2.87$, $r_M = 0.50$; likability: $t(20) = 12.45$, $p < 0.001$, $d = 2.72$, $r_M = 0.48$; danger: $t(20) = 10.00$, $p < 0.001$, $d = 2.18$, $r_M = 0.44$. Again, only the object impressions of beauty were not significantly shared, $t(24) = 1.21$, $p = 0.24$, $d = 0.24$, $r_M = 0.08$.

Shapes

With the presence of consensus, it is possible to meaningfully estimate the extent to which shape parameters contribute to shared impressions of novel objects. Using linear mixed-effects regressions, we regressed the impressions of objects upon their shapes for the three evaluations which showed consensus: danger, approach, and likability. Dangerous evaluations of *family x* yielded the greatest consensus and, consequently, the shape parameters accounted for a high proportion of the variance, $R^2_{\text{fixed}} = 0.66$, $R^2_{\text{total}} = 0.86$. As points and flux increased in value, objects from *family x* appeared more dangerous: $\beta_{\text{points}} = 0.41$, $p_{\text{points}} < 0.001$; $\beta_{\text{flux}} = 1.60$, $p_{\text{flux}} < 0.001$; $\beta_{\text{points:flux}} = 0.29$, $p_{\text{points:flux}} < 0.001$. Similarly, the shape parameters predicted approach evaluations, $R^2_{\text{fixed}} = 0.54$, $R^2_{\text{total}} = 0.74$, and likability evaluations, $R^2_{\text{fixed}} = 0.10$, $R^2_{\text{total}} = 0.18$. Thus, in support of our hypothesis, we observed shapes holding a systematic and isomorphic relationship with impressions of novel objects when there was consensus.

Discussion

Our results demonstrate evidence of shapes' contribution to the nascent impressions people have of novel objects. In the absence of additional information, it appears that first impressions, driven by the shape of a novel object, are lasting impressions. This result is congruent with an observation made by McManus (1980), who observed people evaluating rectangular shapes consistently 2 years later. Another result we observed was that these impressions were sometimes shared across people—that is, impression consensus for novel objects is possible. To the extent that consensus was present across people, their impressions were indeed isomorphic with the underlying parameters that define the shape of each object. What this means is that people are able to share impressions of novel objects, at least in part, because of their shape.

Across our studies, we did observe a few participants with low levels of impression consistency. Low consistency could be interpreted to mean that the task of evaluating novel objects is not meaningfully engaging, or that participants were not paying proper attention in the experiment. Because one of the aims of this study was to assess the extent to which evaluating novel objects constitutes meaningful behavior, no participants were excluded—even if they demonstrated low levels of impression consistency. Our estimates are therefore conservative. The results of our study suggest that people generally find the task of evaluating novel objects to be meaningfully engaging.

We found that people generally agreed on which novel objects were more or less dangerous, but not which were more or less beautiful. This pattern is somewhat congruent with prior literature examining abstract novel objects, wherein consensus occurred among approach/avoid behavior (Duckworth et al., 2002) but not among preferences (Vessel & Rubin, 2010). However, interaction effects of objects and evaluations have been observed before. For example, different rectangular proportions are preferred depending on the type of evaluation solicited (Hekkert, Peper, & van Wieringen, 1994; Russell, 2000). It is therefore possible that certain types of shapes have more relevance for certain types of evaluations.

The novel objects we tested were all generated from a single shape family, *family x*. The underlying procedural generation algorithm describing *family x* was designed only with the intention of generating objects not resembling anything in particular—including standard geometric shapes or *geons* (Biederman, 1987). A posteriori we learned that *family x* bears similitude to the radial frequency (RF) algorithm used to investigate the *bouba-kiki* (formerly *maluma-takete*; Holland & Wertheimer, 1964) phenomenon (Chen et al., 2016). The gist of this phenomenon is that there exists

consensus from some objects generated from RF matching the sound “bouba” and others matching the sound “kiki.” In the RF formulation, three shape parameters (amplitude, frequency, spikiness) distort a 2-D outline of a circle. RF frequency has effects comparable to *family x*’s points—it varies the number of distal extensions. RF amplitude and spikiness together result in effects comparable to *family x*’s flux. Between the two shape families, there are two major differences: RF features radial symmetry, while *family x* does not, because its distally extending points are positioned randomly; and RF generates 2-D line drawings, while *family x* generates 3-D volumes.

The design of the RF algorithm was inspired by prior research (i.e., Gallant, Braun, & Van Essen, 1993; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996) demonstrating neuronal tuning in V4 to parametrically manipulated polar, hyperbolic, and Cartesian gratings (Wilkinson et al., 1998). Considering the similarity of *family x* to RF and its precursor (i.e., polar, hyperbolic, and Cartesian gratings), it could be inferred that the reliability of impressions elicited by the objects generated from *family x* can be explained by the same neural mechanisms—namely, consensus regarding which objects from *family x* seem more or less dangerous, approachable, beautiful, or likable could rely on neuronal tuning at V4.

Nevertheless, it is possible that the underlying shape parameters of *family x* represent a particular set of shapes that exert unique effects upon impressions. Since the primary aim of this article is to document the reliability of a general effect that shapes may have on impressions, it was therefore necessary to examine more types of shapes. Through comparing a larger number of shape families, it is also possible to examine the extent to which impressions of novel objects are shared for only specific evaluative terms (e.g., dangerous) and always idiosyncratic for others (e.g., beauty). Study 1 provided initial evidence of shape effects on impressions. In the subsequent study, we employed a greater diversity of shapes to test whether consensus forms for other types of shapes and whether the level of consensus would vary by shape family for the same type of evaluation.

Study 2

Introduction

In this study, we sought to explicate shape effects upon nascent impressions by including two additional shape families: *family y* and *family z*. We focused on two evaluations, dangerous and beautiful, because they represented the highest and lowest level of consensus

observed in Study 1. We hypothesized that consensus would form for other types of shapes and that it would result from an interaction between the type of shape and the type of evaluation solicited. Such a finding would further substantiate the claim that shapes contribute to impressions of novel objects by demonstrating that systematic shape effects are not solely specific to certain types of evaluations, nor are they unique to *family x*.

Additionally, we ran a separate online survey to check the ease, familiarity, and novelty of our task and its stimuli—three presumptions we made in Study 1. The ease with which an object can be perceived and associated with semantics relates to its perceptual and conceptual fluency (Reber, Schwarz, & Winkielman, 2004). Objects which are processed more fluently are evaluated more positively (Reber et al., 2004). Therefore, we checked the baseline fluency of each shape family using an online survey to determine how the perceived ease or difficulty of an evaluation relates to impressions. We also checked the perceived familiarity of our shape families, because fluency can also relate to how familiar something seems (Whittlesea, 1993). Finally, we asked people about the associations which came to their minds, in order to check the baseline novelty of each shape family.

The design of Study 2 is similar to that of Study 1, except that we reduced the number of repeated exposures to two, since consistency and consensus emerged in the first two blocks of evaluation in Study 1. This alteration allows us to accommodate a larger set of novel objects without increasing the length of the study.

Study 2 was also designed to confirm the results of Study 1 in two ways. First, the initial two blocks of both studies were the same. Second, we took a portion of the objects from *family x* and re-rendered them using different settings to test whether impressions are consistent despite changes in low-level visual qualities between the different renderings of the same object. Finally, we included a question at the end of the experiment gauging how likely participants believed it to be that others would agree with their own impressions. This final question was included to test whether people’s intuitions about the similarity of their impressions to others’ are accurate.

Method

Participants

Using the same sampling methods as in Study 1, 47 volunteers were recruited for Study 2. None of the participants had seen any of the objects before. Nine participants were excluded from the analyses because their evaluative ratings lacked variance across entire shape families. In the final analyses, 20 participants

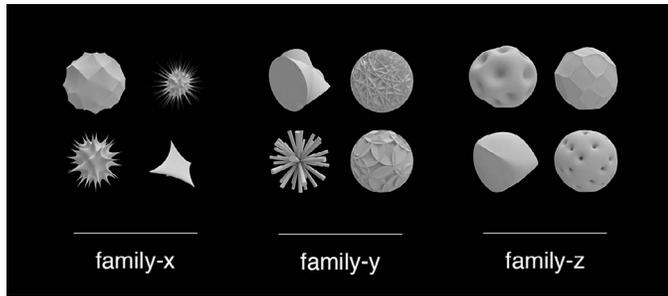


Figure 6. Objects generated from three shape families. Thirty objects were sampled from *family x*, *family y*, and *family z*. Each family defines a morphological relationship among its objects. The *family x* here is the same shape family used in Study 1.

made beauty evaluations and 18 made danger evaluations.

Stimuli

One hundred twenty-six different objects were sampled from *family x*, *family y*, and *family z*. Figure 6 depicts a sampling of objects. As in Study 1, each shape family was designed so as not to appear like anything in particular. Both *family x* and *family z* differ on two parameters: points and flux. The difference between *family x* and *family z* is in how the distal points are expressed. In *family x*, surfaces are produced which push points distally, while in *family z*, surfaces are produced which pull points proximally. There are four parameters incorporated in *family y*: points, flux, length, and radius. Points dictates the number of theoretically distal points. Flux, in this case, shifts the points slightly. Length dictates the magnitude by which the points are projected relative to its volume. Radius dictates the size of a cylindrical edge formed at each distal point.

One hundred fifty-six images were used in this study: 96 images of 66 objects sampled from *family x* and 30 images of 30 objects sampled from both *family y* and *family z*. The reason for the greater number of images than objects from *family x* is because we rerendered the objects. The purpose of doing that was to check whether the impressions were attributable to perceptual properties of the 2-D renderings rather than the morphological properties of the 3-D objects. Figure 7 exemplifies the differences between the renderings.

The new renderings in this study were created using V-Ray (version 2; www.chaosgroup.com). Rhino's rendering engine, used in Study 1, does not share the specific bidirectional reflection distribution function of the default rendering engine. This function describes the ratio of a surface's radiance, captured by a renderer's camera, to the surface's irradiance brought on by a renderer's light source (Bartell, Dereniak, & Wolfe, 1981). Due to the lack of technical specification,

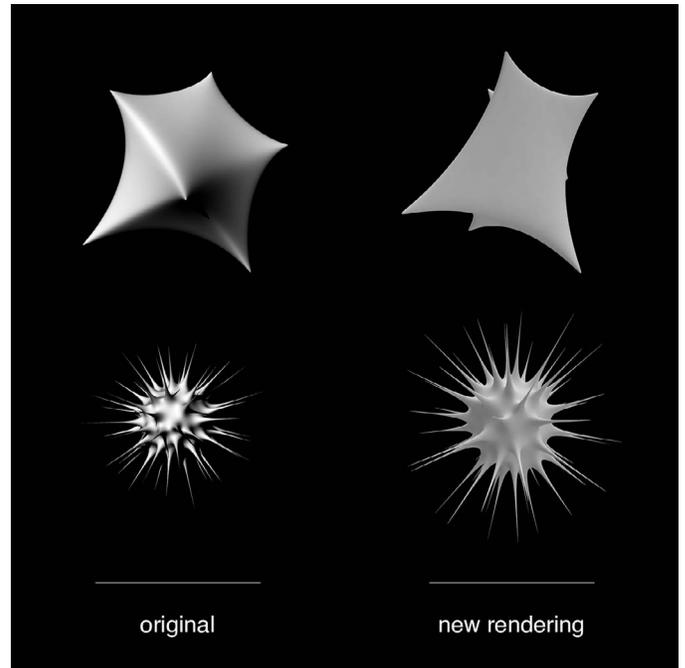


Figure 7. Objects rerendered. Of the original 66 objects sampled from *family x*, 30 were rendered as new images. Horizontally, these are the same objects; vertically, they are different renders. Despite differences between renderings, impressions of each object were significantly correlated.

we moved to using V-Ray. V-Ray also simulates more physically correct illumination of 3-D surfaces than the renderer used in Study 1. The diffuse color of the shader was gray (R = 190, G = 190, B = 190). Similar to the previous settings, the refraction and reflection were set to off, and the index of refraction was 1.0. The bidirectional reflection distribution function can be specified using V-Ray, and was left using the default settings (type = Blinn, anisotropy = 0, rotation = 0, soften = 0, fix dark glossy edges = on, derivation = UVW). The rendering camera in V-Ray can also be specified: F-number = 8.0, ISO = 400; distortion = 0, type = still camera, shutter speed = 200, zoom factor = 1.0, lens shift = 0, vignetting = 0, exposure = on. The indirect illumination settings were left at their defaults; this included the primary light bounces being calculated using an irradiance map (min rate = -3, max rate = -2, color threshold = 0.4, normal threshold = 0.3, distance threshold = 0.1, hemispheric subdivs = 50, interpolation samples = 20, interpolation frame = 2, interpolation type = least squares fit, sample look up = density-based, calc. pass interpolation samples = 15, multipass = on, randomized samples = on, check sample visibility = off) and the secondary light bounces being calculated using a brute-force method (subdivs = 8, bounces = 3). Options for calculating caustics were unnecessary and therefore set to off. The default illumination was used, which includes a skylight that provides indirect

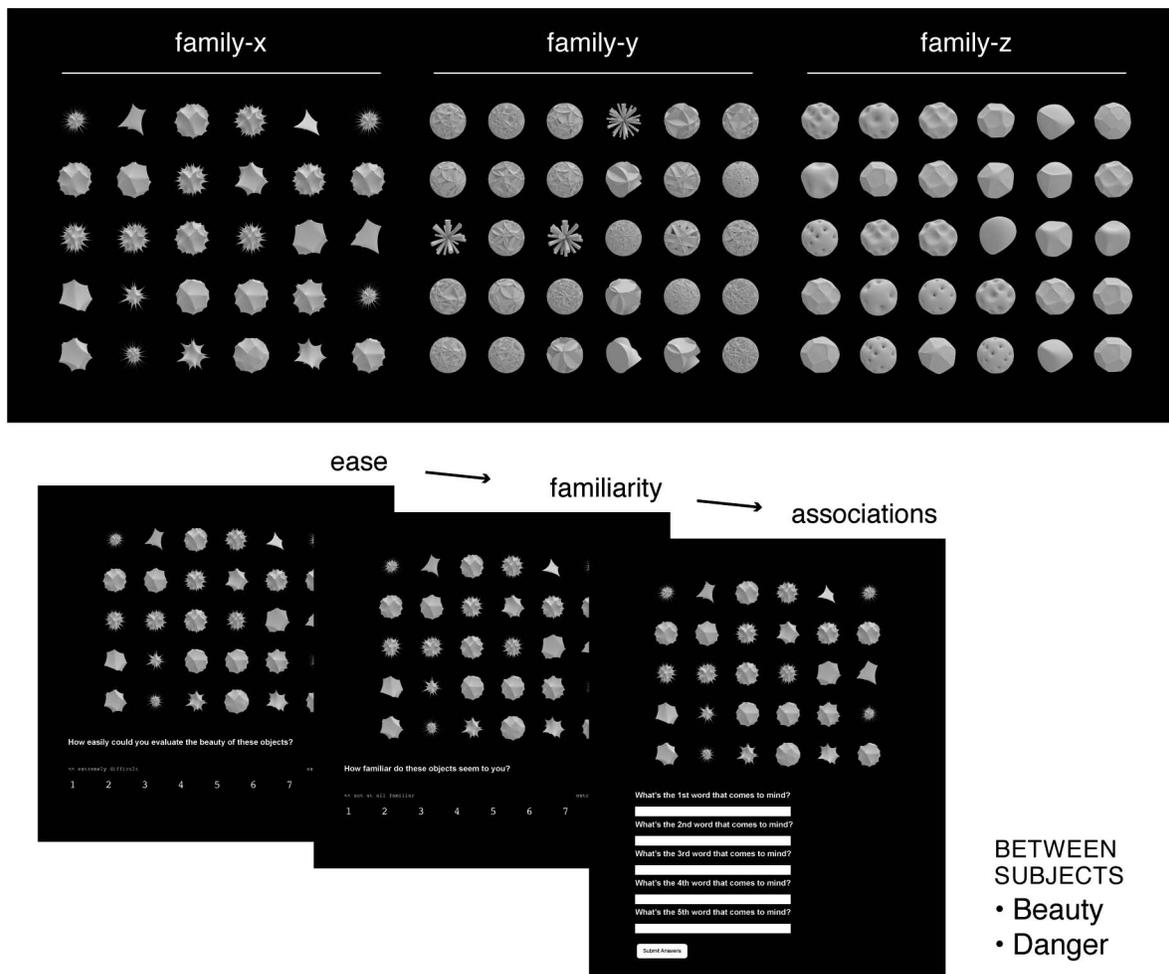


Figure 9. Manipulation check: Experimental design for checking ease, familiarity, and novelty. Shape families were presented as a grid. We checked each shape family for how difficult it was to evaluate, how familiar it seemed, and the diversity of associations it elicited.

Manipulation checks

Replicating Study 1: Study 2's experimental design replicates part of Study 1. The original stimuli were presented in the first two blocks. Beauty and dangerous evaluations were examined again. Based on the results of Study 1, we expected there to be significant impression consistency for both evaluation types; however, only dangerous evaluations would result in significant consensus.

2-D images versus 3-D objects: Other variables aside from 3-D shapes (e.g., image composition, surface reflectance, memory for the stimuli) could explain the reliable impressions. To rule out these possibilities, 30 of the 66 objects from *family x* were re-rendered with a different size and surface quality (e.g., illumination, reflectance, shader) and at a different angle. If impressions remained consistent despite the visual differences between renderings, participants were likely responding to the morphological 3-D properties rather than to specific qualities found within each of the 2-D images. For each participant, we calculated the

Pearson's r correlation between their average impressions of both renderings. We report the average correlation, representing our measure of consistency across renders, across all the participants.

Checking ease, familiarity, and novelty: Three presumptions from Study 1 were checked using an online survey: Evaluating novel objects seems reasonable and not too difficult; the objects were novel insofar as they did not seem significantly familiar or unfamiliar to people; and the objects are novel insofar as different people have different associations which come to mind. Using Amazon Mechanical Turk, 60 participants were surveyed online. However, data from only 57 participants were used in the final analyses, because three participants saw the stimuli in another online experiment or had incomplete data. The stimuli consisted of the same new renderings used in Study 2. For each shape family, every object used in Study 2 was presented together in a 5×6 grid on a black background (see Figure 9). The placement of the objects was randomized across participants.

BETWEEN
SUBJECTS

- Beauty
- Danger

	<i>N</i>	Consistency (r_M)	Consensus (r_M)
Beauty			
Study 1	25	0.58	0.03
Study 2	20	0.55	0.14
Danger			
Study 1	25	0.77	0.94
Study 2	18	0.76	0.86

Table 1. Replication of beauty and danger evaluations of novel objects. *Notes:* The average Pearson's r correlation coefficient representing consistency and consensus is reported.

There were two evaluation conditions tested between participants. One half of the participants were asked “How easily could you evaluate the beauty of these objects?” and the other half were asked “How easily could you evaluate the danger of these objects?” Responses could range from 1 (extremely difficult) to 9 (extremely easily). Then the shape families were presented again, but this time with the question “How familiar do these objects seem to you?” Responses could range from 1 (not at all familiar) to 9 (extremely familiar). Finally, each shape family was presented a third time and participants were asked to list in order the first five associative words which came to mind. Each shape family remained visible until a response was recorded.

Results

Manipulation checks

Replicating Study 1: Confirming our findings from Study 1, consistency and consensus from beauty evaluations and danger evaluations did not meaningfully change; Table 1 lists the average r correlation values across studies. Consistency was not significantly different across studies for beauty evaluations, $t(40.28) = 0.45$, $p = 0.65$, or danger evaluations, $t(32.60) = 0.24$, $p = 0.82$. Consensus was not significantly different across studies for beauty evaluations, $t(33.16) = -1.49$, $p = 0.15$, but it was significantly lower in Study 2 for danger evaluations, $t(38.68) = 4.82$, $p < 0.001$, $d = 1.49$. However, despite the lower level of consensus in the replication, a t test against zero revealed that consensus for danger evaluations in Study 2 was nevertheless highly significant, $t(17) = 19.87$, $p < 0.001$, $d = 4.68$, $r_M = 0.86$.

2-D images versus 3-D objects: Despite changes between the two different renderings of the same 3-D objects, people formed consistent impressions. Correlating impressions of the old renderings with impressions of the new renderings, the level of consistency was significant for both beauty evaluations, $t(19) = 7.76$, $p < 0.001$, $d = 1.74$, $r_M = 0.55$, and danger evaluations, $t(17) = 15.10$, $p < 0.001$, $d = 3.56$, $r_M = 0.85$. These

results suggest that the impressions our participants reported were likely in response to the differences in shapes across the 3-D objects rather than to specific qualities of the 2-D images.

Ease: We first checked the baseline fluency of each condition by asking people to report the ease or difficulty of evaluating a shape family. We calculated planned comparisons for each shape family and evaluation type using t tests against the midpoint. Danger evaluations of *family x* were thought to be significantly easy, $t(28) = 5.07$, $p < 0.001$, $M = 6.55$, $d = 0.94$. Meanwhile, all other conditions were thought to be neither significantly easy nor difficult. The possible implications for the danger evaluations of *family x* are explored further in the Discussion section later.

Familiarity: A second check for imbalanced fluency was to measure the perceived familiarity of each shape family. We ran separate analyses to capture the potential priming effects of mentioning beauty or danger in the prior question posed to the participants. We again calculated planned comparisons using t tests against the midpoint. Participants in the danger condition found *family z* to be significantly familiar, $t(28) = 2.36$, $p = 0.03$, $M = 5.66$, $d = 0.44$. However, all other conditions were not significant. Thus, each shape family was again relatively similar in terms of fluency. This time, the one exception was those evaluating the danger of *family z*; the possible implications are also explored in the Discussion section later.

Novelty: Finally, we checked whether the different shape families indeed elicited diverse associations. Across our sample of 57 participants, we recorded 406 unique words in association with the three shape families. The most diverse associations were noted for *family y*, which elicited 187 different words. This was followed by *family x*, eliciting 170 words, and then *family z*, eliciting 151 different words. These numbers do not add up to 406 because a small subset of words were associated with all three shape families. Our presumption was confirmed—the objects are novel insofar as they elicit diverse associations across people. In fact, none of the shape families elicited the same association across all participants; see Table 2 for a list of the most frequently associated words.

Main analyses

Consistency: We again observed individuals experiencing their impressions consistently. Collapsing across the different types of evaluations and shape families, a t test against zero showed that the level of impression consistency was significant, $t(37) = 21.07$, $p < 0.001$, $d = 3.42$, $r_M = 0.80$. The level of consistency did not significantly differ between beauty evaluations and danger evaluations, $F(1, 36) = 3.10$, $p = 0.087$.

	family x		family y		family z	
	Word	%	Word	%	Word	%
Most frequent association	Sharp	8	Round	5	Round	6
	Pointy	4	Ball	4	Cheese	5
	Spiky	4	Moon	3	Rock	5
	Star	3	Rubber	2	Moon	4
	Spikes	2	Circle	2	Holes	3
Most frequent first association	Sharp	9	Round	9	Cheese	11
	Star	9	Rubber	9	Smooth	7
	Stars	7	Moon	5	Cookie	5
	Pointy	5	Paper	5	Rocks	5
	Spikes	5	Snow	5	Round	5
Most frequent association primed by beauty (N = 28)	Sharp	7	Round	7	Round	7
	Star	4	Circular	3	Cheese	5
	Pointy	4	Moon	3	Rock	4
	Spiky	3	Rubber	3	Holes	4
	Abstract	2	Ball	2	Moon	4
Most frequent association primed by danger (N = 29)	Sharp	8	Ball	6	Cheese	6
	Pointy	5	Paper	3	Rock	6
	Spiky	4	Round	3	Round	6
	Pain	3	Circle	3	Moon	5
	Spikes	3	Moon	3	Clay	3

Table 2. Most frequently associated words for each shape family. Notes: N = 57. Each participant wrote five associations; therefore, the maximum possible number of unique words is 285 for each shape family. Percentages were rounded up to the next integer.

Consensus: Crucially, we observed varying levels of consensus for both evaluations depending upon the shape family being evaluated. That is, the level of consensus among impressions is not dictated solely by the type of evaluation solicited. An ANOVA revealed a main effect of the evaluation type upon the level of consensus, $F(1, 36) = 25.28, p < 0.001, \eta^2 = 0.16$. However, as predicted, the effect of the evaluation type was qualified by an interaction between the evaluation type and shape family, $F(2, 72) = 29.42, p < 0.001, \eta^2 = 0.37$. There was no main effect of the shape family, $F(2, 72) = 2.56, p = 0.08, \eta^2 = 0.05$.

Given the significant interaction between the type of evaluation and shape family, we tested each condition for a significant level of consensus; Figure 10 plots the level of consensus for each experimental cell. Echoing the results of Study 1, family x again led to significant levels of impression consensus among danger, $t(17) = 17.19, p < 0.001, d = 4.05, r_M = 0.89$, and yet not among beauty, $t(19) = -0.21, p = 0.84, d = 0.05, r_M = -0.01$. What is new to Study 2 are the two shape families family y and family z. Both of these shape families yielded shared impressions regardless of the type of evaluation. For family y, the level of impression consensus was significant for both danger, $t(17) = 5.39, p < 0.001, d = 1.27, r_M = 0.53$, and beauty, $t(19) = 6.35, p < 0.001, d = 1.4, r_M = 0.60$. For family z, the level of

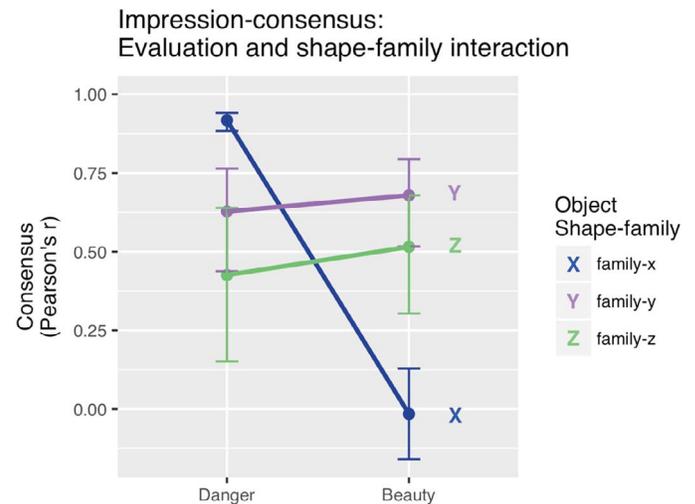


Figure 10. Significant interaction effects in assessing consensus. Impression consensus varies depending on the type of evaluation and shape. The mean consensus values are plotted along with standard-error bars representing the 95% confidence intervals.

impression consensus was lower, but nonetheless significant for both danger, $t(17) = 2.95, p = 0.009, d = 0.70, r_M = 0.31$, and beauty, $t(19) = 4.35, p < 0.001, d = 0.97, r_M = 0.42$. These results therefore show that other types of shapes, beyond those represented in family x, can therefore elicit shared impressions, and that the level of consensus varies across different types of evaluations and shapes.

Shapes: This study included a greater variety of shapes than Study 1 to further test whether differences among impressions and shape parameters are isomorphic to the extent that impression consensus emerges. Following Study 1, we computed a regression model for each shape family and each evaluation which resulted in significantly shared impressions. The greatest consensus occurred among danger impressions of family x and therefore, as predicted, the fixed effects (i.e., shape parameters) explained a large proportion of the variance, $R^2_{\text{fixed}} = 0.53, R^2_{\text{total}} = 0.80$. The trend between the level of consensus and the variance explained by the fixed effects continued with the impressions of beauty from family y, $R^2_{\text{fixed}} = 0.23, R^2_{\text{total}} = 0.61$; danger from family y, $R^2_{\text{fixed}} = 0.15, R^2_{\text{total}} = 0.61$; danger from family z, $R^2_{\text{fixed}} = 0.10, R^2_{\text{total}} = 0.50$; and beauty from family z, $R^2_{\text{fixed}} = 0.09, R^2_{\text{total}} = 0.68$. Thus, the effects shapes have on impressions of novel objects are readily apparent when the impressions are shared.

Intuitions about consensus: Upon finishing the experiment, we asked participants to assess the extent to which others would have similar impressions of novel objects. On average, participants believed they would share similar impressions with others, $M = 6.16$ (on a

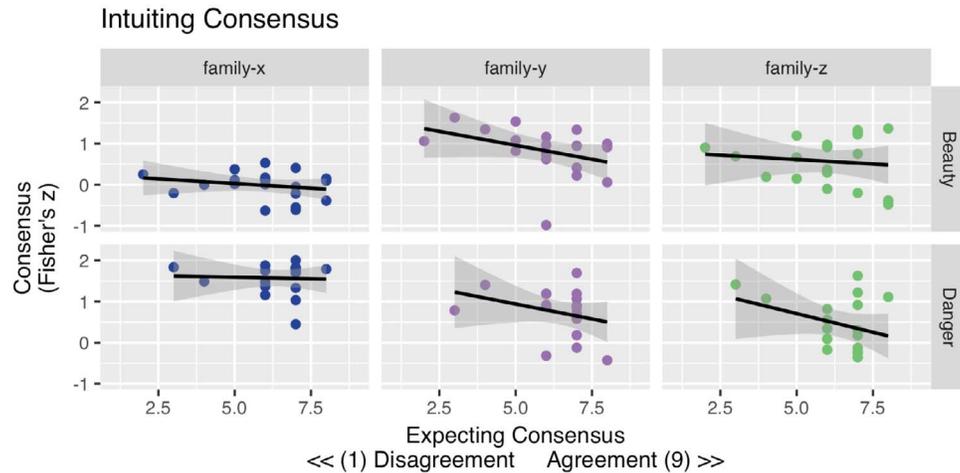


Figure 11. Intuiting consensus. People's intuitions about the degree to which others share their impressions are not necessarily accurate. Plotting the Fisher's r to z transformed scores against people's expectation for agreement among impressions, we found no significant relationship in any condition.

scale from 1 to 9). There was no significant difference between the different types of evaluations, $F(1, 36) = 0.90$, $p = 0.35$. Furthermore, for all conditions (i.e., evaluation type and shape family), there was no significant relationship between a participant's belief that their impressions were similar to others' and actually having similar impressions to others. We tested this by regressing the extent for which each participant contributed to consensus (i.e., correlation with the mean impressions of other participants) upon their intuition for consensus; see Figure 11 for a plot of these linear regressions. While evaluating how beautiful or dangerous something is, participants were inclined to believe that others would provide similar evaluations to their own. However, these intuitions were not accurate. That is, although we observed general consensus, participants overestimated the similarity their own impressions would have with those from others.

Discussion

Impression consensus for novel objects depends upon an interaction between the type of evaluation being solicited and type of shapes being observed. Although in Study 1 we observed impressions of danger to be highly similar across people, Study 2 showed that the extent to which these impressions are shared depends on the types of shapes being evaluated. Furthermore, the claim that some impressions (e.g., dangerous) are always less subjective than others (e.g., beautiful) is evidently falsifiable. Contrarily, we observed similar levels of consensus among beauty impressions and danger impressions with some shape families. A more appropriate claim would be that

different impressions vary in consensus depending upon the types of shapes being viewed.

Not only did the average level of impression consensus vary across conditions, but the standard error around the mean was also highly variable (see Figure 10). For example, danger evaluations of *family x* results in not only a higher level of impression consensus than danger evaluations of *family z* but also a smaller range of standard error. Because of the way in which impression consensus is calculated, the variability in standard error is attributable to participants reporting impressions that are in more or less accordance with the average (i.e., normative) impressions. It is possible that this method of calculating impression consensus using the correlation with the mean inflates the consensus relative to the pair-wise correlation method. For example, consensus among danger evaluations of *family z* results in a much weaker correlation using the pair-wise method, $r = 0.11$, than correlation with the mean, $r = 0.31$. We focused on the latter method because it allows for tests of significance. However, regardless of the method, the primary inference made in Study 2 remains intact: The levels of consensus are not solely dependent upon the type of evaluation, such as danger, but rather vary across different types of evaluations and shapes.

Among impressions of dangerousness, we observed the most consensus with *family x* and the least with *family z*. It is possible that this is related to the different baseline levels of perceptual fluency across these conditions. Recall the results of our manipulation checks: Danger evaluations of *family x* were found to be significantly easy, while *family z*, if primed by the word "danger," was found to be significantly familiar. Taken together, this could mean that the ease of evaluating the dangerousness of *family x* is related to

why people are likely to share impressions, while the perceived familiarity of *family z* is related to why people are not likely to share impressions. Nonetheless, what is important to note here is that there was variation in consensus across beauty evaluations of different shape families although they were balanced in terms of baseline perceived fluency. Therefore, our results show that perceived fluency alone cannot supplant our hypothesis that shapes contribute to shared impressions.

By testing impressions across multiple renderings of the same object, we further demonstrated the importance of shapes in the evaluation of novel objects. Initial impressions were unperturbed by changes in 2-D rendering qualities—for example, image size, surface shader, and viewing angle. The permanence of shape’s effect upon impressions may be supported by neurons selectively tuned to 3-D shapes over 2-D qualities (Yamane et al., 2008). It is plausible that shapes take precedence over other perceptual qualities in evaluating novel objects. Infants as young as 4.5 months have been found to individuate objects using their shape, months before the ability to individuate objects based on their patterns or colors is demonstrated (Wilcox, 1999). And it has been shown that infants as young as 6 months are already equipped with a region of their brain selectively activated to changes in shape (Emberson, Crosswhite, Richards, & Aslin, 2017).

At the end of the experiment, we assessed participants’ general intuitions about the degree to which others share their impressions. Participants on average expected to share similar evaluations with others. However, they overestimated the similarity of their own impressions to the impressions of others. Therefore, the expectation of sharing similar evaluations of novel objects with others constitutes a specific instance of the false-consensus effect—that is, participants overestimate the degree of similarity between themselves and others (Marks & Miller, 1987).

General discussion

In Study 1, we examined four types of evaluation (approachable, beautiful, dangerous, likable) and two types of stimuli (novel objects, faces). People proved to have consistent impressions across all types of evaluations and stimuli. Impression consensus, however, was not equally reliable between both types of stimuli. Faces always resulted in a reliably significant level of consensus, whereas novel objects resulted in varying levels of consensus, with only some types of impressions (dangerous and approachable) being evidently shared. Regardless, we observed evidence in support of our main hypothesis—that the shape parameters would

be isomorphic with shared impressions when there is impression consensus. In Study 2, we replicated the core findings of Study 1 while also testing a more diverse set of novel objects to show that impression consensus is not solely driven by specific types of shapes, such as *family x*, nor by a specific type of evaluation, such as dangerousness. The results suggest that shared impressions could occur for any novel object, depending upon its shape and the evaluation being solicited.

The question of why consensus is found only among specific combinations of shapes and evaluations remains an open one. Based on prior literature, there are many possible explanations. Here we offer three in no particular order: (i) Certain shapes have certain affordances, and impression consensus occurs when they assess a relevant affordance (for an overview of affordances, see Chemero, 2003). (ii) Some shape families, despite being abstract, bring to mind commonly shared semantic associations, and therefore certain types of evaluations will lead to consensus (e.g., Vessel & Rubin, 2010). However, in Study 2 we showed that people have different associations that come to mind for the present set of objects. Although this could serve as contradictory evidence, semantics could still offer an explanation if the diversity of associative words we recorded actually solicits the same semantic network. (iii) Based on the notion of micro-valence (generally defined as a “subtle affective valence” by Lebrecht, Bar, Barrett, & Tarr, 2012), each shape parameter and evaluative term (e.g., dangerous, beautiful) may be, to varying degrees, positive or negative (i.e., micro-valence). From this perspective, impression consensus may be explained by the likelihood of a similar mapping of valence for the evaluative term and shape parameters. This explanation, along with the others, may be tested in future research by leveraging procedural generation. For example, shape families could be designed to manipulate affordances, semantics, and valence. Studying such explanations could lead to an understanding of how all shapes may relate to one another and how impressions of novel objects may be codified.

The results of our studies follow what has long been evidenced in more applied realms of psychological research. In examining which shapes should symbolize aircraft in the military, rounded figures were more likely to be perceived, easily learned, and accurately recalled as friendly entities (Provins, Stockbridge, Forrest, & Anderson, 1957). Specific interactions between fonts and English words result in faster identification of meaning (Lewis & Walker, 1989). Trees and schematics similar in shape are evaluated similarly (Summit & Sommer, 1999). Curvilinear, over rectilinear, architecture is more likely to be evaluated as being beautiful (Vartanian et al., 2013). Products

advertised with a curvilinear rather than a rectangular logo are expected to be more comfortable (Jiang, Gorn, Galli, & Chattopadhyay, 2015). Our studies contribute to this literature by demonstrating the reliability of shape's effects through the manipulation of the most complex, yet controlled, novel objects to date.

Exploratory principal-components analysis

What does it mean when impression consensus is not found? We reported that there was no general consensus regarding the beauty of *family x*. Using exploratory principal-components analysis (PCA) with varimax rotation on the level of participants, we were able to test the extent to which beauty impressions were idiosyncratic. If completely idiosyncratic, then there should be an equivalent number of principal components as there were participants entered into the analysis. The results showed four components reliably emerging with an eigenvalue over 1. These four factors explained 87% of the total variance in Study 1 and 78% in Study 2. Applying a *k*-means clustering analysis, we found a high level of consensus among two clusters of participants within both Study 1 and Study 2. In other words, even if general consensus was not found, impressions were not completely idiosyncratic. Instead, there exist subgroups sharing similar evaluations. Of the total variance explained by these four-factor PCA models with two clusters, 54.7% of the variance can be explained by the shape parameters in Study 1 and 49.7% in Study 2. It remains unknown what causes people to evaluate objects in more or less similar ways. The results of this exploratory PCA demonstrate the possibility of studying mediating factors (e.g., social, contextual, situational, personality, prior experiences) whenever general consensus is not found.

Individualized models

Another way to examine shape's effect when consensus is absent is to analyze the impressions at the level of every individual. We therefore calculated individualized linear models—every participant's average impression was regressed on the scaled values of the shape parameters. From each model, we extracted the beta coefficients to plot against the level of consistency exhibited by the same individual represented by the model. The resulting plots confirm the relevance of shape parameters even in the absence of impression consensus (see Figure 12).

A relationship can be observed in Figure 12 between an individual's consistency and the extent to which shape parameters were significantly predictive of their impressions. Individualized regression models where

none of the shape parameters were significantly predictive of impressions tended to be models of impressions with poor consistency. A prime example occurred in Study 1, where one participant's impressions of likability of *family x* were not predicted by the shape parameters and also were not consistent ($r_M = 0.00$) across re-evaluations.

Another relationship observable in Figure 12 is the pattern for the shape parameters. For example, the pattern in which impressions map to shape parameters appears regular across participants for danger evaluations of *family x* but not beauty evaluations of *family y*. Recalling the results of Study 2, these two conditions also represented the highest and lowest levels of impression consensus. Taken together, Figure 12 provides a qualitative look at how shape parameters work when consensus is present and when it is absent. Among beauty evaluations of *family y*, points mapped positively to beauty for some people but negatively for others. These results are congruent with the exploratory PCA results in that there appear to be clusters of participants sharing similar mappings between shape parameters and impression. Thus, even if a participant does not share the same impression of a novel object with most other participants, they may still share their impression with some others because of the shape of the objects.

Conclusion

How do people begin to form impressions of objects? We hypothesized that the shape of objects contributes to initial impressions. The results of these studies demonstrate that the impressions individuals have of objects they have not seen before, while not clearly identifiable, are meaningful insofar as they have them consistently. In other words, a person does not need to know what something is to have particular feelings or cognitions about it. We show that one contributing factor to these impressions of novel objects is their shape.

This article also contributes to the current literature by introducing the procedural-generation method of stimulus creation to psychological research. To our knowledge, the stimuli presented here represent the most complex parametrized abstract novel objects examined in psychological research. What is meant by “complex” is that the underlying shape parameters vary in more dimensions than has previously been accomplished. Procedural generation provides exciting opportunities for future research attempting to elucidate the way complex visual stimuli are perceived and evaluated by people. Additionally, the process we implement—assessing consistency, consensus, and

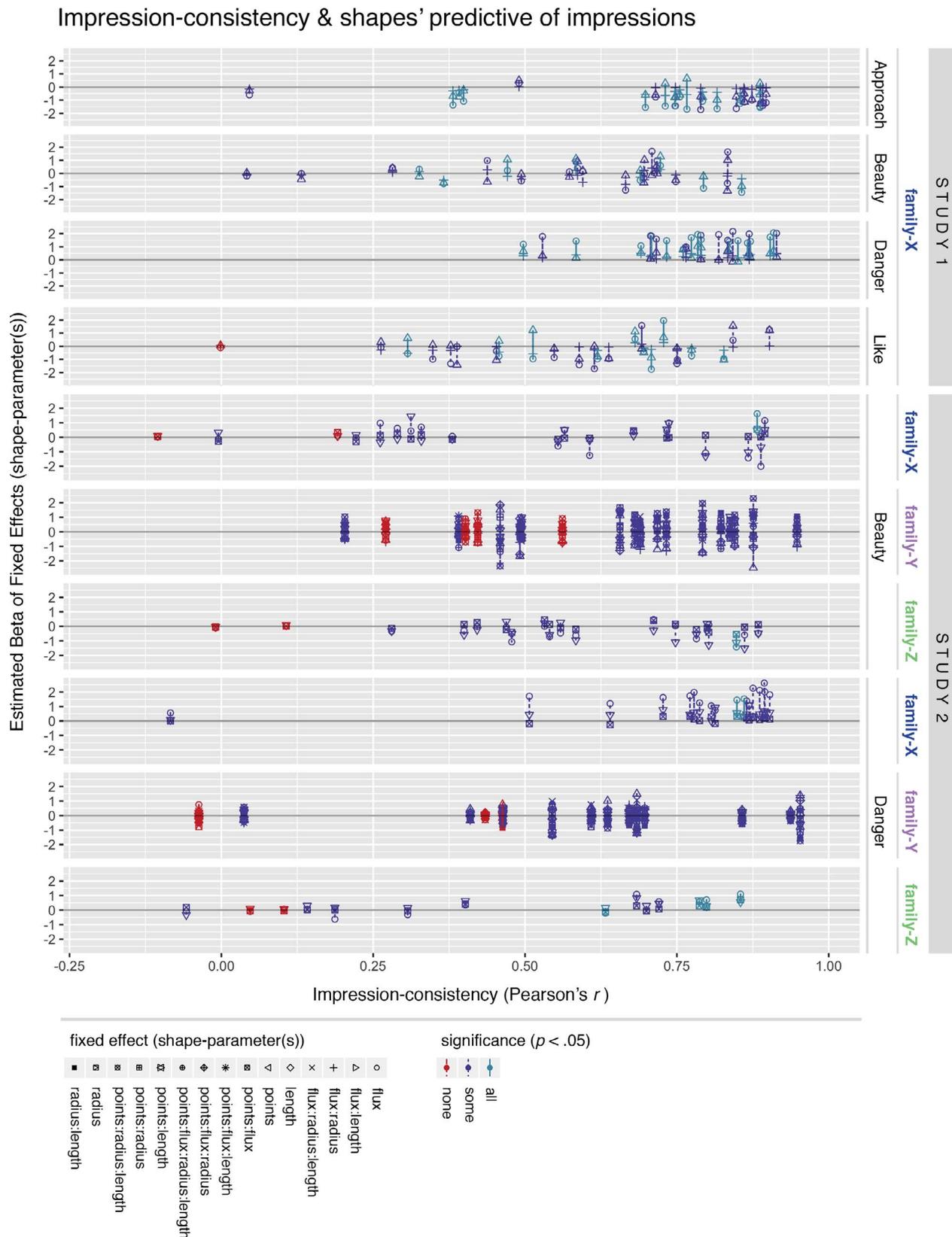


Figure 12. Impression consistency and shapes predictive of impressions. Every individual's impressions were regressed on the shape parameters. Plotted are the beta coefficients (y-axis) against an individual's consistency (x-axis). The beta coefficients describe how participants are individually affected by each shape parameter, and their interactions, when forming their impression. If at least one

→

←

shape parameter was significantly predictive ($p < 0.05$) of an individual's impressions, it is marked as “some”; otherwise, the color of a data point denotes that all or none of the shape parameters were significantly predictive. Note that in Study 2, five data points could not be plotted because of a lack of variance across impressions of *family z*: one beauty evaluation and four danger evaluations. Using this plot, it is possible to glean a qualitative account of how similar individuals are evaluating the novel objects. Similar patterns of fixed effects mean a similar model of impression. Individual models in which none of the shapes were significantly predictive of impressions coincided with lower levels of impression consistency.

modeling impressions using the parametric values underlying each stimulus's shape—may serve as a framework for studying the meaningful metrics by which shapes alter impressions of novel objects. For example, there is something about *family x*'s shape parameters that reliably elicits shared impressions of dangerousness across people. There are numerous possible explanations, such as how pointy the objects look, the ratio between surface area and volume, the average acceleration of the object's surface, and the acceleration of the contour formed by the two-dimensional silhouette of the object. The approach we outline here could be used to rigorously test more specific hypotheses such as these while helping to build an enriched understanding of how people evaluate their visual environments.

Keywords: psychophysics, 3-D surface and shape perception, form perception, shape and contour, shape perception, visual cognition, object recognition, impression, evaluation, procedural generation

Acknowledgments

We are grateful to John Wolfe for helping collect this data, and Joel Martinez for advice with some analyses.

Commercial relationships: none.

Corresponding author: Aaron Kurosu.

Email: akurosu@princeton.edu.

Address: Department of Psychology, Princeton University, Princeton, NJ, USA.

References

- Amir, O., Biederman, I., & Hayworth, K. J. (2011). The neural basis for shape preferences. *Vision Research*, *51*(20), 2198–2206, doi:10.1016/j.visres.2011.08.015.
- Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, *17*(8), 645–648, doi:10.1111/j.1467-9280.2006.01759.x.
- Bartell, F. O., Dereniak, E. L., & Wolfe, W. L. (1981). The theory and measurement of bidirectional reflectance distribution function (BRDF) and bidirectional transmittance distribution function. In *Proceedings of SPIE 0257, radiation scattering in optical systems* (pp. 154–161). Bellingham, WA: SPIE.
- Bertamini, M., Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology*, *107*(1), 154–178, doi:10.1111/bjop.12132.
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147, doi:10.1037/0033-295X.94.2.115.
- Chemero, A. (2003). What is an affordance? *Ecological Psychology*, *15*, 107–114, doi:10.1207/S15326969ECO1502_1.
- Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2016). When “Bouba” equals “Kiki”: Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, *6*(1), 26681, doi:10.1038/srep26681.
- Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science*, *13*(6), 513–520, doi:10.1017/S1466046609990378.
- Emberson, L. L., Crosswhite, S. L., Richards, J. E., & Aslin, R. N. (2017). The lateral occipital cortex is selective for object shape, not texture/color, at six months. *The Journal of Neuroscience*, *37*(13), 3698–3703, doi:10.1523/JNEUROSCI.3300-16.2017.
- Engell, A. D., Haxby, J. V., & Todorov, A. T. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519, doi:10.1162/jocn.2007.19.9.1508.
- Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*(4), 814–823, doi:10.1093/cercor/bhs067.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, *23*(12), 5235–5246.

- Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science*, 259(5091), 100–103.
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and cartesian grating in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76(4), 2718–2739.
- Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training “greeble” experts: A framework for studying expert object recognition processes. *Vision Research*, 38(15–16), 2401–2428, doi:10.1016/S0042-6989(97)00442-2.
- Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A. M., Smoller, J. W., Kwok, H., . . . Wilmer, J. B. (2015). Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Current Biology*, 25, 1–6, doi:10.1016/j.cub.2015.08.048.
- Goldstone, R. L., & Styvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, 130(1), 116–139, doi:10.1037/0096-3445.130.1.116.
- Gómez-Puerto, G., Munar, E., & Nadal, M. (2016). Preference for curvature: A historical and conceptual framework. *Frontiers in Human Neuroscience*, 9, 1–8, doi:10.3389/fnhum.2015.00712.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, 16(2), 152–160, doi:10.1111/j.0956-7976.2005.00796.x.
- Hekkert, P., Peper, C. L. E., & van Wieringen, P. C. W. (1994). The effect of verbal instruction and artistic background on the aesthetic judgment of rectangles. *Empirical Studies of the Arts*, 12(2), 185–203, doi:10.2190/RMEM-K5KL-R2RN-7A2Y.
- Holland, M. K., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or, maluma and takete revisited. *Perceptual and Motor Skills*, 19(1), 111–117, doi:10.2466/pms.1964.19.1.111.
- Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 199–209, doi:10.1037/0096-1523.32.2.199.
- Jiang, Y., Gorn, G. J., Galli, M., & Chattopadhyay, A. (2015). Does your company have the right logo? How and why circular and angular logo shapes influence brand attribute judgments. *Journal of Consumer Research*, 42, 709–726, doi:10.1093/jcr/ucv049.
- Krispel, U., Schinko, C., & Ullrich, T. (2015). A survey of algorithmic shapes. *Remote Sensing*, 7(10), 12763–12792, doi:10.3390/rs71012763.
- Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: Perceiving affective valence in everyday objects. *Frontiers in Psychology*, 3, 1–5, doi:10.3389/fpsyg.2012.00107.
- Lewis, C., & Walker, P. (1989). Typographic influences on reading. *British Journal of Psychology*, 80(2), 241–257, doi:10.1111/j.2044-8295.1989.tb02317.x.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces—CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet. [https://doi.org/ISBN 91-630-7164-9](https://doi.org/ISBN%2091-630-7164-9)
- Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, 102(1), 72–90, doi:10.1037/0033-2909.102.1.72.
- McManus, I. C. (1980). The aesthetics of simple figures. *British Journal of Psychology*, 71(4), 505–524, doi:10.1111/j.2044-8295.1980.tb01763.x.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133–142, doi:10.1111/j.2041-210x.2012.00261.x.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, USA*, 105(32), 11087–11092, doi:10.1073/pnas.0805664105.
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13, doi:10.1016/j.jneumeth.2006.11.017.
- Phillips, F., Norman, J. F., & Beers, A. M. (2010). Fechner’s aesthetics revisited. *Seeing and Perceiving*, 23(3), 263–271, doi:10.1163/187847510X516412.
- Provins, K. A., Stockbridge, C. W., Forrest, D. W., & Anderson, D. M. (1957). The representation of aircraft by pictorial signs. *Occupational Psychology*, 31, 21–32.
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver’s processing experience? *Personality and Social Psychology Review*, 8(4), 364–382, doi:10.1207/s15327957pspr0804_3.
- Rodway, P., Kirkham, J., Schepman, A., Lambert, J., & Locke, A. (2016). The development of shared

- liking of representational but not abstract art in primary school children and their justifications for liking. *Frontiers in Human Neuroscience*, *10*, 21, doi:10.3389/fnhum.2016.00021.
- Russell, P. A. (2000). The aesthetics of rectangle proportion: Effects of judgment scale and context. *The American Journal of Psychology*, *113*(1), 27–42.
- Summit, J., & Sommer, R. (1999). Further studies of preferred tree shapes. *Environment and Behavior*, *31*(4), 550–576.
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118, doi:10.1016/j.cognition.2012.12.001.
- Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, *25*(3), 395–410, doi:10.1080/02643290802044996.
- 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*(15), 1–27, doi:10.1146/annurev-psych-113011-143831.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460, doi:10.1016/j.tics.2008.10.001.
- Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., ... Skov, M. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *Proceedings of the National Academy of Sciences, USA*, *110*, 10446–10453, doi:10.1073/pnas.1301227110.
- Vessel, E. A., & Rubin, N. (2010). Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, *10*(2):18, 1–14, doi:10.1167/10.2.18. [PubMed] [Article]
- Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253, doi:10.1037/0278-7393.19.6.1235.
- Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition*, *72*(2), 125–166, doi:10.1016/S0010-0277(99)00035-9.
- Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research*, *38*(22), 3555–3568, doi:10.1016/S0042-6989(98)00039-X.
- Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition*, *7*(1–3), 297–322, doi:10.1080/135062800394829.
- Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., Connor, C. E., & Charles, E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Neuroscience*, *11*(11), 1352–1360, doi:10.1038/nn.2202.