In order to estimate the shape of objects, the visual system must refer to shape-related regularities in the (retinal) image. For opaque objects, many such regularities have already been identified, but most of them cannot simply be transferred to transparent objects, because they are not available there at all or are available only in a substantially modified form. We here consider three potentially relevant regularities specific to transparent objects: optical background distortions due to refraction, changes in chromaticity and brightness due to absorption, and multiple mirror images due to specular reflection. Using computer simulations, we first analyze under which conditions these regularities may be used as shape cues. We further investigate experimentally how shape perception depends on the availability of these potential cues in realistic scenes under natural viewing conditions. Our results show that the shape of transparent objects was perceived both less accurately and less precisely than in the opaque case. Furthermore, the influence of individual image regularities varied considerably depending on the properties of both object and scene. This suggests that in the transparent case, what kind of information is usable as a shape cue depends on a complex interplay of properties of the transparent object and the surrounding scene.

Introduction

Perceiving the spatial extent and shape of objects is a fundamental ability that allows us to identify objects and successfully interact with them. How the visual system infers the shape of opaque objects has already been investigated in a large number of theoretical and empirical works, and several essential shape cues have been identified—for example, contours and edges, texture, shading, mirror images, and highlights (see Figure 1a). However, there are only a few studies that have investigated visual shape perception in the case of transparent objects (e.g., Chen & Allison, 2013; Chowdhury, Marlow, & Kim, 2017; Interrante, Fuchs, & Pizer, 1995, 1997; Kersten, Stewart, Troje, & Ellis, 2006; Wijntjes, Vota, & Pont, 2015). Most work on perceptual transparency deals with the transmission properties of simple flat filters and the color or brightness relations in the image that lead to perceptual transparency (Anderson, 2015; Beck, 1978; Beck, Prazdny, & Ivry, 1984; Faul, 2017; Faul & Ekroll, 2002, 2011, 2012; Faul & Falkenberg, 2015; Khang & Zaidi, 2002a, 2002b; Ripamonti, Westland, & Da Pos, 2004; Robilotto, Khang, & Zaidi, 2002).

The question arises whether the shape cues identified in the opaque case can somehow be transferred to the case of transparent objects. If one considers the substantial differences in light transport between those two material classes, this appears at least questionable. Light interacts with transparent objects in a much more complex way than with opaque objects: It can interact several times with the surface of the object, its interior, and the environment before it reaches an observer. Depending on the actual material properties, this can change not only the spectral properties of the light but also its direction of propagation. These differences in light transport in the opaque and transparent cases lead to differences in the information available in the image that may potentially be used as shape cues. Thus, the shape cues identified in the opaque case cannot simply be transferred to transparent objects.

In this work, we perform both theoretical analyses and empirical investigations to clarify how shape recognition of transparent objects differs from that of opaque objects.

In the first part, we examine several regularities that are potentially related to the shape of transparent objects. For the most part, we restrict our analysis to single images—that is, to static situations in which objects, illuminations, and observers are stationary. While some cues from the opaque case remain the same (contour) or exist in modified form (mirror images), others are no longer available (e.g., shading, texture). On the other hand, there exist potential cues that are...
specific to transparent objects. These include background distortions caused by refraction and changes in chromaticity and brightness due to internal absorption (see Figure 1b). These regularities are not uniquely related to shape but depend also on many other properties of the object and the scene. While this is a common problem of most visual cues, it is especially pronounced in the case of transparent objects. From a theoretical point of view, it therefore appears likely that the usability of these potential cues depends more heavily on the specific situation than with opaque objects, and that shape recognition differs significantly in both material classes. Since the potential shape cues depend in complex ways on numerous properties of transparent materials, it does not seem appropriate to analyze individual image regularities in isolation. Instead, understanding shape perception in the transparent case appears to require a more comprehensive approach. Further, it must be taken into account that some regularities may be used as shape cues only in certain situations, while in others they appear less appropriate for this purpose.

In the second part of this work, we investigate experimentally how well the visual system infers the shape of transparent objects and on which image regularities it relies in this process. In an experiment, we measured the perceived shape of realistic transparent and opaque objects under natural viewing conditions while varying the availability of different potential shape cues. The results indicate that subjects are able to infer the shape of transparent objects, but less accurately than with opaque objects of identical shape. In addition, some of the image regularities had opposite effects on the accuracy of shape perception, depending on whether the transparent objects were massive or hollow. These results provide strong empirical evidence that the shape perception of opaque and transparent objects actually depends on visual processes that are specific for each material class.

Cues from background distortions due to refraction

Transparent materials generally change the direction of propagation of the light they transmit. This so-called refraction occurs both when light enters and leaves the material. Snell's law describes quantitatively how the degree of refraction depends on the angle of incidence of the light and on the relative optical density of the material and the medium surrounding it. In the image, refraction usually leads to optical distortions of the background visible through a transparent object. This effect is particularly pronounced in massive objects with curved surfaces (cf., Figure 1b).

Some previous works have already dealt with the question of what role these background distortions
play in the perception of transparent materials. For example, Kawabe, Maruya, and Nishida (2015) and Kim and Marlow (2016) have shown that certain aspects of these distortions help to distinguish between transparent and opaque materials. Fleming, Jäkel, and Maloney (2011) suggested that background distortions can serve as a specific cue for the refractive properties of the material (but see Schlüter & Faul, 2014, 2016).

In this work, we focus on the question whether background distortions may also indicate the shape of transparent materials. This problem has so far been investigated mainly in machine vision (Ben-Ezra & Nayar, 2003; Hata, Saitoh, Kumamura, & Kaida, 1996; Morris & Kutulakos, 2011; Murase, 1990, 1992). However, the corresponding findings cannot easily be generalized to the human visual system, because often highly artificial observation situations are assumed and certain scene parameters that are hidden to a human observer are assumed to be known.

To derive the relationship between background distortions and shape, we will consider a simplified situation in which an observer looks through a single slightly waved refracting surface on a flat background. We further assume an orthographic projection, because in this case the differences in the degree of refraction of the light that reaches the observer from different surface locations can uniquely be attributed to differences in the local curvature of the surface. With perspective projection the situation is more ambiguous, because the angle of incidence and thus the degree of refraction depends not only on the surface orientation but also on the viewing angle, which varies across the image. In the retinal image, spatial changes in the degree of refraction lead to optical magnifications and compressions of the background texture that is seen through the transparent surface (since compressions are negative magnifications, we will sometimes refer to both simply as magnifications). Because a local surface patch is usually curved differently in different directions, the degree of local magnification in general also depends on the directions. For each image location, the magnification is maximum in one direction and minimum in a direction perpendicular to it (for details, see Figure 2a). There is an obvious structural similarity between local magnifications and local curvature, which suggests that the two magnification components $M_{\text{min}}$ and $M_{\text{max}}$ are somehow related to the local principal curvatures $K_{\text{min}}$ and $K_{\text{max}}$ of the surface.

To get an insight into how local image distortions and local surface shape are related, a representation of the principal curvatures proposed by Koenderink and van Doorn (1992) seems especially suitable. This representation distinguishes between a qualitative and a quantitative aspect captured by the shape index $s$ and curvedness $c$, respectively. The shape index $s$ describes the type of local surface shape on the continuum, from spherical concave ($s = -1$, “Cup”) to saddlelike ($s = 0$, “Saddle”) to spherical convex shapes ($s = +1$, “Cap”). The curvedness $c$ describes the strength of local curvature independent of the type of surface shape (i.e., the shape index $s$). Figure 2b illustrates how specific patterns of minimum and maximum magnifications are related to the local shape type. Based on this relationship, we propose that given the vector $\mathbf{r}$

$$r_a = (a - r_0)$$

$$r_b = (b - r_0)$$

$$M_{\text{min}} = -(a - r_0)$$

$$M_{\text{max}} = -(b - r_0)$$

Figure 2. Conceptual analysis of the relationship between optical background distortions caused by a light-refracting surface and its shape. (a) Illustration of the light paths of six arbitrary light rays reaching an observer in a hexagonal configuration. The geometry of the underground depicted by the undistorted rays can be approximately described by a circle (blue dashed circle). Its radius $r_0$ is given by the eigenvalues of the covariance matrix of the reflection points on the underground. The background area depicted by the distorted rays can vary in size, position, and shape. For sufficiently small bundles of light, the form of this background patch can be approximated by an ellipse. Its half-axes $a$ and $b$ are related to the minimum and maximum magnifications $M_{\text{min}}$ and $M_{\text{max}}$ with which the ray bundle depicts the underground. More specifically, $M_{\text{min}} = -(a - r_0)$ and $M_{\text{max}} = -(b - r_0)$. (b) Illustration of how the geometry of an optically distorted background patch (bottom), and thus its minimum and maximum magnifications $M_{\text{min}}$ and $M_{\text{max}}$ are related to the shape type of the refracting surface (top). Like in (a), the blue dashed circles denote the undistorted background patches, while the red circles/ellipses denote the background patches optically distorted by refraction. Specific patterns of minimum and maximum magnifications are related to qualitatively different surface shapes.
The orientation of the vector $\tilde{M} = (M_{\min}, M_{\max})$ approximates $s$ and its length approximates $c$.

$\tilde{M} = (M_{\min}, M_{\max})$, the shape index $s$ can be approximated by the orientation and the curvedness $c$ by the length of this vector $\tilde{M}$. More specifically, the angular range $[(1/4)\pi, (5/4)\pi]$ of $\tilde{M}$ is mapped to the range $[1, -1]$ of the shape index $s$. Figure 3 depicts the results of a simulation which confirms the approximate validity of this relationship. Since shape index and curvedness are just an alternative representation of the principal curvatures $K_{\min}$ and $K_{\max}$ of the surface, the magnifications can also be used to estimate the principal curvatures.

**Numerical experiment: Estimating shape from background distortions due to refraction**

To analyze how the magnifications caused by a three-dimensional light-refracting surface are related to its principal curvatures and thus its shape, we conducted a numerical experiment in which the interaction of light rays with a three-dimensional water surface was simulated. Estimates of the local surface shape and the strength of curvature were then derived from the optical distortions of the surface beneath the water. To test how robust these estimates are, we varied both the amplitude of the water surface and its distance to the underground.

**Stimuli**

The stimuli were computer-generated images of a water surface that was modeled with the 3-D modeling software Blender (Blender Foundation, 2015). The surface was created from a plane of size $140.4 \times 140.4$ mm, which consisted of 24,200 triangular faces. The vertices of this plane were displaced along their normal direction (Displace modifier with parameters direction $= \text{normal}$, texture coordinates $= \text{local}$, midlevel $= 0.5$, and strength $= 0.5$). The amount of displacement was determined by the intensity of a Perlin noise texture (texture “Cloud” with parameters noise basis $= \text{original Perlin}$, size $= 1$, and depth $= 0$, and options “Grayscale” and “Soft” selected). The resulting maximum amplitude of the water surface was $a = 16.9$ mm. The mean distance to the underground was set to $d = 70.7$ mm.

In two additional conditions, either the amplitude of the water was adjusted to $4a$ by adjusting the size parameter of the Perlin noise texture accordingly or the distance to the underground was set to $8d$.

An orthographic camera was located above the water surface. The scene was rendered as a high-dynamic-range image with the physically based Cycles renderer (image size $520 \times 520$ pixels, 2,048 samples/pixel, color depth $= 16$ bits/channel). The material properties of the wavy surface have been adjusted to match those of water. However, to isolate the effects of optical distortions, Fresnel effects and specular reflections have been ignored (Refraction BSDF material shader with parameters distribution $= \text{sharp}$, color $= \text{RGB}[1, 1, 1]$, roughness $= 0$, and IOR $= 1.33$).

**Procedure**

To gain information about the actual light paths, we provided the underground with color gradients, so that each point of the underground had a unique color. Using the color of a pixel in the rendered image as an index, it was therefore possible to determine the point on the underground that it depicts. To determine the degree of displacement due to refraction, the same scene was rendered with and without the water surface. Then the positions of pixels of the same color in both rendered images were compared. In this way, both the light paths (beginning with the corresponding points on the underground) and the displacement due to refraction (given by the distance of corresponding points in the images) could be calculated.

Based on this procedure, the minimum and maximum magnifications ($M_{\min}$ and $M_{\max}$) of a total of 492,246 evenly distributed bundles of light $B_i$ passing through the water surface were calculated. Each bundle of light consisted of six arbitrary light rays arranged hexagonally (cf., Figure 2a). The distance between adjacent rays was $0.27$ mm. The orientation and length of the vector $\tilde{M}_i = (M_{\min,i}, M_{\max,i})$, which denotes the magnifications for bundle of light $B_i$, were then used to estimate the corresponding shape index and curvedness, respectively. The veridical principal curvatures were calculated using a procedure proposed by Rusinkiewicz (2004). From them, the veridical shape index and curvedness were also calculated.

It should be noted that the procedure used here is based on the inverse of the real image-generation
process, where light propagates after a diffuse reflection at the underground in the direction of the observer. However, this has no influence on the distortions of the background in the retinal image. This fact is exploited in the path-tracing algorithm used in much render software, which also starts from the viewing point and follows the light rays backward to surfaces and light sources in the scene.

**Results**

The stimuli and results of the numerical experiment are shown in Figure 4.

*Slightly curved surface:* The leftmost two diagrams in Figure 4a show a subset of the light paths simulated for the slightly curved water surface. The minimum and maximum components of the curvature are strongly correlated with the minimum and maximum components of the magnification (see Figure 4a, third diagram from the right). Estimating the shape index and curvedness from the magnification components leads to only minor deviations from ground truth (see Figure 4a, rightmost two diagrams). It should be noted, however, that estimated and veridical curvedness values can differ from each other by an a priori unknown factor. This factor depends on arbitrary parameters, such as the diameter of the simulated bundle of light. Estimating the absolute strength of curvature from the magnifications would therefore require an appropriate anchoring.

*Strongly curved surface:* If the amplitude of the water surface is quadrupled, the higher curvature causes the light rays to converge and diverge more strongly than with the less wavy water surface (see Figure 4b, leftmost two diagrams). As a consequence, the rays from the underground to the water surface may cross. This happens whenever the positive curvature of a surface area exceeds a certain value. Below this critical
value, an increasing curvature is accompanied by an increasing magnification. At the critical curvature, the maximum possible magnification is reached and all light rays of the corresponding image area converge to a single point on the background. If the curvature is further increased beyond this critical value, then corresponding points on the underground start to move away from each other and the magnification begins to decrease again (we will refer to this as magnification inversion). At extreme curvatures, the magnification can even become negative and turn into a compression. If a magnification inversion occurs, the correlation between magnification and curvature is no longer monotonic but inversely V-shaped (see Figure 4b, third diagram from right). If shape type and curvature strength are estimated without taking the magnification inversion into account, this results in larger deviations from the veridical values (see Figure 4b, rightmost two diagrams).

Increased background distance: Magnification inversion can also occur if the distance to the underground is increased (see Figure 4c, leftmost two diagrams). Although the direction changes at the water surface are relatively small, the rays can nevertheless cross each other because of their greater travel distance. Correspondingly, the correlation between curvature and magnification is no longer monotonic (see Figure 4c, third diagram from right) and there are larger deviations between the estimated and veridical shape parameters (see Figure 4c, rightmost two diagrams). In this case, however, the strength of curvature can be estimated more accurately than the shape type.

Conclusion

The results of the numerical experiment show that in certain situations, distortions of the underground can be used to estimate the shape of the water surface that lies above it. However, our analysis has also revealed that an approximately linear relationship between shape and background distortions can only be assumed in cases where light rays do not intersect and only simple optical compressions and magnifications occur.

Generalization and open questions

Estimating the shape of transparent objects from background distortions caused by light refraction is a very complex problem. The approach described in the foregoing, which explores the usability of distortions as a shape cue on a rather fundamental level, can therefore only be a starting point for more thorough investigations. In Appendix A we discuss a number of topics that in our view need to be addressed in further research. First we address the question of how the veridical optical magnifications, which are not directly accessible to an observer, could be estimated from the image, and to what extent orientation maps could help with this. Then we address the point that under natural viewing conditions, optical distortions indicate not necessarily the intrinsic local curvature of a surface but its curvature relative to the observer—that is, the rate at which the surface orientation changes for successive locations in the image.

Use of distortions as a cue in more complex situations

In the transparent case, the light transport can in principle be of any complexity. For example, in a typical transparent object with a closed outer surface, light reaching the observer from the background is usually refracted not once but at least twice. The question arises how background distortions are related to shape in such cases. An exemplary analysis of the corresponding light paths shows how complex the depiction of the underground and thus the distortions can be in such cases (see Figure 5a). If the light is refracted twice, the distortion at each image point is simultaneously determined by the shape of two different parts of the object's surface. Inferring the shape of one of the surfaces separately is difficult, if not impossible, because the distortions in the image are not merely the sum of two individual distortions that are independent of each other. If background distortions can contribute at all to the perception of shape in such situations, it seems at least necessary to make additional assumptions about the range of possible shapes that may occur in a given context.

In a way, the situation becomes even more complex when light reaching the observer is refracted more than twice, as for example in the case of hollow objects. However, the larger number of refractions does not necessarily lead to stronger distortions. With relatively small wall thicknesses, the distortions can be much weaker than for massive objects of the same shape (see Figure 5b). Although the fundamental problem of inferring the shape of one of the surfaces that contribute to the image distortion remains the same, there are two favorable circumstances that may facilitate the estimation of object shape. First, the correlation between distortions and curvature is barely affected by magnification inversions, as most light rays that reach the observer do not cross. Second, two of the four surfaces involved in the distortion often have a very similar shape (outer and inner front and back of the object). Their influence on the distortion is therefore relatively similar and could thus serve as a joint cue for the shape of each pair of surfaces.

Further complexity arises from the fact that in the general case, the background behind the distorting material does not have to be flat but can consist of
The distortions then also depend heavily on the properties of the background scene. This includes not only the shape and material of objects seen in the background but also their position relative to the transparent object. Using distortions as a shape cue in such situations would be even more difficult.

In addition to the aspects considered so far, temporal changes in the scene can influence background distortions. Although this dynamic further increases the complexity of the scene, it may also provide additional shape cues, as is suggested by works from machine vision (Morris & Kutulakos, 2011; Murase, 1990, 1992). In addition, dynamics could help to identify individual shape cues in the image that are confounded in a static stimulus. For example, the contributions of the distorted background and the distorted mirror images might be more easily separated, because the image features related to these causes would in general move differently in the image.

Cues from changes in intensity and chromaticity due to absorption

Many transparent materials absorb parts of the transmitted light. Absorption is described by the Bouguer–Lambert–Beer law and can change both the intensity of the light and its spectral distribution. In the image, this leads to a darkening and to changes in chromaticity. Since the strength of these effects depends on the length of the light path inside the object, they are related to the thickness of the object and may thus indirectly contribute to the recognition of the object’s shape. For simple transparent filters, Faul and Ekroll (2011) have already shown that estimating the thickness from transmission and saturation can succeed in certain situations. In this section, we will investigate to what extent these findings can be generalized to objects of more complex shape and to more natural viewing conditions. Since estimates based on changes in luminance have proven to be much more robust than estimates based on saturation changes (Faul & Ekroll, 2011), we will focus on the role of absorption-related darkening.

Darkening due to absorption can only be used as a shape cue if it can be discerned from other causes of darkening in the image, such as spatially varying background-reflectance properties or changes in transmittance and reflectance due to Fresnel effects. Comparable problems arise in identifying the shading of opaque objects, because this requires isolating the darkening in the image due to orientation changes of the surface from other causes of darkening, and it is quite possible that similar perceptual mechanisms are used in both cases.

A further problem arises from the fact that the relationship between darkening and thickness is significantly influenced by the absorption properties of the material and by factors such as light refraction and total reflection. Objects of different thicknesses can therefore lead to similar absorption-induced darkening, if the absorption coefficients of their materials are properly chosen. This means that the absolute thickness of a material can only be estimated if the absorption coefficient is known. Without this knowledge, at most information about the relative thickness can be obtained. Anchoring such relative-thickness informa-
tion would then require integration with other shape
cues. The situation becomes particularly complex when
the absorption properties of the material are spatially
inhomogeneous. In such cases, even relative-thickness
information can no longer be deduced unambiguously
from the darkening. In addition, due to refraction or
total reflections inside the object, it cannot be assumed
that the light rays reaching the observer have crossed
the object along a straight line in the respective viewing
direction. As a consequence, the distance traveled by
the light within the material is generally not identical to
the thickness of the object along the respective line of
sight.

Numerical experiment: Estimating shape from
intensity changes due to absorption

To analyze the correlation between absorption-
induced darkening and object thickness, and to test
how robust thickness values estimated from the
darkening are against influences from refraction and
total reflection, we performed a numerical experiment.

Stimuli

The stimulus material consisted of computer-generated
images of 15 randomly shaped transparent objects.
The objects were created with Blender. The object mesh
was based on an icosahedron whose faces were
subdivided six times. The resulting icosphere consisted
of 81,920 triangular faces and was adjusted to a
diameter of 100 mm. The icosphere was deformed by
translating its vertices along their normal direction
(Displace modifier with parameters direction = normal,
midlevel = 0.5, and strength = 1). The amount of
displacement was determined by the intensity of three-
dimensional Perlin noise (texture “Cloud” with pa-
rameters noise basis = original Perlin, size = 1, and
depth = 0, and options “Grayscale” and “Soft”
selected). To gain different shapes, the noise was
probed at different locations.

A perspective camera was placed at a distance of 400
mm from the center of the object (vertical field of view
= 44.10°). The objects were rendered as high-dynamic-
range images with the Cycles renderer (image size 1,040
× 1,040 pixels, 2,048 samples/pixel, color depth = 16
bits/channel). An infinitely distant spherical emitter
was set to homogeneous white (Background surface
shader with parameters color = RGB[1, 1, 1] and
strength = 1).

Procedure

Each of the 15 objects was rendered three times with
different material properties to gain information about
the veridical object thickness, absorption-induced
darkening, and areas of total reflection.

In the first run, object and scene parameters were
chosen so that the image contained information about
the veridical object thickness. To this end, the object
was defined not to have a distinct surface but to consist
exclusively of a homogeneous absorptive material (no
surface shader; Volume Absorption volume shader
with parameters color = RGB[0.8, 0.8, 0.8] and density
= 0.03). The resulting image shows the decrease in light
intensity caused by absorption. The intensity values of
this image were transformed to represent the actual
object thickness along the respective viewing directions.
To this end, each pixel of the grayscale image
was transformed so that $p'_{xy} = -1/a \times \log(p_{xy}/1)$, where $a$
corresponds to the absorption coefficient that was
defined during the rendering. With $C = 0.8$ and $D = 0.03$
corresponding to the values of the parameters color
and density used by the Cycles renderer, $a$ is given as $a$
$= (1 - C) \times D$.

In the second run, object and scene parameters were
chosen so that the image represented absorption-
induced darkening. To this end, a surface shader was
added to the object material that takes refraction and
total reflections into account but ignores nontotal
specular reflections and Fresnel effects (see Appendix B
for the Blender node setup that was used to define the
Cycles material). The intensity values in the image then
corresponded to the amount of darkening. This can be
described as the ratio $I(d, \lambda)/I_0(\lambda)$, where $I_0(\lambda)$
denotes the intensity of light of wavelength $\lambda$ before it enters the
material and $I(d, \lambda)$ the intensity of the exiting light
after it travels a distance $d$ inside the material. Based on
this image, local estimates of the object thickness were
calculated. Due to the exponential relationship between
darkening and light-path length inside the object, the
local thickness of the objects was estimated by
$t = -1/a(\lambda) \times \log(I(d, \lambda)/I_0(\lambda))$. This can be consid-
ered as the estimate of an ideal observer, who can
unambiguously identify the darkening due to absorption
and knows the absorption coefficient of the
material.

In the third run, object and scene parameters were
chosen so that image areas with total reflection could
be identified. To this end, the object material was
changed so that it refracts light (without taking Fresnel
effects into account) but does not reflect or absorb it
(Refraction BSDF surface shader with parameters
BSDF = sharp, color = RGB[1, 1, 1], roughness = 0,
and IOR = 1.49; no volume shader). With this setup, all
pixels of the image hit by light rays that were totally
reflected within the object at least once along their path
were displayed in black. All other pixels were white.
This black-and-white image was then used as a mask to
separately consider areas with and without total
reflections.
Results

Figure 6 summarizes the simulation results. While both refraction and total internal reflection reduce the correlation between darkening in the image and object thickness for all 15 simulated objects. The saturation of a point corresponds to the frequency with which a certain combination of darkening and object thickness occurred. The correlation for light that has been totally reflected at least once (red points) is substantially weaker than for light that has only been refracted (blue points). For comparison, the dashed gray line in the plot shows the relationship between darkening and thickness for (hypothetical) light that is neither refracted nor totally reflected. (b) Typical spatial error distribution demonstrated at one of the simulated objects. In places where the light path includes total reflection (red areas), the error (represented as saturation) is much greater than where the light has only been refracted (blue areas). Since total reflections occur mainly near the object’s rim, this is where the errors are largest. The negative influence of reflection is also largest near the rim of the object, where light hits the surface at a shallower angle. (c) Distribution of the error for image areas with (red) and without (blue) total reflections for all 15 simulated objects. In 94% of the areas with total reflection, the estimated thickness deviates by more than 100% from the veridical one (i.e., $|t - \hat{t}|/t > 1$). In the area without total reflection this applies to only 11% of the cases.

Conclusion

The results of the numerical experiment show that absorption-related shape cues seem less complex than the distortion-related ones already discussed, there are still many open questions that need to be addressed in future research. In Appendix C, we discuss whether absorption-induced darkening can provide shape information in cases where a transparent object is hollow, as well as how thickness information could be transformed into a specific object shape.

Generalization and open questions

Although absorption-related shape cues seem less complex than the distortion-related ones already discussed, there are still many open questions that need to be addressed in future research. In Appendix C, we discuss whether absorption-induced darkening can provide shape information in cases where a transparent object is hollow, as well as how thickness information could be transformed into a specific object shape.

Cues from mirror images due to specular reflections

Light that hits a transparent material is usually not completely transmitted. Instead, a part of the light is specularly reflected at the material’s surface. The relative amount of reflected and transmitted light is described by the Fresnel equations and depends on the angle of incidence and the refractive properties of surround and material. Like in the opaque case, the microstructure of the surface determines how strongly the light scatters around the mirror direction. Here we focus on ideally smooth surfaces that reflect the incident light solely in the mirror direction. In the retinal image, these ideal specular reflections lead to a sharp mirror image of the environment on the outer surface of the transparent object (Figure 1b, right-most).
Opaque and transparent objects of the same shape and surface structure produce geometrically identical mirror images. This suggests that existing findings on the role of gloss in the perception of opaque objects’ shape (Adato, Vasilyev, Ben-Shahar, & Zickler, 2007; Adato, Vasilyev, Zickler, & Ben-Shahar, 2010; Fleming, Torralba, & Adelson, 2004; Muryy, Welchman, Blake, & Fleming, 2013; Oren & Nayar, 1997; Savarese, Chen, & Perona, 2004, 2005; Savarese, Fei-Fei, & Perona, 2004; Savarese & Perona, 2001, 2002) can be transferred to transparent objects. However, reflections that occur with transparent objects can be substantially more complex. For example, light can hit the surface of transparent objects not only from the outside but also from the inside. These internal reflections differ from the external ones in that they depend in another way on the angle of incidence and that total reflections can occur at larger angles of incidence. Like the exterior ones, internal reflections also produce mirror images of the surround. There can therefore be as many mirror images as there are reflective surfaces the light rays interact with on their way to the observer. In the case of massive transparent objects, there are typically two mirror images (see Figure 7a, top, and 7b, top). Both mirror images are superimposed additively in the image (see Figure 7c, top). Because the light reflected from the front surface reaches the observer, there are in addition reflections of third and fourth order (see Figure 7a, bottom, and 7b, bottom), which are also superimposed additively in the image (see Figure 7c, bottom).

Transparency-specific problems in using reflections as a cue

The existence of higher order reflections in transparent objects poses specific problems that do not occur in the case of opaque objects. We will briefly discuss the most severe of them.

First-order reflections

In order for first-order reflections to serve as a shape cue, like in the opaque case, they first have to be isolated from higher order reflections. In machine vision this problem has already been discussed in more detail (Morris & Kutulakos, 2007). However, such approaches cannot easily be transferred to human vision, as they typically rely on special observation conditions. In general, the isolation of first-order reflections could be facilitated by the fact that higher order reflections are affected by the absorption properties of the object’s material and therefore usually differ from first-order reflections in both intensity and spectral distribution. Furthermore, reflections of dif-
different orders move differently in the image when the object or observer moves.

**Higher order reflections**

Higher order reflections must not, however, be regarded merely as a source of interference in shape recognition. In principle, they could provide further information about the shape. This could particularly be the case with hollow objects. As long as the wall thickness is relatively small, first- and second-order reflections are often very similar. Since both reflections are superimposed in the image, the second-order reflections can indirectly increase the visibility of first-order reflections. The similarity between the mirror images could also be used in more complex ways. For example, Shih, Krishnan, Durand, and Freeman (2015) have shown how small shifts between two mirror images could be used to detect them and remove them from the image.

Higher order reflections might also be used to estimate the shape of surface areas that are not directly visible. In principle, each reflection in the image could be used to estimate the shape of the surface on which that reflection occurs. In the case of a massive object, for example, second-order reflections could be used to estimate the shape of its rear surface. Similarly, third- and fourth-order reflections could be used to estimate the shape of the inner and outer rear surfaces of a hollow object. From a computational point of view, however, such a procedure would be very complex. While the light of second-order reflections is already refracted twice, the light of third- and fourth-order reflections is refracted four or six times at different surface locations of the object before it reaches the observer. Each of these refractions distorts the transmitted light pattern and changes its intensity distribution. As in the case of (multiple) optical distortions of the background, these distortions of the mirror images interact with each other in a very complex way. In addition, light that is reflected by a specific surface might be not only refracted but also reflected several times on its previous or subsequent path. This can lead to reflections of reflections. One reason for this can be the total reflection already discussed.

Higher order reflections depend not only on the shape of the surface that causes them but also indirectly on the shape of all other surfaces of the object with which the light interacts. In order to gain information about the shape of a specific surface, one would have to isolate the influence on the image of that particular surface reflection from other influences. This is difficult, because both the number of the remaining interactions and their type (reflection, total reflection, refraction) are a priori unknown.

### Summary of the theoretical analyses

Our analyses reveal that distortions of the background in the retinal image caused by transparent objects can be closely related to the shape of these objects: In simple situations there is often a direct correlation between optical magnifications and the local principal curvatures of a surface, and in such cases it seems possible to estimate the shape from the direction-dependent texture densities in the image. In general, however, the relationship between background distortions, image magnifications, and shape is considerably more complex. Distortions are then related less to the intrinsic curvatures of a surface but rather to the curvatures relative to the observer. If the refraction effects are so strong that light rays cross on their way to the observer, this correlation becomes increasingly ambiguous. In cases where light is refracted at several surfaces, it is difficult if not impossible to determine the shape of one of the surfaces involved in the distortion. The extent to which background distortions contribute to the perception of shape therefore seems also to depend on the ability to identify situations in which they do not provide reliable information about shape.

Absorption inside a transparent material can change the intensity and spectral distribution of the transmitted light. The longer the distance that the light travels inside the object, the greater the influence of absorption. In principle, the absorption-induced darkening and chromaticity changes of the background can indicate the thickness of an object and thus indirectly contribute to the recognition of the object’s shape, but this darkening is often difficult to identify. Moreover, light refraction and total reflection normally weaken the correlation between darkening and thickness. Absorption could indirectly influence shape perception, because it can affect the visibility of other potential shape cues.

Transparent objects can reflect light at multiple surfaces, and the corresponding mirror images are superimposed in the image. The mirror images on the front surface facing the observer are identical to those of opaque objects and could therefore serve as a shape cue in a similar way. The use of the remaining mirror images would be computationally much more complex, since they are usually influenced by several interactions of the light with the surface and the material of the object.

### Experiment: Testing the contribution of different cues to shape perception

In the first part of this work, we identified regularities in the image related to the shape of transparent objects. While some of the regularities
known from opaque objects are no longer present (e.g., shading and texture), others remain unchanged (contour) or are present in a similar way (mirror images). Other regularities are specific to transparent objects, such as background distortions due to refraction and changes in chromaticity and brightness due to absorption. The correlations between these image regularities and shape are in most situations substantially more complex than in the case of opaque objects. The different sets of potential shape cues that are available in the opaque and transparent cases suggest that the mechanisms used in shape perception depend also on these material classes. This raises two main questions: How well can the shape of transparent objects be recognized, and how good is the performance compared to that found with opaque objects? And do the image regularities that we just analyzed theoretically play a role in the perception of shape?

To investigate these questions, we conducted an experiment in which we presented subjects with randomly shaped, bloblike objects, which have often been used for investigating shape perception in the opaque case (e.g., Todd, 2004). To determine the perceived shape, we asked the subjects to indicate the local surface orientation of the objects by adjusting small measuring probes that were projected onto their surfaces (gauge-figure task; Koenderink & van Doorn, 1992). In contrast to simple identification tasks, in which subjects have to identify a particular object among several distractor objects, this gauge-figure task has the advantage that it can generally not be solved solely on the basis of the shape information given by the contour. Furthermore, the data obtained with this method can be used to reconstruct the surface shapes perceived by the subjects.

Based on our computational analyses and simulation results, we used massive and hollow versions of transparent objects, since it is plausible to assume that they are processed differently. In addition, we used objects of identical shape that were made of opaque materials to compare the shape perception between the two material classes. Like in our computational analyses, we limited ourselves to static stimuli and used transparent objects with smooth surfaces that had no subsurface scattering.

To test whether the image regularities that we analyzed computationally play a role in perceiving transparent objects, we manipulated their availability in the image. To this end, we omitted in each condition one image regularity from the full set.

We aimed to investigate shape perception for realistic scenes under natural viewing conditions. We therefore chose physically plausible materials (i.e., realistic refractive properties and realistic absorption spectra) and used stereo-scopic stimuli. As a result, all image regularities were present in their usual form with binocular disparity. In the General discussion, we discuss the role that disparity information might play in perceiving the shape of transparent objects.

**Stimuli**

The stimuli were computer-generated stereoscopic images of randomly shaped transparent and opaque objects placed on a floor. In the transparent case, objects were either massive or hollow. The object meshes were created with Blender, and the actual stimulus images were rendered with the physically based Mitsuba renderer (Jakob, 2013). Both modeling and rendering were performed in RGB color space.

The object meshes were based on an icosahedron that was subdivided six times. The resulting icosphere consisted of 81,920 triangular faces and was adjusted to a diameter of 100 mm. A total of seven deformed instances of this icosphere were created by translating its vertices along their normal direction (Displace modifier with parameters direction = normal, midlevel = 0.5, and strength = 1). The amount of displacement was determined by the intensity of three-dimensional Perlin noise (texture “Cloud” with parameters noise basis = original Perlin, size = 1, and depth = 0, and options “Grayscale” and “Soft” selected). To achieve different shapes, the noise was probed at different locations. We avoided locations that would have led to shapes with extensive self-occlusions, because this would have unnecessarily complicated the later reconstruction of perceived surface shapes. The seven meshes used in the experiment are shown in Figure 8. While the position of the objects in the scene remained constant, the vertical position of the floor was adjusted to compensate for the varying vertical extent of the objects. Hollow objects had a wall thickness of 1 mm and were created by eroding their interior without changing the outer shape (Solidify modifier with parameters thickness = 2 and offset = −1, and options “Even Thickness” and “High Quality Normals” selected). After modeling, the objects were exported and the scene rendered with the Mitsuba renderer. In the rendering, all objects were surrounded by (and, if hollow, filled with) nonabsorptive air (refractive index $R \approx 1$).

In eight different cue conditions, four for opaque objects and four for both massive and hollow transparent objects, we manipulated the availability of different known or potential shape cues (see Figure 9). For all except one cue condition, this was achieved by choosing appropriate material properties for both the object and its background. In the remaining cue condition (Mirr−−), we explicitly manipulated the image generation. For both transparent and opaque objects,
we defined base conditions that contained all cues that were manipulated in the remaining conditions.

In the base condition for the massive transparent case (Trns:Mass:Full), the objects were made of red-tinted acrylic glass with a smooth surface (Mitsuba plugin “dielectric,” refractive index \( R = 1.49 \), absorption coefficient \( a_{\text{Mass}} = \text{RGB}[0.0048, 0.0072, 0.0072] \) 1/mm). In the base condition for hollow transparent objects (Trns:Holl:Full), a higher absorption coefficient was used (\( a_{\text{Holl}} = \text{RGB}[0.176, 0.264, 0.264] \) 1/mm) in order for massive and hollow objects to appear similarly tinted. In the remaining three cue conditions for the transparent case, potential shape information from distortions, darkening, and mirror images were omitted individually. To omit optical distortions of the background (cue condition Dist–), we changed the color of the background to a uniform gray with a reflectance of \( \text{RGB}[0.2, 0.2, 0.2] \). To omit darkening due to absorption (cue condition Dark–), the object’s material was set to not absorb any light (\( a_{\text{Dark}} = \text{RGB}[0, 0, 0] \) 1/mm). To omit mirror images (cue condition Mirr–), specular reflections at any of the object’s surfaces were disabled during the rendering process.

In the opaque base condition (Opq:Full), the objects were made of plastic (polypropylene) with a smooth surface (Mitsuba plugin “plastic,” refractive index \( R = 1.49 \)) and had a red granitelike texture based on three-dimensional Perlin noise (texture “Cloud” with parameters noise basis = original Perlin, size = 1.5, and depth = 2, and options “Grayscale” and “Soft” selected). The reflectance of the texture ranged from \( \text{RGB}[0.144, 0.096, 0.096] \) to \( \text{RGB}[0.336, 0.224, 0.224] \). Each object was textured individually, so that the intrinsic texture density remained constant despite the different shapes. In two additional cue conditions for the opaque case, shape information from texture and mirror images were omitted individually. In the fourth condition, information from mirror images was presented in isolation from the other cues. This condition was implemented because it corresponds to the large material class of reflective metals, which has already been studied in more detail in some of the works on the perception of the shape of opaque objects. To omit texture information (cue condition Tex–), the objects’ surface color was changed to a uniform red with a reflectance of \( \text{RGB}[0.24, 0.16, 0.16] \). To omit mirror images (cue condition Mirr–), the objects were made of an ideal diffuse material with a Lambertian reflectance (Mitsuba plugin “diffuse”). To isolate mirror images (cue condition Mirr+), the objects were made of a metallike material that reflects incoming light specularly (Mitsuba plugin “conductor,” reflection coefficient \( r = \text{RGB}[0.72, 0.48, 0.48] \)).

Unless otherwise stated, the floor below the objects showed gray graph paper that was made of two superimposed grid textures of different size. This texture was well suited to depict a wide range of optical magnifications and compressions as they were caused by the different refractive and reflective materials used here. The background was made of an ideal diffuse material with a Lambertian reflectance (Mitsuba plugin “diffuse”). The widths of the two grids were 1 and 10 mm. The reflectance of the grid lines (\( \text{RGB}[0.28, 0.28, 0.28] \)) was slightly higher than that of their surround (\( \text{RGB}[0.12, 0.12, 0.12] \)).

The scene was illuminated by an environment map (Mitsuba plugin “envmap”). The illumination texture was a high-dynamic-range image (color depth = 16 bits/channel) of a natural daylight outdoor scene with a partly cloudy sky (Yimm & Bell, 2008). This environment map was considered a typical representative of realistic and natural ambient lighting.

The camera settings were chosen to correspond to the actual experimental setup (distance to the center of the object = 400 mm, vertical field of view = 44.10°, lateral offset = ±32 mm). Thus, the stimuli appeared in virtually the same way as in a corresponding real scene. The stimuli were first rendered as high-dynamic-range images (color depth = 16 bits/channel) with the Mitsuba renderer (extended volumetric path tracer with maximum path depth = 64; Hammersley sampler with 2,048 samples/pixels; Gaussian reconstruction filter with \( SD = 0.5 \); image size = 839 × 1,200 pixels).
Note that more complex effects resulting from the dispersion or polarization of light were not taken into account.

To compensate for the limited dynamic range of the display device used in the experiment, the stimuli were tone mapped to low-dynamic-range images (color depth = 8 bits/channel) according to the procedure described by Reinhard and Devlin (2005). All stimuli were tone mapped with the same set of parameters. This refers to both the initial parameters (contrast = 0.1, intensity = 1.5, chromatic adaptation = 0.1, light adaptation = 1) and the implicit image-dependent parameters that were first gained for one high-dynamic-range stimulus that contained high luminance values (drawn from the opaque cue condition Mirr+) and subsequently used for the remaining stimuli. Due to the limited horizontal field of view of the mirror stereoscope used in the experiment, the stimulus images for the right and left eye were trimmed at their right and left edges, respectively. The final size of a half image was 720 × 1,200 pixels. To slightly increase the contrast of the images, they were gamma corrected with an exponent of γ = 1.2, which is slightly lower than the value (1.6) that was used by Reinhard and Devlin (2005).

Subjects

A total of 42 subjects (38 women, four men) participated in the experiment. Their ages ranged from 18 to 34 years. All subjects were unaware of the purpose of the experiment. They reported normal or corrected-to-normal visual acuity, and showed no color-vision deficiency, as tested by Ishihara plates (Ishihara, 1969).
Procedure

To keep the duration of the experiment within reasonable bounds, each of the 42 subjects performed only a subset of all conditions. Each subject was assigned to one of two groups, and each group was presented with seven out of 12 stimulus conditions, consisting of massive and hollow transparent and opaque objects in different cue conditions. The first group was presented with three opaque conditions (Opq:Full, Opq:Tex−, and Opq:Mirr−) and all four massive transparent ones (Trns:Mass:Full, Trns:Mass:Dist−, Trns:Mass:Dark−, and Trns:Mass:Mirr−). The second group was presented with two opaque conditions (Opq:Full and Opq:Mirr+), one massive transparent one (Trns:Mass:Full), and four hollow transparent ones (Trns:Holl:Full, Trns:Holl:Dist−, Trns:Holl:Dark−, and Trns:Holl:Mirr−). The base cue conditions for the opaque and the massive transparent case were shared by both groups, to control for any systematic effects between the groups. In both groups, the seven randomly shaped objects were balanced across the seven stimulus conditions and the 21 subjects. As a result, every object was combined with every stimulus condition. Although each subject was presented with all seven objects, each object was presented in only one stimulus condition. This ensured that subjects did not see objects of identical shapes in different stimulus conditions.

The stimuli were presented on an LCD screen (Eizo ColorEdge CG243W, Eizo Corporation, Hakusan, Japan; display area = 518.4 × 324.0 mm; resolution = 1,920 × 1,200 pixels, with 3.704 pixels/mm; color depth = 8 bits/channel) and were viewed through a mirror stereoscope (SA200 ScreenScope Pro; Stereo Aids, Albany, Australia; optical viewing distance = 400 mm; interocular distance = 64 mm). The size of each half image (720 × 1,200 pixels) corresponded to 194.4 × 324 mm on the screen.

In each trial, the subjects were asked to indicate the orientation of the normal (gauge-figure task; Koenderink & van Doorn, 1992) at one of 160 surface points (see Figure 10). Inputs were made by mouse and keyboard. The measurement points were evenly distributed in a triangular grid so that they fitted into the respective object area. Because the outlines of the various objects differed, the resulting number of triangular faces varied slightly between 272 and 275. At the rim of the objects, the inclination of the surface to the observer was maximal (i.e., its slant was 90°). To avoid trivial settings, the measurement points were located at least 5 pixels away from the rim. The gauge figure was highlighted in green (RGB[0, 0.98, 0.60]) and had a maximum base diameter of 24 pixels, a maximum rod length of 12 pixels, and a line width of 2 pixels.

A problem when using the gauge-figure method with stereoscopic stimuli is to decide at what depth the gauge figure should be positioned. If it is positioned at an arbitrary depth (e.g., at the image plane of the screen), this may make the adjustment more difficult, as the gauge figure may not necessarily appear to lie on the surface of the object but rather in front of or behind it. If it is positioned correctly on the actual surface of the object, its perceived depth may indirectly provide information about the shape of the object and thus interfere with the information provided by other cues. In order not to provide the subjects with stereoscopic cues to the depth of the gauge figure, we therefore presented it to the right eye only. In a preliminary experiment in which we used this monocular presenta-
tion mode, we did not observe the systematic overestimation of the perceived surface slant that was found by Bernhard, Waldner, Plank, Solteszova, and Viola (2016), who combined a gauge figure without disparity with stereoscopic stimuli.

To gain experience with the gauge-figure task, subjects performed numerous practice trials until they felt up to the task. The shapes of the objects shown in the practice phase differed from those used in the experiment. In the actual experiment, each subject performed 1,120 trials in a randomized order (7 stimulus conditions × 160 measurement points). This resulted in three repetitions (each by a different subject) of the 160 individual measurements belonging to each of the 84 combinations of seven object meshes and 12 stimulus conditions. In the 14 combinations involving the two base conditions Opq:Full and Trns:Mass:Full, three additional repetitions were made. On average, a subject required roughly 65 min. to perform all of the trials. Due to the additional practice phase and possible rests, the experiment was divided into multiple sessions depending on the speed of the subject.

After the experiment, the subjects were asked about their material and massiveness impression. To this end, they were presented with printed copies of stimulus images that showed an object similar to that used in the experiment under different stimulus conditions. The subjects were asked to indicate whether the respective object material appeared transparent or opaque to them. In addition, the subjects of the second group were asked whether the object shown in the respective conditions appeared solid or hollow.

Results

The subjects’ settings were analyzed in different ways that each focused on a specific aspect of the perceived shape and its deviation from the actual shape. We started by considering each setting as an independent measurement of local surface orientation. To evaluate the relative performance in the 12 stimulus conditions, we compared means across different local error measures observed under these conditions. In a second step, we reconstructed the surface shapes perceived by the subjects from their local settings. Here the error is given by the deviation of the reconstructed surface from the actual one. This more global approach allows us to distinguish between qualitative and quantitative errors and analyze systematic misjudgments of the local shape.

Since we found no systematic differences between the two base conditions Trns:Mass:Full and Opq:Full that were shared between the two groups of subjects, the redundant settings of the second group were discarded in order to maintain equal group sizes.

Systematic and random local errors

An obvious way to evaluate the local gauge-figure settings is to compute the normal error—that is, to directly compare the unit normal vectors indicated by the gauge figures with the veridical unit normal vectors (see Appendix D). A more sophisticated approach is to analyze the variance of the adjusted normals about the veridical ones and decompose it into accuracy and precision components, to distinguish between systematic and random errors (see Figure 11a). We calculated the total variance and this decomposition separately for each point of measurement, each object, and each stimulus condition. Figure 11b shows the pattern of the total normal variances across the different stimulus conditions, which closely resembles the pattern found for the normal error (compare Figure A2b). The variances are considerably higher for massive transparent objects than for hollow ones. The lowest variances can be found for opaque objects. In addition, the data indicate that in the transparent conditions, the accuracy variance tends to take a greater share of the total variance than the precision variance (roughly about 60%). Apart from the case of stimuli without texture, the relative contribution of accuracy and precision variance tends to be more balanced for opaque objects throughout (see Figure 11c, which explicitly shows the share of the accuracy variance of the total variance for different stimulus conditions).

The relative influence that each potential cue had on the settings is shown in Figure 12. It turns out that the negative influence of optical background distortions that we found for massive transparent objects is mainly due to an increase in the systematic error (i.e., the accuracy variance). In contrast, its slightly positive influence for hollow objects is caused by a rather equal decrease in both systematic and unsystematic variance. For both massive and hollow transparent objects, mirroring affected the accuracy variance to a larger extent than the precision variance. However, the influence was more positive in the massive transparent case than it was negative in the hollow case. Its positive effect was even stronger than the positive effect of the texture cue in the opaque case. In contrast to the transparent case, the mirroring cue had virtually no influence on shape perception in the opaque case. While the positive effect of absorption-induced darkening observed for massive transparent objects is due to a decrease of both systematic and unsystematic errors, the small effect for hollow objects is due to a decrease of the accuracy variance only.

Effect of contour information

Besides the shape cues that were manipulated in the experiment, the objects’ contour is always present as an additional shape cue. To examine how strong the
relative influence of the contour is, we plotted the size of the angular normal error (cf., Appendix D) against the distance of the measurement points from the contour. Figure 13 shows that in the case of opaque objects, the normal error is barely affected by the proximity of the measurement points to the object contour. For transparent objects, however, the normal error tends to increase with contour distance. This trend can be found for both massive and hollow transparent objects, where the difference in the error level is approximately constant for all contour distances. For very small contour distances, the error level found for hollow transparent objects almost decreases to the level found for opaque objects.

**Local slant and tilt errors**

Previous work has shown that subjects tend to underestimate the surface slant—that is, the angle between the perceived surface normal and the line of sight (e.g., Bernhard et al., 2016; De Haan, Erens, &...
To test for this effect and further potential differences between the stimulus conditions, we reparametrized the normals indicated by the gauge figure in terms of spherical slant and tilt coordinates (see Figure 14a). Stevens (1983) argued that this parametrization corresponds well to how the visual system represents the orientation of surfaces. The general pattern of slant and tilt errors (see Figure 14b) is similar to that found for the angular normal error (cf., Appendix D). However, the tilt error tends to be substantially larger than the slant error. The interpretation of surface tilt is difficult, if adjusted and veridical surface normals are both close to the line of sight, because then even small angular differences can lead to large tilt errors. However, for the convex objects used in the experiment, such cases are very rare and thus cannot explain the relatively large tilt errors. As expected, the slant bias was negative for every stimulus condition, which means that on average the slant was underestimated for both transparent and opaque objects (see Figure 14c). In addition, the degree of this underestimation differs between the stimulus conditions. The corresponding pattern of results is roughly inversely proportional to the previous error measures. On average, the underestimation of slant was most pronounced for massive transparent objects and smallest for opaque objects. This general underestimation of slant contradicts the results of Bernhard et al. (2016), who found a systematic overestimation of the perceived surface slant when a gauge figure without disparity was used with stereoscopic stimuli. In the tilt dimension, no systematic bias was found (i.e., $B_\tau \approx 0$ for each stimulus condition).

Reconstruction of perceived surfaces

One drawback of local error measures like the ones discussed so far is that they do not directly indicate what surface shape subjects perceived when they were presented with a specific object. In particular, it is
possible that surfaces that are perceived differently lead to identical averages of local errors. Consider, for example, a flat surface that is erroneously perceived as concave by one subject and as convex by another. Obviously, the average normal error of both subjects can nonetheless be the same. To allow for a more global interpretation of subjects’ settings, we integrated the local gauge-figure data into triangular meshes that were meant to reflect the perceived object shapes (see Figure 15a). This surface reconstruction involves adding a third dimension to the image space and assigning to each point of measurement a depth value that fits the data best (right). Because extreme gauge-figure settings with a slant value of 90° can lead to reconstructed surfaces with infinite depth expansion, we limited the range of the adjusted slant values so that \( \tilde{\sigma}_i = \min(\hat{\sigma}_i, 89°) \). Note that the reconstructed depth values are defined along the respective viewing directions of the surface points (black arrows). While different viewing directions run parallel to the \( z' \)-axis in the image space, they diverge in world space due to the perspective projection. (b) To compare the reconstructed surfaces with the veridical ones (right), we subsequently transformed them into the world space (left; see Appendix E for details). To this end, the reconstructed surfaces were anchored at a specific distance from the observer, assuming that their centers of gravity coincide with the respective veridical surfaces. This corresponds to the assumption that the subjects were able to accurately judge the overall distance of the objects. For the analysis of the data, the resolution and range of the veridical mesh were reduced to match those of the reconstructed surface.

**Qualitative and quantitative shape errors**

Adjusted and veridical surface shapes can be analyzed by comparing the depth differences between them (see Appendix F). However, there are cases where such measures do not adequately represent the goodness of shape perception. If, for example, convex and concave surface patches are correctly perceived as such but their strength of curvature is underestimated, adjusted and veridical surfaces can differ, although the type of local shape is judged correctly. Furthermore, as a consequence of a global surface reconstruction, large normal errors at one point can indirectly lead to large depth errors at points where the normal errors are actually lower. It therefore seems more sensible to compare adjusted and veridical surfaces in terms of their qualitative and quantitative shape similarity. To this end, we analyzed local shape indices and curvedness values, two measures that describe the type of local surface shape and the strength of local curvature, respectively (Koenderink & van Doorn, 1992). The calculation of the principal curvatures was done using a procedure proposed by Rusinkiewicz (2004).

One way to analyze the accuracy of the reconstructed shapes is to correlate adjusted and veridical shape index and curvedness values. The top row of Figure 16a shows the correlations between adjusted and veridical shape indices for the transparent and opaque base conditions. The correlation is highest for the opaque base condition (\( R = 0.75 \), right), considerably weaker for the hollow transparent one (\( R = 0.32 \), center), and almost absent in the massive transparent one (\( R = 0.13 \), left). If the correlation coefficients of all stimulus conditions are considered, it turns out that the resulting pattern is roughly inverted relative to the pattern of the normal error \( \Delta n \) (see Figure 16b, left).
A similar pattern can be found for curvedness (see Figure 16a, bottom, and 16b, right). The correlation is strongest for objects of the opaque base condition (\(R = 0.57\)) and substantially lower for the hollow and massive transparent base conditions (\(R = 0.21\) and \(R = 0.13\), respectively).

Another way to analyze the accuracy of the reconstructed shapes is to consider the error distribution of the shape index and curvedness values. We define the shape-index error as the absolute difference between the adjusted and the veridical shape indices (\(D_{\text{s}i} = |\hat{s}_i - s_i|\), where \(i\) denotes a specific vertex of the reconstructed surface). The larger \(D_{\text{s}}\), the more the local shape of the adjusted surface differs from that of the veridical surface. An error of \(D_{\text{s}} = 0.5\) occurs, for example, if a cylindrical surface is misjudged to be saddlelike (and vice versa) or if a convex (or concave) surface is misjudged to be cylindrical. While \(\Delta s < 0.5\) for 91.5% of the surface locations in the opaque base condition, this is only true for 78.4% and 67.7% in the hollow and massive transparent base conditions, respectively (see Figure 17, left). Accordingly, the average \(\Delta s\) is higher for massive transparent objects (\(\Delta s_{\text{Tms:Mass:Full}} = 0.439\)) than for hollow transparent (\(\Delta s_{\text{Tms:Holl:Full}} = 0.319\)) or opaque ones (\(\Delta s_{\text{Tms:Mass:Full}} = 0.180\); Figure 17, right). This pattern of results is similar to that found for the normal error \(\Delta n\) (cf., Appendix D). In contrast to the depth error (cf., Appendix F), the shape-index error \(\Delta s\) is only slightly reduced if the surface reconstruction is based on the mean adjusted normals \(\bar{n}\), instead of the individually adjusted normals \(\hat{n}_i\).

Analogous to the shape-index error, we defined the curvedness error as the absolute difference between the adjusted and veridical curvedness (\(\Delta c_i = |\hat{c}_i - c_i|\), where \(i\) denotes a specific vertex of the reconstructed surface). The results show that \(\Delta c\) is very similar in all massive and hollow transparent stimulus conditions (see Figure 18a). Differences mainly occur in the opaque stimulus conditions. While \(\Delta c\) is about 30% smaller in the opaque base condition than in the transparent conditions, the relative strength of the influence in the different opaque cue conditions is similar to that indicated by the normal error (cf., Figure A2). Like \(\Delta s\), \(\Delta c\) is only slightly reduced if the adjusted surfaces are reconstructed from the mean adjusted normals.

In order to detect any systematic under- or overestimation of curvedness, we further analyzed the curvedness bias (\(\text{B}_c = \hat{c}_i - c_i\), where \(i\) denotes a specific vertex of the reconstructed surface). As Figure 18b shows, the mean \(\text{B}_c\) is negative for all stimulus conditions. This indicates that subjects generally underestimated the objects’ curvedness. Although this underestimation occurs for both opaque and transparent objects, there are differences between the stimulus conditions. For example, the bias is more pronounced for hollow transparent than for opaque and massive transparent objects. If surfaces are reconstructed from the mean adjusted normals, \(\text{B}_c\)
Spatial error distribution

The evaluation of the gauge-figure settings and the quantities derived from them, such as the shape index, has so far been carried out mainly in a summarized form—that is, by averaging the corresponding parameters over the various objects shown in the experiment or individual points of measurement and the subjects. In contrast, we will now analyze the spatial distribution of these parameters at the level of individual objects. This approach makes it possible to identify specific image areas in which the subjects make larger errors than in others. In particular, increased systematic deviations from the veridical values in some regions can provide clues as to how the subjects used certain image information for their settings.

With respect to the normal error $\Delta n$, it can be seen that its spatial distribution across the surface is relatively inhomogeneous, especially for some of the massive transparent objects. As an example, Figure 19a shows the distribution for the object Obj2 (cf., Figure 8). While $\Delta n$ tends to be smaller in the peripheral areas of this object (cf., Figure 13), there is a ring-shaped area in the middle of the object where $\Delta n$ is considerably higher ($\Delta_{n_{\text{max}}}=102.4^\circ$). Near the center of this ringlike area the normal error is again much smaller ($\Delta n=45^\circ$). A separate analysis for the three subjects who adjusted the gauge figures in this case reveals that only two of them show this specific distribution of $\Delta n$ (see Figure 19b, subjects AEMA and DUUN). The distribution for the remaining subject is much more homogeneous, and the overall error level smaller (see Figure 19b, subject RARA). Apparently, the high absolute precision variance that we found for massive transparent objects (cf., Figure 11b) has a systematic cause itself, namely that different subjects perceived the shape of the object differently.

To investigate how the subjects perceived the shape of object Obj2, we compared the individual surfaces reconstructed from their settings (see Figure 20, top row). The results suggest that the two subjects with an inhomogeneous spatial distribution of $\Delta n$ (AEMA and DUUN) saw a surface that has an indentation near the object’s center, whereas the remaining subject with a spatially more homogeneous $\Delta n$ (RARA) correctly perceived a bulge in the corresponding area. In the other cue conditions, some of the subjects also saw an indentation, except in the cue condition in which background distortions were missing (Dist−). Systematic misinterpretations of local shape also occurred with other objects. This is, for example, clearly visible for Obj3 (see Figure 20, bottom row). Here, two of three subjects also erroneously judged a convex surface area in the middle of the object to be concave. In contrast to Obj2, this systematic misjudgment also occurred when background distortions were missing (Dist−), whereas it did not occur when darkening was omitted (Dark−). We will discuss possible explanations for these systematic misinterpretations of the surface shape in more detail in the Discussion.

Material and massiveness ratings

Figure 21 shows the results of the follow-up survey. In both the opaque and the hollow transparent base conditions (Opq:Full and Trns:Holl:Full, respectively), the material of the example object was correctly identified by all subjects. In the massive transparent base condition (Trns:Mass:Full) this is true for only 90.5% of the subjects. In the transparent case, the largest number of misclassifications occurred for the massive object without mirror reflections.
and the hollow object without background distortions (Trns:Holl:Dist—). In these cases, 47.6% and 42.9% of the subjects incorrectly judged the sample object to be opaque. In the opaque case, only the completely reflecting object (cue condition Mirr+) was not perceived as opaque by all subjects; 23.8% incorrectly classified it as being transparent.

In the three base conditions, the massiveness was correctly identified by 47.6% of the subjects for the massive transparent object, 76.2% for the hollow transparent one, and 95.2% for the opaque one. In the cue conditions of the hollow transparent case, the proportion of correct massiveness estimates was either equal to or slightly lower than those in the corresponding base condition. The lowest value was found in the cue condition without reflections (Mirr—): Only 52.4% of the subjects correctly recognized that object as being hollow. The completely reflecting opaque example object (cue condition Mirr+) was incorrectly identified as being hollow by 33.3% of the subjects.

**Discussion**

In this experiment, we presented subjects with stereoscopic images of randomly shaped transparent objects, either hollow or massive, and asked them to indicate the orientation of the normal at various surface points. We varied the availability of three potential shape cues (background distortions due to refraction, darkening due to absorption, and mirror images due to specular reflections) by altering either...
scene and material properties or the image generation. For comparison, we also presented subjects with opaque objects and varied the availability of corresponding shape cues.

Our computational analysis revealed that the potential shape cues in the transparent case are particularly complex. The main reason is that they not only are related to the shape of the object but also depend on properties of the whole scene. It is therefore not surprising that the shapes of transparent objects were judged less accurately than those of opaque ones. On average, the errors made for transparent objects were approximately twice as large as those made for opaque objects. This suggests that the visual system processes shape-related image information differently in the transparent case than in the opaque case. We also found that hollow transparent objects were perceived considerably more accurately than massive transparent objects. This result is also not surprising, as our analysis has shown that massive objects often lead to

Figure 20. Analysis of intersubject differences in the perceived shape of two massive transparent objects (top row: Obj2; bottom row: Obj3). Left to right: The respective stimulus images (right eye only, trimmed), the shape indices of the surfaces that were reconstructed from the gauge-figure settings of the respective subjects, and the veridical shape indices of the object meshes. Both the reconstructed and veridical surfaces are shown with the same perspective projection as the stimulus images shown at the left.

Figure 21. Results of the follow-up survey, in which subjects were asked to indicate the material and massiveness of an example object shown in different stimulus conditions. Massiveness ratings were performed only by subjects of the second group, to whom hollow objects were shown during the experiment. The ratings are based on printed copies of stimulus images. Furthermore, all ratings refer to the same example object and not to the objects actually seen by the subjects in the experiment. (a) Stacked bar plot showing the relative frequency of the material ratings for each stimulus condition, averaged across all subjects. (b) Stacked bar plot showing the relative frequency of the massiveness ratings for each stimulus condition, averaged across all subjects of the second group.
much larger background distortions than hollow objects of identical shape. From a computational point of view, the estimation of shape therefore appears to be much more complex in such situations.

For both massive and transparent objects, shape perception was influenced by the three manipulated potential shape cues. Depending on whether the objects were solid or hollow, some of the potential shape cues had opposite effects. For massive transparent objects, inclusion of both specular reflections and absorption enhanced shape perception, whereas for thin-walled objects of the same outer shape, absorption had almost no effect, and specular reflections even had negative effects. The effect of background distortions was very small for hollow transparent objects and negative for massive ones. These results indicate that in the transparent case, the particular influence of an image regularity appears to depend more on the specific situation than in the opaque case. A possible explanation of this finding is that the visual system can use the image regularities for shape perception only if they remain within certain bounds. An interesting question for further research would be to examine more explicitly how the accuracy of shape perception depends on the specific manifestation of the different image regularities—for example, the degree of optical distortion.

The outer contour of an object seems to play a similar role for shape perception in both the opaque and the transparent case. The closer to the rim of an object, the more similar the error levels are for both material classes. This is not surprising, because the outer contour of an object—and thus the shape information provided by the corresponding edge in the image—does not depend on the material. With increasing distance from the rim, the contribution of other shape cues appears to become more dominant. At least in the transparent case, however, these cues seem to be less reliable than the contour.

Although the specific role that an image regularity plays for perceiving the shape of transparent objects appears to depend on a complex interplay of properties of the object itself and its surround, our results provide a first insight into how different image regularities are processed by the visual system. However, the interpretation of the results is by no means trivial. This applies not only to those cases in which a certain image regularity had a positive effect on shape perception but also to those cases in which no effect was visible or shape perception was negatively influenced. In the following, we will briefly discuss these different patterns of results and the interpretations they are compatible with. In particular, there is evidence that some image regularities were used in a way that is not suitable for estimating the shape of transparent objects.

As a first example, we consider cases where cues had no effect on shape perception. This was the case with mirror images caused by opaque objects and absorption-induced darkening caused by hollow transparent objects. This appears to indicate that these two cues were not used for shape perception. However, it is also possible that these image regularities actually served as a shape cue but that their influence was not noticeable because they provide shape information that is consistent with the information already provided by other shape cues.

If the presence of an image regularity has a positive effect on shape perception, the interpretation may appear less equivocal. However, even in this case there are several ways to interpret the results. This can be illustrated by considering the darkening caused by absorption that had a positive influence on the perception of massive transparent objects. One interpretation is that the visual system actually used the thickness information provided by the darkening to estimate the shape. An alternative explanation is that the darkening has only an indirect positive effect by enhancing the influence of another shape cue. For example, the chromaticity and brightness changes caused by absorption might have reduced the visibility of higher order reflections and increased the visibility of the particularly informative first-order reflections.

The presence of background distortions with massive transparent objects and mirror images with hollow transparent objects had a clearly negative effect on shape perception. Again, there are various possible interpretations for these observations. First, the two image regularities might not serve as shape cues themselves but instead have only an indirect (negative) influence on other shape cues. For example, the strong background distortions of massive transparent objects could have made it difficult to detect the mirror images, which themselves had a clearly positive influence on shape perception. This explanation is not implausible, because background distortions and mirror images are usually reflected in the image in a similar way. Both lead to spatially varying, direction-dependent magnifications and compressions of the background or the mirrored environment. If two such regularities are superimposed in the image, using one of them as a shape cue (in this case, the mirror images) could be impeded.

Conversely, a negative influence of an image regularity on shape perception does not necessarily mean that it is not used by the visual system as a shape cue. The visual system might refer to the regularity, but not in a way that is appropriate for estimating the shape correctly. One reason for this could be that the image regularity exists in a form that the visual system cannot process in a meaningful way. As outlined already, the visual system might, for example, be able to use optical distortions of individual refractive
surfaces as a shape cue, but not the much more complex distortions of objects with multiple refractions. If the same mechanisms are nevertheless also used in more complex situations, this could lead to errors and negative effects on shape perception. Apparently, such an erroneous use of image information is preceded by what can be regarded as a misinterpretation of the information available in the image. In the current example, for instance, distortions caused by multiple refractions might have been misinterpreted as distortions of a single refractive surface. Our results provide some evidence that such confusions do not have to be limited to the same material category. In our local analyses of the adjusted surface shapes, we described several cases in which systematic, spatially limited deviations from the veridical shape occurred. In three of four cue conditions of the objects Obj2 and Obj3, two out of three subjects erroneously perceived a convex bulge in the middle of the objects as a concave indentation (see Figure 20). In both cases, the deviating settings are compatible with the interpretation that the subjects at least partially interpreted the available image information as if it had been caused by opaque objects (see Appendix G for a detailed discussion on this topic).

**General discussion**

In this work, we dealt both theoretically and empirically with the visual perception of the shape of transparent objects. On the theoretical level, we analyzed how the shape of transparent objects is reflected in properties of the (retinal) image. In particular, we considered several image regularities associated with shape that are specific for transparent objects: optical distortions of the background caused by light refraction, changes in chromaticity and brightness caused by absorption, and distorted mirror images of the environment caused by specular reflections at each surface that separates spatial regions with different refraction indices. Our computational analyses showed that the relationship between these image regularities and shape are often substantially more complex than in the opaque case. Furthermore, the analyses showed that the common problem that cues interact with each other and cannot be considered in isolation occurs in a particularly pronounced form with transparent objects.

The substantial differences in shape-related information that is available in the images of transparent and opaque objects strongly suggest that shape perception works differently for objects from these two material classes. This raised the questions of how well the visual system can recognize the shape of transparent objects at all and how well this is possible in comparison to opaque objects. A further question was whether the potential shape cues that were identified in our theoretical analysis actually play a role in shape perception. To investigate these questions, we conducted an experiment in which we used a gauge-figure task to measure the accuracy and precision of shape perception depending on the availability of potential shape cues. Our results show that subjects’ settings in the transparent case were both less accurate and less precise than in the opaque case. Furthermore, the influence of individual image regularities in the transparent case was sometimes opposite, depending on whether they originated from massive or hollow objects.

These observed differences between solid and hollow objects are a consequence of the fact that for transparent objects, it is not only the material and the visible part of the outer shape that are crucial for shape perception but also the surfaces that are not directly visible and the properties of the interior. The interior could in principle be of almost arbitrary complexity. Thus, the massive and hollow objects used in the experiment must be considered as just two exemplars from the set of all possible objects that, given a specific outer shape, can be defined by varying the wall thickness and the number of enclosed surfaces, not as representatives of two disjunct object classes.

The present results also indicate that the influence that image regularities have on shape perception can vary greatly within the set of possible objects. Our finding that some transparent objects were misinterpreted as opaque objects suggests that the differences related to the object type are partly due to the fact that in some cases strategies that are only suited for shape perception in the opaque case were applied to transparent objects (cf., Appendix G). It seems that a general model of shape perception in the transparent case must also include a model about how the object type influences the use of specific image regularities. A preliminary step for solving this general problem would be to identify for each object type the ranges of relevant object and scene parameters in which the regularities can be used at all. A pragmatic approach for determining such parameter ranges would be to determine empirically for each situation whether and in which parameter ranges shape perception works with acceptable accuracy. Based on such data, more abstract principles could be identified that determine these parameter ranges. On the basis of our theoretical analysis of background distortions, it seems plausible to assume that the absence of intersections of light rays along their path is such an abstract principle.

Although our results indicate an influence of some image regularities on the perception of shape, it is difficult to tell in which way this happens. This applies
in particular to the question whether the respective image regularities were actually used as a shape cue. Especially with massive transparent objects, our results provide some evidence that the image regularities caused by refraction and absorption were partially misinterpreted as opaque shape cues. In addition, some image regularities might have had only a moderating effect on other shape cues. Therefore, a primary goal of subsequent studies should be to obtain further information about how certain image regularities influence the shape perception of transparent objects. For this purpose, cases in which the perceived shape deviates markedly from the veridical shape are particularly diagnostic. Different types of deviations promise specific insights into the mechanisms underlying shape perception but also require different methodological approaches.

One type of deviation results from unsystematic errors. This means that the shape parameters set by the subjects scatter randomly around the veridical shape parameters. In the cases we investigated, the size of such deviations varied considerably depending on the image regularities that were present. Although different degrees of unsystematic errors are partly compatible with several interpretations, a deeper analysis of the pattern of such errors can nevertheless help to obtain further information about the respective role of the image regularities under scrutiny. If an image regularity is actually used as a shape cue, then it is to be expected that the size of the unsystematic error depends on the amount of usable information the cue provides.

Variations in the size of random errors can then be indicative of the degree of uncertainty in estimating the shape based on this cue. Accordingly, intentional changes in the amount of information should lead to corresponding changes in the error level. This could be achieved by varying the background texture, the illumination, or the absorption properties of the transparent material. For example, the contrast of the background texture could be altered, or more localized light sources could be used instead of complex ambient illumination. If the systematic errors are reduced in situations that presumably have a higher information density, this would provide further evidence that the corresponding image regularity is actually used as a shape cue. In addition, such manipulations are a proper means of identifying the parameter ranges within which the individual image regularities can be used as a shape cue by the visual system.

Not only unsystematic errors are diagnostic for the mechanisms underlying shape perception—so are systematic errors. Examples are the systematic misinterpretation of the local surface shape of two massive transparent objects that we found in the present investigation. A possible interpretation of these deviations is that the visual system did not use the image regularities properly but instead misinterpreted image regularities caused by refraction and absorption as originating from opaque objects and accordingly used them in a way that would only be appropriate for inferring the shape of opaque objects. In future work, the validity of this explanation should be examined in more detail. A first interesting pair of questions is how regular these misinterpretations are and whether the perceived local shapes can actually be divided into only two disjunct categories. Using a larger number of subjects might help to obtain more reliable information about the frequency of those misinterpretations. In addition, an attempt should be made to model the misinterpretations more precisely. This could be based on a more detailed analysis of the information and regularities available in the image. On this basis, more specific hypotheses could be generated as to which aspects of the image could be responsible for the misinterpretations. Ideally, it should be possible to make precise predictions as to which shape an image regularity would be compatible with if it were incorrectly interpreted as an opaque shape cue. These hypotheses could then be verified by deliberately manipulating these regularities in the image. In the cases observed in the present investigation, for example, the plausibility of an opaque interpretation of the background distortions could be increased by omitting all other transparency-specific image regularities. The proportion of subjects with systematic misinterpretations of the shape should then increase. An opposite strategy would be to remove image aspects that supposedly trigger the misinterpretation; the systematic misinterpretations should then no longer occur. Alternatively, the stimuli could be modified in order to contain additional cues to the material. Such cues can arise from, for example, movements of the object, parts of the scene, or the observer. The dynamics and their regularities in the image not only are known to influence the activation of transparency-specific mechanisms (Kawabe & Kogovšek, 2017; Kawabe et al., 2015; Tamura, Higashi, & Nakauchi, 2018) but could also provide cues to the shape itself (e.g., Ben-Ezra & Nayar, 2003). From a theoretical point of view, it therefore appears promising to also consider moving stimuli and to analyze their shape-related image regularities in more detail.

In the present experiment, we used stereoscopic stimuli to simulate natural viewing conditions. It is an interesting question whether and how binocular disparity that is available in this case influences the processing of individual image regularities and shape perception in the transparent case. A natural assumption seems to be that it contributes positively to shape perception, as in the case of opaque objects (Doorschot, Kappers, & Koenderink, 2001). However, it is also possible that the multiple disparity patterns related
to different image regularities within mirror images or background distortions are too complex to be useful for the visual system. It is even possible that these complex disparity patterns have a negative effect on shape perception by hampering the evaluation of other shape cues. To investigate the actual influence of binocular disparity in the transparent case, it would be interesting to compare the results obtained with stereoscopic and monoscopic stimuli. We are currently replicating the present experiment with such monoscopic stimuli—that is, we are presenting the same images to both eyes. Preliminary results of this study suggest that the overall effect of this change of viewing mode on shape perception is not very large, in the sense that the general pattern of results is rather similar (see Figure 22). However, there are also noticeable changes in the direction and strength of some effects. For instance, including the distortion cue in the case of massive transparent objects decreases the normal error slightly in the monoscopic case, whereas it leads to an increase of the error in the stereoscopic case.

A major challenge in investigating the perception of the shape of transparent objects is that there are numerous interactions between different image regularities associated with shape. Due to these interactions, it was difficult in our experiment to decide whether an image regularity had an effect because it was actually used as a shape cue or merely because it had an indirect positive or negative influence on other image regularities. Although such interactions complicate conclusions about individual image regularities, they also open up further methodological options for investigating the role that these image regularities play. The basic idea is to manipulate different aspects of certain image regularities separately from each other. This includes, on the one hand, the aspect associated with the shape (shape-related part) and, on the other hand, the aspect indirectly influencing other shape cues (moderating part). A major advantage of this approach is that shape-related and moderating parts can largely be described and analyzed at the level of image generation. The hypothesis that a certain image regularity has only an indirect effect without serving as a shape cue itself could then be tested empirically by removing its shape-related part but preserving its moderating part. Using the example of absorption-induced darkening, this would mean that the higher order reflections are darkened in the image but with a constant degree of darkness across the stimulus, so that it does not provide systematic information about material thickness.

Conclusions

Our empirical results have shown that subjects can at least approximately recognize the shape of transparent objects. We have analyzed several image regularities related to transparent objects with respect to their suitability as shape cues. All of these potential cues had, at least in some cases, a noticeable effect on the accuracy and precision of shape perception. However, the shape perception of transparent objects was substantially worse than that of opaque objects. In addition, performance depended on the actual object type and was considerably better for thin-walled hollow objects than for massive ones. We discussed several strategies for how the analysis of local and global errors may be used to derive more specific hypotheses about the role played by individual image regularities in shape perception.

Keywords: perceptual transparency, shape perception, material perception

Acknowledgments

This research was supported by Deutsche Forschungsgemeinschaft Grant FA 425/2-2 (F. F.).

Commercial relationships: none.
Corresponding author: Nick Schlüter.
Email: schluerer@psychologie.uni-kiel.de.
Address: Institut für Psychologie, Christian-Albrechts-Universität zu Kiel, Kiel, Germany.

References


Electrical and Electronics Engineers, https://doi.org/10.1109/VISUAL.1995.480795.


---

**Appendix A: Cues from background distortions due to refraction—generalization and open questions**

**Estimation of distortions from the image**

Under normal viewing conditions, veridical optical magnifications are not directly accessible to an observer. The question therefore arises whether optical distortions can be estimated from the retinal image alone. Provided that the background has sufficient structure and contrast to reflect the optical distortions, they are visible in the image as direction-dependent variations in texture density, where optical magnifications tend to produce lower texture densities than optical compressions. The minimum and maximum texture densities $D_{\min}$ and $D_{\max}$ in a small region around a specific location in the image could therefore be used as estimates of the corresponding maximum and minimum optical magnifications $M_{\max}$ and $M_{\min}$, respectively. As is shown in the text, the latter are closely related to the shape index $s$ and curvedness $c$, which in turn uniquely determine the maximum and minimum principal curvatures $K_{\max}$ and $K_{\min}$.

However, inferring optical magnifications from texture densities is difficult for several reasons. First, texture-density variations in the image cannot always be attributed to differences in optical magnification. They may also be caused by properties of the background itself (e.g., variations of its intrinsic texture density). Reliable inferences therefore require that either the background be sufficiently regular or that further information about the undistorted background be available. Second, it is obvious that a certain texture density in the image can only be interpreted as a...
magnification or a compression if the texture density corresponding to the undistorted state is known. If parts of the undistorted background are directly visible at another point of the image or at another point in time, this information could be used as a reference (Fleming, Jäkel, & Maloney, 2011), provided that the background is sufficiently spatially and temporally homogeneous. If the background is entirely located behind the transparent material, as in the case of the water surface analyzed in the text, one can refer only to image statistics—for example, estimating the undistorted background texture density by the mean value of minimum and maximum texture densities across the image.

Orientation maps as a potential shape cue

Chen and Allison (2013) have suggested that the shape of transparent objects could be estimated based on the orientation maps of the background distortions they cause. Orientation maps were initially proposed by Fleming, Torralba, and Adelson (2004) as a shape cue for opaque objects (see also Fleming, Holtmann-Rice, & Bülthoff, 2011). An orientation map describes spatial image structures (irrespective of their underlying cause) by indicating at each pixel the dominant orientations of the texture (i.e., in the notation used in the foregoing, the directions of $D_{\text{min}}$ and $D_{\text{max}}$) and the direction-related strength of the texture density (i.e., the relative size of the texture density in different directions). Chen and Allison surmised that orientation maps could serve as a shape cue in the transparent case also, because the image distortions caused by transparent and specularly reflecting opaque objects are similarly related to object shape. In both cases, larger surface curvatures cause larger distortions in the mapping of the environment or background to the image. At present, it is an open empirical question whether orientation maps actually play a role in the perception of transparent objects. However, there is some reason to be skeptical, because according to the derivation outlined here, orientation maps do not contain sufficient information for a successful estimation of object shape: While the dominant orientation of the texture at a certain image point indicates the orientation of the minimal curvature, information about the sign and absolute size of the principal curvatures is lacking. The direction-dependent strength of the texture, which is estimated from the minimum and maximum local filter responses, provides information only about their relative size, and this is not sufficient to estimate the shape type (i.e., the shape index) of a surface. This means, for example, that convex surface areas cannot be distinguished from concave ones.

Intrinsic versus observer-dependent shape

So far, we have assumed that optical distortions depend solely on the local curvature of a surface. However, under natural viewing conditions this is generally not the case. If the observed surface is not oriented perpendicular to the viewing direction, factors related to the optical projection may also play an important role. As a consequence, magnifications in the image may be less correlated with the intrinsic curvature of a surface than with its curvature relative to the observer—that is, with the rate at which the surface orientation changes “in the image.” At places where the surface is strongly inclined to the viewing direction—for example, near the rim of an object—it is therefore difficult to infer the intrinsic curvature of the surface from magnifications in the image. A similar issue was highlighted by Fleming et al. (2004) in the case of mirror objects. It is not yet clear what the consequences of this dependency are on the observer or how an estimate of the intrinsic, observer-independent shape of a surface can be gained from viewpoint-dependent information.

Appendix B: Analysis of absorption-induced darkening

See Figure A1.

Appendix C: Cues from changes in intensity due to absorption—generalization and open questions

Hollow objects

With hollow objects, the correlation between darkening caused by absorption and object thickness is normally considerably weaker, because light that passes through such objects is refracted more often and only some sections of its path inside the outer surface of the object lead through the object’s material. The length of these sections can deviate so much from the object thickness that the darkening due to absorption does not provide any useful information about it. Thus, the best strategy seems to be to simply ignore the darkening as a thickness indicator. How hollow objects can be identified is a separate problem. The identification could, for example, be based on very weak background distortion; however, this would be only a rough heuristic, because massive objects can also cause weak
background distortions, for example if they are rather thin or only slightly curved.

**Integration of thickness information**

In order to serve as a shape cue, the thickness information obtained from the darkening in the image would have to be integrated into a specific object shape. This is difficult because the thickness indicated by the darkening is compatible with an arbitrary number of pairs of different front and back surface shapes. For example, thickness information is invariant against an exchange of the front and back surfaces. These ambiguities could be reduced by general heuristics about object shapes—for example, shapes that are symmetrical in depth may be preferred over asymmetrical shapes. Another way to reduce the space of compatible shapes would be to integrate the thickness information with the information from other shape cues. If these other cues provide sufficient information about the front shape of a transparent object, absorption-induced darkening could then in principle be used to estimate the rear shape.

Furthermore, it is possible that darkening additionally (or even exclusively) plays a “passive” role in the perception of shape—that is, that it indirectly influences other potential shape cues. For example, more absorptive materials can reduce the visibility of the background and internal mirror images while enhancing the visibility of the reflectance on the front side. However, it is difficult to predict the influence of such interactions on shape perception.

**Appendix D: Analysis of the angular normal error**

To compare the gauge-figure settings with the expected settings given perfect shape perception, we computed the angular difference between the unit normal vector indicated by the gauge figure and the veridical unit normal vector (see Figure A2a). Figure
Figure A2. Analysis of the angular normal error. (a) The normal error $\Delta n$ is defined as the (unsigned) angular difference between the unit normal vector indicated by a gauge-figure setting $i$ ($\hat{n}_i$, yellow) and the veridical unit normal vector ($n_v$, green). Note that both $n$ and $\hat{n}$ are defined with respect to the image plane. Thus, $n$ is not identical to the veridical-surface normal in the world space. (b) Angular normal error ($\pm 95\%$ confidence intervals) for each stimulus condition, averaged across all objects, points of measurement, and subjects. The error levels of the base conditions (Trns:Mass:Full, Trns:Holl:Full, and Opq:Full) are emphasized by dashed horizontal lines. Due to the restricted adjustment range of the gauge figure, the maximum averaged normal error $\Delta n_{\text{max}}$ that would occur if the subjects made the greatest possible error for each of their settings depends on the distribution of the veridical-surface normals. For the stimuli used in this experiment, $\Delta n_{\text{max}} = 112.52^\circ$. If the subjects’ settings were random, the expected averaged normal error would be $\Delta n_{\text{random}} = 90^\circ$.

A2b shows that this normal error $\Delta n$ is considerably higher for the massive and hollow transparent base conditions ($\Delta n_{\text{Trns:Mass:Full}} = 38.74^\circ$ and $\Delta n_{\text{Trns:Holl:Full}} = 28.39^\circ$, respectively) than for the opaque base condition ($\Delta n_{\text{Opq:Full}} = 14.11^\circ$). In the transparent case, the omission of background distortions and mirroring has opposite effects for massive and hollow objects. Without background distortions (cue condition Dist$-$), the normal error decreases by 3.08° for massive objects but slightly increases by 1.10° for hollow ones. In contrast, the normal error increases by 8.96° when mirroring is omitted for massive objects (Mirr$-$), while it decreases by 3.15° for hollow ones. The omission of absorption-induced darkening (Dark$-$) has negative effects for both the massive and hollow cases (2.19° and 0.56°, respectively). Although the normal error is lowest in the opaque base condition, it increases to a similar extent to that found for hollow transparent objects when information from texture is missing (Tex$-$) or the object is made of a fully reflective material (Mirr$+$). In contrast to the transparent case, the omission of mirroring (MIRr$-$) has no effect on normal error in the opaque case.

Appendix E: Transformation of reconstructed surface from image space to world space

To convert the reconstructed surfaces from the image space into the world space, we transformed both their vertices and their normals. During the transformation, the normals were considered not as location-independent directions but as vectors originating from a specific vertex. To this end, we treated the tips of the normal vectors as regular points and transformed them in the same way as the vertices.

In the first step, we translated the vertices $v$ and the normal tips $t$ so that the zero point of the coordinate system corresponds to the location of the virtual camera. In addition, we reversed the $z$-axis of the coordinate system:

$$v' = \begin{pmatrix} v_x - (w/2) \\ v_y - (h/2) \\ -v_z + (d/r) \end{pmatrix}$$ (A1)

$$t' = \begin{pmatrix} t_x - (w/2) \\ t_y - (h/2) \\ -t_z + (d/r) \end{pmatrix}.$$ (A2)

Here, $w = 839$ pixels and $h = 1,200$ pixels are the width and height of the stimulus image, respectively; $d = 400$ mm corresponds to the viewing distance of the camera; and $r = 0.27$ mm/pixel corresponds to the resolution of the LCD screen used in the experiment.

In the next step, we defined camera ray vectors $c_v$ and $c_t$ that point to the locations where the vertices and the normal tips are projected on the image plane, respectively:

$$c_v = \begin{pmatrix} v'_x \\ v'_y \\ d/r \end{pmatrix}$$ (A3)

$$c_t = \begin{pmatrix} t'_x \\ t'_y \\ d/r \end{pmatrix}.$$ (A4)

The world-space coordinates of the vertices and the normal tips were then calculated by adjusting the length of the camera ray vectors to the $z$ component of the transformed vertex and normal-tip vectors:

$$v'' = (c_v/||c_v||) \times v'_z$$ (A5)

$$t'' = (c_t/||c_t||) \times t'_z.$$ (A6)

After this, the $z$-axis was reversed and the zero point of the coordinate system was changed to correspond to the center of the object. In addition, the unit was...
Appendix F: Analysis of surface similarity

To compare adjusted and veridical surfaces, we first analyzed the depth differences between them, which we call $\Delta d$ (see Figure A3a). In contrast to the local error measures discussed previously, the largest depth errors occur not for massive but for hollow transparent objects (see Figure A3b, black points). Furthermore, the variances found in these conditions are substantially higher than for most of the remaining conditions. To determine whether this is due to the susceptibility of the reconstruction algorithm to extreme individual gauge-figure settings, we also analyzed the depth error for surfaces that were reconstructed from the averaged adjusted normals $\hat{n}_i$ instead of the individual adjusted normals $\hat{n}_{ik}$. The underlying idea is that the reconstruction algorithm delivers more reliable results if the reliability of its input increases. In fact, the depth errors of the surfaces reconstructed in this way are smaller in every stimulus condition. Furthermore, the depth error for hollow transparent objects is now smaller than that for massive ones (see Figure A3b, blue points). This contrasts with the results just discussed but is consistent with the results pattern we found for the normal error $\Delta n$ (cf., Figure A2b). Another indication that the deviating results pattern of individually reconstructed surfaces is mainly due to the susceptibility of the reconstruction to outliers is that the depth deviation between the differently reconstructed surfaces is most pronounced for hollow transparent objects. While one could assume that this is due to the fact that the precision of the subjects is particularly low in these conditions, the variance decomposition of the normal error shows that this is not the case. Instead, the precision variance for hollow transparent objects is lower (i.e., the precision was higher) than for massive transparent objects. This shows that an analysis based on individually reconstructed surfaces runs the risk of overestimating the actual depth errors. There exist other measures for the difference between two surfaces, such as the modified Hausdorff distance (Dubuisson & Jain, 1994), that presumably are less susceptible to outliers. However, an analysis of this measure showed the same pattern of results that we found for the depth error $\Delta d$.

Appendix G: Erroneous use of shape cues from the wrong material category?

As a first example we consider object Obj2. Our results suggest that the local shape errors can be attributed to the influence of background distortions, because such errors did not occur in situations without distortions (i.e., cue condition Dist–). A possible explanation for the systematic shape errors observed in this case is that the background distortions of the transparent object have been at least partially misin-
interpreted as the distorted texture of an opaque object. This would not be implausible, because optical distortions of the background and shape-induced variations of the texture density of opaque objects are hard to discern in the image. Both lead to spatially varying, direction-dependent magnifications or compressions, in one case of the background and in the other case of the surface texture. If Obj2 is presented as a massive transparent object, there is a ring-shaped area around its center along which there are strong directional distortions of the background (see Figure A4, left). In the case of textured opaque objects, such an image regularity would be expected at image regions where the inclination of the surface to the observer (i.e., its slant) is particularly high (see, e.g., Fleming, Holtmann-Rice, & Bülthoff, 2011). Without further information, however, the local orientation of the surface (i.e., its tilt) is ambiguous. The surface along the ring-shaped distorted area could therefore be inclined either inward or outward. In the first case, the area enclosed by the distortions would be a concave indentation; in the second case, a convex bulge.

Figure A4. Possible misinterpretation of background distortions of a transparent object as shape-induced variations of the texture density of an opaque object. The left side shows a massive transparent object used in the experiment (Obj2; right eye only, trimmed). In the center of the object can be seen a ring-shaped area, along which the background is distorted (cyan). Next to the stimulus, this image regularity is shown in isolation. In the case of textured opaque objects, such an image regularity would be expected at image regions where the inclination of the surface to the observer (i.e., its slant) is particularly high (see, e.g., Fleming, Holtmann-Rice, & Bülthoff, 2011). Without further information, however, the local orientation of the surface (i.e., its tilt) is ambiguous. The surface along the ring-shaped distorted area could therefore be inclined either inward or outward. In the first case, the area enclosed by the distortions would be a concave indentation; in the second case, a convex bulge.

Figure A5. Possible misinterpretation of the darkening of a transparent object caused by absorption as shading of an opaque object. The left side shows a massive transparent object used in the experiment (Obj3; right eye only, trimmed). In the center of the object, the background is markedly darkened due to absorption (cyan dashed area). Next to the stimulus, this image regularity is shown in isolation. If the visual system is unable to correctly identify the absorption-induced darkening as such, it might misinterpret it as the shading of an opaque object. In the opaque case, such a darkening would be expected, for example, in strongly concave and correspondingly shaded surface areas.

A misinterpretation of background darkening as shading of an opaque object could also be responsible for the systematic shape errors we found with object Obj3 (see Figure A5). Here, two out of three subjects in three of four cue conditions also erroneously perceived convex bulges in the middle of the object as concave indentations. In this area, the background has been darkened by absorption. The systematic shape errors did not occur if the object had no absorption that darkened the background (cue condition Dark−–). The potential misinterpretations of image information outlined so far are not necessarily the only ones that have occurred. For both Obj2 and Obj3, it is also possible that parts of the bright or strongly saturated areas with total reflections were misinterpreted as mirror images or highlights caused by ordinary surface
reflections. Some of these total reflections are located exactly at the border of the area where the systematic shape errors occurred (compare Figure A4, left, and A5, left). Mirror images or highlights with a similar appearance usually occur in strongly curved, convex areas and thus indirectly support the interpretation of a concave indentation.

Whether the convex bulge set by the remaining subjects was based on an opaque or transparent interpretation of the image information cannot easily be determined. Based on the current results, it is also difficult to judge whether these misinterpretations are merely a spatially limited phenomenon or even a general characteristic of the perception of transparent materials. To answer this question, more detailed predictions for potential opaque misinterpretations would be required. For the complex-shaped objects with multiple reflections and refractions used in this experiment, such predictions can be ambiguous and can partly overlap with those for transparent interpretations.