Material category of visual objects computed from specular image structure

Alexandra Schmid, Pascal Barla, Katja Doerschner

▶ **To cite this version:**

Alexandra Schmid, Pascal Barla, Katja Doerschner. Material category of visual objects computed from specular image structure. Nature Human Behaviour, 2023, 7 (7), pp.1152-1169. 10.1038/s41562-023-01601-0 . hal-04173993v2

HAL Id: hal-04173993
https://inria.hal.science/hal-04173993v2

Submitted on 6 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

TITLE:

Material category of visual objects computed from specular image structure

AUTHOR LIST:

Alexandra C. Schmid[1]* (email: Alexandra.Schmid@nih.gov; *Corresponding author)

Pascal Barla[2]

Katja Doerschner[1,3]

AFFILIATIONS:

[1]Department of Psychology, Justus Liebig University Giessen, Germany

[2]INRIA, University of Bordeaux, France

[3]Center for Mind, Brain and Behavior, Justus Liebig University Giessen & Philipps University Marburg, Germany

April 12th 2023

ABSTRACT

Recognising materials and their properties visually is vital for successful interactions with our environment, from avoiding slippery floors to handling fragile objects. Yet there is no simple mapping of retinal image intensities to physical properties. Here, we investigated what image information drives material perception by collecting human psychophysical judgments about complex glossy objects. Variations in specular image structure – produced either by manipulating reflectance properties or visual features directly – caused categorical shifts in material appearance, suggesting that specular reflections provide diagnostic information about a wide range of material classes. Perceived material category appeared to mediate cues for surface gloss, providing evidence against a purely feedforward view of neural processing. Our results suggest that the image structure that triggers our perception of surface gloss plays a direct role in visual categorisation, and that the perception and neural processing of stimulus properties should be studied in the context of recognition, not in isolation.

Editor's summary: Schmid et al. examine how humans perceive material classes from image structure produced by surface reflectance properties.

INTRODUCTION

Our visual experience of the world arises from light that has been reflected, transmitted, or scattered by surfaces. From this light we can tell whether a surface is light or dark, shiny or dull, translucent or opaque, made from platinum, plastic, or pearl (Figure 1a). The quick and accurate recognition of materials and their intrinsic properties is central to our daily interactions with objects and surfaces, from inferring tactile information (is it smooth, heavy, soft?) to assessing function (is it edible, fragile, valuable?). Yet material perception is not trivial as the structure, spectral content, and amount of light reaching our eyes depends not only on surface reflectance, transmittance, and scattering properties, but also on complex interactions with the 3D shape, position, and orientation of surfaces with respect to each other, light sources, and the observer. Thus, our ability to effortlessly discern a wide variety of materials that each have potentially unlimited optical appearance is remarkable, and serves as a compelling example of an important but unresolved challenge in visual neuroscience: How does the brain disentangle the conflated contributing factors to the retinal image to perceive our visual world?

While there is a growing body of work investigating the visual perception of material properties like colour, lightness, transparency, translucency, and gloss[1–6], there is comparatively little work investigating the recognition of different material classes like plastic, pearl, satin, steel, etc.[7–17]. For example, previous research has discovered a limited set of image conditions (photogeometric constraints) that trigger the perception of a glossy versus matte surface, involving the intensity, shape, position, and orientation of specular highlights (bright reflections[18–24]), and lowlights (dark reflections[25]) with respect to diffuse shading. However, it remains unknown what image information triggers our perception of different materials. A likely reason for the disparate focus on material properties versus classes is that studying properties like colour and gloss seems more tractable than discovering the necessary and sufficient conditions for recognising the many material classes in our environment. The challenge is that the perceptual space of materials is unspecified; there are many different optical "appearances" that can look like steel (think polished, scratched, rusted), or plastic (smooth and glossy or rough and dull).

Furthermore, a traditional feedforward view of neural processing is often assumed in which the recognition of objects and materials proceeds from the processing of *low-level* sensory information (image features) to the estimation of shape and surface properties (often referred to as *mid-level* vision[2,3]) to the *high-level* recognition of object and material categories[26] (*feedforward hypothesis*; Figure 1b (i)). Within this framework it would make sense to first study how the visual system computes mid-level material properties like surface gloss and colour from images, as material class is thought to be subsequently computed from the high-dimensional feature space defined by these component mid-level properties[9–12,27,28].

An alternative view suggests that the visual system does not try recover physical surface properties like diffuse reflectance (seen as surface colour) and specular reflectance (seen as gloss) per se., but rather learns about statistical variations or regularities in image structure from which both material properties like gloss and material categories like plastic can be "read out" simultaneously[5,6,29,30] (*simultaneous hypothesis*, Figure 1b (i)). For example, in Figure 1c, systematic differences in the surface reflectance properties of otherwise identical objects produce variations in the appearance of specular highlights, such as their contrast, sharpness, and coverage (see Figure 2).

These variations in highlight appearance not only affect the quantitative level of surface gloss perceived (shiny to dull), but also the *quality* – or material category – of the surface altogether (rubber, plastic, ceramic, or metal), and it is possible that the same visual features, or *cues*, directly underlie both processes. As it stands, however, the relationship between image structure, perceptual properties like surface gloss, and material recognition is unclear.

Here, we test the precise role of image structure produced by specular reflections in material recognition by rendering complex but carefully controlled stimuli in natural illumination fields, measuring and manipulating aspects of specular image structure, and collecting human psychophysical judgments about surface appearance in a series of experiments. Our results show that the visual system is highly sensitive to specular structure for material recognition, even in the absence of information from other physical sources like surface texture, transmittance, and subsurface scattering, suggesting that specular reflections play a more extensive role in visual recognition than previously appreciated. Furthermore, the data reveal that, rather than materials being derived via the estimation of material properties like gloss (*feedforward hypothesis*), it is more likely that visual cues for gloss perception are constrained by material class, implying that the perception and neural processing of gloss and other stimulus properties should be studied in the context of recognition rather than in isolation. Moreover, our results demonstrate that material category is directly computable from measurable visual features of specular structure, and that manipulating these features transforms perceived category, suggesting that specular structure provides direct diagnostic information about an object's material. We discuss a simultaneous account of material perception (*simultaneous hypothesis*) in which stimulus properties like gloss are co-computed with (i.e., constrained by) our qualitative holistic impressions.

[Figure 1 approximately here]

[Figure 2 approximately here]

RESULTS

Specular appearance yields a wide range of material classes

If the visual system is sensitive to specular reflections for material recognition beyond whether a surface is shiny or matte[24], then altering the appearance of specular reflections should lead to changes in perceived material. To test this, we computer-rendered glossy objects with different surface reflectance properties under natural illumination (Figure 1d). We parametrically manipulated base colour (lightness and saturation of the diffuse component) in addition to five specular reflection parameters (specular strength, specular tint, specular roughness, anisotropy, and anisotropic rotation) to control the appearance of specular reflections with respect to diffuse shading, resulting in 270 stimuli (Supplementary Figure 1). We collected unbiased participant-generated category terms for each stimulus using a free-naming task (Experiment 1) where participants (n=15) judged what material each object was made from with no restrictions.

After processing for duplicates and similar terminology (see Analyses), over two hundred terms were used to describe the materials, more than half of which were category terms (i.e., nouns like porcelain, gold, plastic, stone, ceramic, chocolate, pearl, soap, wax, metal, bronze, rubber, fabric, velvet, etc.; Supplementary Figures 2 and 3). Although the use of each term was distributed well among participants (i.e., category labels did not come from the same few participants), semantic labels are only relevant to the extent that they capture qualitative perceptual differences in visual material appearance. For example, dark brown stimuli with medium-clarity specular reflections might be labelled as "melted chocolate" or "mud" but would be qualitatively visually equivalent to one another. Such stimuli would have a different visual quality to those with low-clarity, dim reflections, which might be labelled as "latex" or "rubber" (Figure 3a). Therefore, we sought to reveal the latent perceptual space of materials for our stimulus set, with the following steps.

First, we reduced the set of category labels generated from the free-naming task to those that were used by at least five participants, and merged visually or semantically similar terms, guided by correlations between the categories (see Analyses and Supplementary Figure 4). The reduced set of 18 category terms is shown in Figure 3b. We decided to not reduce the number of category terms further to offer a wide range of choices to subjects in the next experiment.

Second, a separate set of participants (n=80) completed a multiple-alternative-forced-choice task (18-AFC task; Experiment 2) where they were asked to choose the material category that best applied to each stimulus (Supplementary Figure 5). For this experiment, the stimulus set was extended to include a larger range of reflectance parameters, two shapes (dragon and bunny), and two lighting environments (kitchen and campus), resulting in 924 stimuli (Supplementary Figure 1). Participants (n=20 per stimulus) provided confidence ratings (converted to a score between 1-3), which allowed them to indicate their satisfaction with the category options presented. The confidence ratings for each category for each stimulus were summed across all participants, providing a distribution of category responses for each stimulus (*category profiles*) and a distribution of stimuli for each category (*stimulus profiles*; Figure 3b). For many stimuli, more than one category term applied; for example, Stimulus 1 was almost equally classified as "glazed ceramic" and "covered in wet paint", Stimulus 6 was classified as both latex/rubber and plastic, etc. (Figure 3b). This might be due to redundancies in terminology and/or imperfect category membership driving different decision boundaries (e.g., "it looks a bit like plastic but also a bit like rubber"). Indeed, we found that the stimulus profiles for some of the categories correlated with one another (Supplementary Figure 6).

Third, the data were subjected to a factor analysis to extract the common variance between categories and reveal orthogonal perceptual dimensions for our stimulus set. Figure 4a shows that there was no clear plateau in shared variance explained with each additional factor, and twelve factors were needed to account for at least 80% of the shared variance between stimuli (the upper limit based on degrees of freedom; see Analyses). Figure 4b shows example stimuli from the emergent dimensions, which were highly interpretable (twelve plus one dimension that emerged from the negative loadings; Figure 4c). Retaining eight or ten factors accounted for approximately only 60% and 70% of the common variance between categories, respectively, demonstrating that changing the appearance of specular reflections yields a diverse range of perceived materials that cannot be reduced to a small number of perceptual dimensions.

Surprisingly, the materials that were perceived extended beyond those expected based on the reflectance function used to generate them. In the real world, materials like porcelain, pearl, soap, wax, velvet, and many others produce image structure that is caused by the extent and way in which they transmit, internally scatter, and disperse light in addition to pigment variations and mesoscale details like fibres in fabrics, or scratches in anisotropic metals. Yet, participants reported seeing these materials for surfaces that were only defined by (uniform) diffuse and specular reflectance properties, despite the absence of these other potentially diagnostic sources of image structure. Thus, our results reveal that the human visual system is highly sensitive to the image structure produced by specular reflections for material recognition, even for complex materials. Note that some categories (like gold metals) occurred due to our arbitrary choice of yellow for the coloured stimuli, and the outcome would be different for other colours. Nevertheless, as we show later, such materials are additionally defined by specular image structure, and colour information alone is not sufficient to discriminate between materials.

[Figure 3 approximately here]

[Figure 4 approximately here]

Material class is not determined by but may constrain gloss

While manipulating surface reflectance properties changed image structure in a way that participants interpreted as different materials, how this image structure relates to material recognition and the underlying mechanism is unclear. A feedforward approach assumes that the appearance of specular reflections determines surface glossiness, which in turn combines with other estimated mid-level properties (such as object shape, colour, translucency, etc.) to define the material category (Figure 1b (i)). If this is true, then we should be able to identify visual features (*cues*) that predict gloss perception, and the material categories from Experiment 2 should be associated with a particular level of gloss. We tested this in Experiment 3 in which a separate group of participants (n=22) rated the perceived glossiness (i.e., gloss level) of each of the 924 stimuli from Experiment 2. We directly measured visual features of the stimuli that describe the appearance of specular reflections and are based on generative constraints on how light is reflected and scattered by a surface's reflectance properties (Figure 2). Three of these features – *coverage*, *sharpness*, and *contrast* – have previously been found to predict participants' judgments of surface glossiness[32,36], so we refer to them as *gloss cues*. However, whereas Marlow and colleagues used perceptual judgements of each cue to predict gloss, here we operationalised the cues using objective, image-based measures computed from dedicated render outputs (see Analyses and Supplementary Figure 9). Intuitively, *coverage* is the extent to which an object's surface is covered in specular highlights (bright reflections; Figure 2a); *sharpness* refers to the rapidness of change between brighter and dimmer regions within and at the boundary of those highlights, and is usually related to the distinctness of the reflections (i.e., the clarity of the reflected environment; Figure 2b); and *contrast* is the variance in bright and dim regions caused by specular reflections, and usually relates to how bright the specular highlights look in relation to the surrounding regions (Figure 2c).

We found that inter-subject agreement for gloss ratings was high (median r=0.70), and overall a linear combination of the gloss cues (coverage, sharpness, and contrast) accounted for 76% of the variance in participant's gloss ratings (Figure 5a), R^2=0.76, F(3,916)=973.74, p<0.001. This links perceived gloss with objective image-based measures of specular reflection features for a wide range of reflectance conditions. However, the material dimensions defined in Experiment 2 were not associated with a particular level of gloss, contrary to what would be predicted by the *feedforward hypothesis* (Figure 1b (i)). Instead, stimuli from the same material class exhibited a wide distribution of gloss levels, and stimuli from very visually distinct classes like ceramic and (gold and uncoloured) metals had completely overlapping gloss distributions (histograms in Figure 5b). We investigated whether the wide gloss distributions could be caused by the continuous nature of the material dimensions (e.g., perhaps more metallic-looking materials are glossier, more rubber-looking materials are less glossy, etc.). If this is true, then gloss ratings should correlate with the loading of stimuli onto material dimensions (i.e., material factor scores). This was the case for only seven of the thirteen material dimensions (scatter plots in Figure 5b, correlation coefficients highlighted in red), with the other six showing no significant correlation between gloss ratings and material score (black coefficients), suggesting that gloss level is overall not a good indicator of how well stimuli load onto a material dimension.

The lack of a clear relationship between perceived gloss and material raises the possibility of a direct link between visual features and perceived material. Indeed, we found that material (factor) scores correlated directly with one or more of the cues for gloss (i.e., the visual features coverage, sharpness, and contrast) for eleven of the thirteen material dimensions (Figure 5c, left). This relationship between cues and material scores can explain the extent to which gloss ratings predicted material scores in Figure 5b (shown in Figure 5d, r=0.50, p=0.001), suggesting that an object's material could be computed from features of specular image structure directly rather than via gloss estimation. Supplementary Analysis A supports this idea by showing that a model predicting material score from gloss level performed worse than a model predicting material score directly from a linear combination of the cues (Supplementary Figure 12). Moreover, the predictiveness of cues to perceived gloss varied across different material classes (black bars in Figure 5c), suggesting that, rather than gloss mediating cues to material, the visual cues used to estimate gloss could instead be constrained by perceived category. In the General Discussion we explore the possibility that surface gloss and material could be co-computed and mutually constrained by the same image structure (see also Supplementary Analysis A). Collectively, these results do not support the idea that gloss mediates the perception of materials from specular reflection image features (*feedforward hypothesis*; Figure 1b (i)) but are in line with the idea that surface gloss and material class could be simultaneously "read out" from these features (*simultaneous hypothesis*; Figure 1b (i)).

[Figure 5 approximately here]

Features of specular image structure predict material class

If visual features directly determine the perceived material of an object, this likely includes colour information. Indeed, many material dimensions that emerged from our stimulus set seem to

be defined by specific colour information produced by reflectance properties, including gold, brown, and uncoloured metals; melted and solid chocolate; glazed and unglazed porcelain; and waxy materials (Supplementary Figure 8). We measured three visual features – *highlight saturation*, *lowlight saturation*, and *lowlight value* – that capture colour information within and outside of the specular highlight regions (Figure 2d-f). These features, or *colour cues*, are linked to different aspects of a material's colour appearance (e.g., surface pigment, or specular "tint" seen in coloured metals). Intuitively, *highlight saturation* is the colour saturation within the specular highlight regions (bright spots); *lowlight saturation* is the colour saturation outside of the highlight regions (referred to as lowlight regions[25]; and *lowlight value* is the brightness of the colour within the lowlight regions.

To test the prediction that materials can be discriminated directly from features of specular image structure (*simultaneous hypothesis*; Figure 1b (ii)), we sought to predict material class from the measured visual features (the three gloss cues and three colour cues; Figure 2) for the stimuli that loaded most strongly onto each material dimension (Figure 4). Figure 6a plots the distribution of visual features for these stimuli (coloured violin plots) for each material. Visualised in this way, the features provide "material signatures" for each class. The data were subjected to a linear discriminant analysis (LDA) that classified materials based on linear combinations of features. We used a leave-one-condition-out approach, where the classifier was trained on three out of the four shape/lighting conditions (e.g., dragon-kitchen, bunny-kitchen, dragon-campus) and tested on the remaining condition (e.g., bunny-campus). Figure 6b plots the accuracy of the model for each material, combined over the four training-test combinations. Overall accuracy was 65%, which is well above chance (7.7%, red dotted line). A further cross-validation test showed that the model generalised across shape and lighting conditions (Figure 6c, mean accuracy=64%, SD=5.2), demonstrating that features of specular structure can predict human material categorisation behaviour for our stimulus set. Figure 6d and e illustrate that the discriminations made by the model are perceptually intuitive.

Interestingly, some materials were classified better than others (Figure 6b). The materials with the highest classification accuracies (Figure 6b) were those with the most distinct material signatures (Figure 6a; waxy materials=100%, uncoloured metals=95%, unglazed porcelain=94%, solid chocolate=89%, gold metals=86%, glazed porcelain=72%, brown metals=69%, melted chocolate=61%). That is, there are at least a few features that seem to "characterise" those materials (e.g., silvers must have uncoloured highlights and lowlights; wax must be light and coloured with low-contrast reflections). The materials with the lowest classification accuracies had a less distinct set of features (ceramics=36%, pearlescent materials=31%, plastic=28%, fabrics=28%). One possible reason for this is that there are, for example, many types of plastics or pearlescent materials (Figure 4b), and different specific (nonlinear) combinations of features define these subtypes – something that LDA does not capture. A second possibility is that the stimuli might not fall nicely into perceptually discrete classes and would be better represented as a smooth continuation from one material dimension to another. To investigate this, we tested whether variations in category profile between stimuli (Figure 3b) could be predicted by variations in the measured visual features (Figure 2). Using representational similarity analysis (RSA)[37] we found that visual features predict material changes better than gloss perception (Supplementary Analysis B and Supplementary Figure 15).

[Figure 6 approximately here]

Manipulating specular structure transforms material class

Thus far our analyses have been correlational in nature. If material class is computed from combinations of the visual features that we measured (Figure 2), then directly manipulating these features should transform perceived category in predictable ways. This would better test whether the measured features are causally responsible for the perceived category shifts, or whether they merely correlate with other changes in image structure that are important for material perception that we did not measure.

To this end, we attempted to directly manipulate visual features to transform stimuli that were perceived as glazed ceramic in Experiment 2 into each of the remaining materials, for each shape and light field (Figure 7a). We created two stimulus sets to test the success of our feature manipulations. The first set contained "simple" (linear) feature manipulations, which closely correspond to the previously measured visual features (see Supplementary Figure 16). The second set contained "complex" (nonlinear) feature manipulations, which accounted for particular types of contrast needed for some materials that we empirically noticed were not captured by the first set of manipulations (Supplementary Figure 16). For example, the velvety/silky stimuli from Experiment 2 were exclusively defined by very rough, anisotropic specular reflections, which caused a high degree of directional blur. This gives the effect of elongated, low-clarity specular reflections that, despite this low clarity, have rapidly changing (sharp/high contrast) boundaries between highlights and lowlights relative to isotropic surfaces (see Figure 3a, Figure 4b, and Supplementary Figure 8). On the other hand, the perception of uncoloured metals appears to require the presence of high contrast reflections that appear all over the surface (i.e., there is high contrast image structure in the lowlights, not just the highlights; see also Norman et al.[13]). These particular types of contrast could not be achieved through a simple linear transformation of image intensities from the glazed ceramic stimuli but were accounted for after applying more complex additional filters to non-linearly manipulate pixel intensities (Figure 7b and Supplementary Figure 17).

In Experiment 4, a new set of participants (n=22) performed an 18-AFC task identical to Experiment 2, but with the feature-transformed stimuli. The stimulus profiles were converted to factor scores to see which material dimensions best applied to the new stimuli. The average factor scores plotted in Figure 7c and d show that participants agreed with our informal observations. For the first stimulus set (linear feature manipulations; Figure 7c), stimuli that were intended to be silver metals and fabric actually had the quality of glazed ceramic and plastic, respectively. For the second stimulus set (non-linear feature manipulations; Figure 7d), the perceived materials align much better with the intended classes.

[Figure 7 approximately here]

DISCUSSION

In the present study we showed that the image structure produced by specular reflections not only affects how glossy a surface looks; it can also determine the quality, or category, of the material that is perceived. Critically, specular reflections provide more information than just distinguishing between surfaces that are matte versus glossy[24], or plastic versus metal[16]; we found that the perceptual dimensionality of glossy surfaces defined by a diffuse and specular component is much larger than has previously been suggested[38–41]. This is likely because prior studies on gloss perception have manipulated stimulus parameters within a narrow range and used restricted tasks; for example, stimuli are simple, smooth shapes and/or their reflectance properties only encompass the range of plastic- and ceramic-looking materials, and participants judge the level of gloss or relative similarity between surfaces. Here, we used complex shapes, manipulated reflectance parameters within a wider range, and asked participants to judge each object's qualitative material appearance. This greatly expanded the number of dimensions required to account for perceptual differences between glossy surfaces relative to prior studies. Note that we did not sample the whole perceptual space of glossy objects defined by a diffuse and specular component (e.g., we manipulated colour saturation within only one hue). The dimensionality of gloss appearance is likely to expand further upon sampling a wider range of reflectance parameters, shapes, mesostructure detail, and environment lighting conditions.

Importantly, we found that changes in specular structure – caused by either generative sources or direct feature manipulation – led to qualitative shifts in material appearance beyond those expected by the reflectance function used. This demonstrates that features of specular image structure can be diagnostic for recognising a wide range of materials. A potential reason for this is that a material's surface reflectance properties create some of its most salient optical characteristics, and, since most surfaces reflect some light specularly, relying on such characteristics could carry ecological importance when other cues to material are not available. For instance, the visual effects of translucency can be greatly diminished with frontal lighting[42]; in such cases, visual features caused by specular reflections might remain diagnostic of translucent materials (e.g., porcelain). Similarly, meso-scale details like fibres on cloth or scratches on metal might not be visually resolvable when seen at a distance; yet such materials might still be recognised from specular image structure. Indeed, these additional sources of image structure (from mesostructure or translucency) are absent from the stimuli used in the present study, and although such details might render the materials more compelling (when say compared to a photograph), the stimuli nevertheless convincingly resemble silk-like, porcelain-like, wax-like, brushed metal-like, etc. materials. Interestingly, these results are in line with findings from the computer graphics literature that show that visual effects from different types of fabrics come predominantly from specular reflections, and diffuse reflection and shadowing-masking play a much less pronounced role even for relatively matte-looking fabrics[43].

Our data do not support the notion of a feedforward path to recognition whereby the visual system combines estimates of physical properties that are first "recovered" from images; instead, our results are in line with the idea that vision assigns perceptual qualities to statistically varying image structure[2,5,6,44,45]. Specifically, we found that gloss was not a good indicator of a material's class; instead, materials were differentiated directly based on measurable image-based specular reflection features, suggesting that material qualities like gloss are a perceptual outcome with –

rather than a component dimension of – our holistic impressions. Indeed, the perception of material qualities like gloss might even be influenced by these holistic impressions, as suggested by the fact that the contribution of different cues to gloss was not stable across different materials, but co-varied with the cues that predicted material class. That is, the regions of feature space occupied by different material classes (i.e., different qualitative appearances) seemed to mediate the processing of those same features when estimating surface glossiness.

One potential mechanism is that emergent categories from our stimulus set could reflect cognitive decisions, and this cognitive interpretation has a "top-down" influence on which features are used to judge surface gloss. For example, "chocolate" and "plastic" could have similar contrast and sharpness of specular highlights (similar "gloss types") and the different labels might result from a cognitive decision by participants based on body colour (brown versus yellow). However, we cannot think of a principled reason why a feature's influence on material category would affect how people choose to use that feature for gloss judgments.

Furthermore, we argue that such a clear perceptual scission of image structure into different layers (e.g., a gloss layer and a body colour layer) and then subsequent "cognitive reassembly" is unlikely, especially for the complex-shaped stimuli and wide sampling of reflectance parameters used in the present study. This kind of layered appearance of gloss, which can be construed as a form of transparency perception, likely only applies to smooth surfaces with little variation in surface curvature, like spheres[25]. For the present stimuli, specular image structure seems to interactively combine with diffuse shading to create a gestalt-like quality for each material, such that we may not even have perceptual access to individual cues (e.g., see Supplementary Figure 1). Like with object recognition, this material quality can be given a label like "gold" or "pearl", but nonetheless reflects a holistic impression, rather than cognitive combination of cues (also see Okazawa et al.[46]). Such holistic impressions are likely why we could successfully manipulate visual features to transform one category into another (Experiment 4), but why linear models like LDA, and RSA (Experiment 2) did not better account for the data.

We propose an alternative mechanism to explain the co-variation between cues for surface gloss and material class, inspired by converging evidence from independent lines of psychophysical and neuropsychological research that suggests computations of stimulus properties are inherently *coupled*[3,47]. Specifically, monkey physiology studies have found that neurons in primate area V4 respond only to specific combinations of texture and shape but not to these properties separately, suggesting joint representations of shape and surface properties[47]. In line with this, a series of human psychophysics studies have demonstrated that percepts of 3D shape and surface properties (e.g., gloss or translucency) are mutually constrained by specific image gradient-contour relationships (i.e., the cues for each are not separate)[3]. Similarly, transparency impressions triggered by low contrast centre-surround displays are mutually constrained with the perception of surface lightness: a medium-grey patch surrounded by a slightly darker (homogenous) surface can appear whiteish with the quality of a flimsy transparent sheath, which occurs to the extent that luminance within the central patch is attributed to a foreground (patch) or background (surround) layer[48]. In the present study, the co-variation between cues for gloss and material class could also reflect a mutual dependency (i.e., their computations could be coupled). We suggest that a surface's qualitative (categorical) appearance constrains how specular image structure is perceptually allocated to different aspects of material appearance such as specular reflections (versus say bright

pigment, or illumination, or 3D shape[49,50]), and thus which cues are *perceptually available* to make a surface look "glossy" or "shiny".

Interactions between categorical, or "holistic", material impressions and the perception of different stimulus properties warrants further investigation because the underlying mechanism will influence how we study perception in multiple domains. Recently, unsupervised deep neural networks (DNNs) have generated excitement as a potential tool to uncover purported mid-level visual processes, for example by providing evidence that the brain could potentially spontaneously discover the existence of properties like gloss, illumination direction, and shape without prior knowledge about the world, but by learning to efficiently represent variations in image structure produced by these properties[51]. However, here we showed that with more complex stimuli image information cannot be neatly mapped onto perceptual properties like gloss; this mapping is constrained by the qualitative, or categorical appearance of surfaces. Some material classes occupy relatively homogeneous feature spaces (e.g., gold) compared to other, more diverse classes (e.g., plastic), which likely emerges through behavioural relevance. Thus, for unsupervised learning to generate useful hypotheses about perceptual processes, future work needs to consider how the behavioural relevance of stimuli might influence spontaneous learning about statistical image variations.

Neuropsychological results should also be interpreted in the context of qualitative, categorical material perception. For example, neurons in the lower bank of the superior temporal sulcus (STS) in monkeys have been shown to preferentially respond to specific combinations of manipulated gloss parameters[52]. It is possible that different categorical or holistic material impressions are triggered by these specific combinations, driving this response. Moreover, research into object recognition[53] often investigates how visual features like texture[54] and form[55] influence object category representations, but thus far have neglected the role of surface properties in recognition, for example by not controlling for changes in surface properties[55–57], or "scrambling" the images in a way that breaks the perception of surfaces[54]. Our results suggest that the specific photogeometric constraints on image structure that trigger our perception of surface gloss play an important role in visual categorisation and recognition, and that a fruitful direction for neuropsychological research could be to focus on identifying the neural mechanisms that represent objects holistically[58].

However, we think that this role extends beyond object identification. Just as surface gloss is not a component dimension of material perception, material perception is not merely a component dimension of scene perception. A material's qualitative appearance is useful for different aspects of navigation, such as identifying where we are (beaches have sand and water), choosing which paths to take (icy vs. muddy vs. concrete), and deciding which obstacles to avoid (solid rock vs. flexible vegetation). We also need material perception for locating items among clutter (finding a metallic pot in the cupboard); evaluating an object's function, usefulness, or motor affordances (Can I eat it? Is it expensive or fake? How do I pick it up?), and predicting tactile properties and future states of objects (Is it heavy, sticky, or wet? Will shatter, stretch, or bounce?). Our results shed light on how image structure is transformed into our representations of surfaces with particular material "appearances", thereby making an important contribution towards bridging the gap between the early stages of visual information processing and behavioural goals such as categorisation, action, and prediction[59,60].

METHODS

Ethics

This study was approved by the local ethics review board of the Justus Liebig University Giessen (LEK FB 06) and strictly adhered to the ethical guidelines put forward by the declaration of Helsinki (2013). All participants gave written informed consent prior to the experiments and were told about the purpose of the experiments. All participants were compensated for their participation at a rate of 8 €/hour.

Participants

Volunteer participants were students at Justus Liebig University Giessen in Germany. Fifteen participants completed the free-naming experiment (Experiment 1, mean age: 23.7, female 80%, male 20%). Eighty native-level German speakers participated in the 18-AFC experiment (Experiment 2, mean age 24.9, female: 83.3%, male: 16.7%), 22 participants took part in the gloss rating experiment (Experiment 3, mean age: 25.3, female: 58.3%, male: 41.7%), and 22 participants took part in the feature manipulation experiment (Experiment 4, mean age: 23.4 , female: 81.8%, male: 18.2%). Different participants were recruited for each experiment.

Stimuli

Stimulus generation: Experiment 1 (free-naming), Experiment 2 (18-AFC), and Experiment 3 (gloss ratings). We generated our stimulus set by computer rendering images of complex glossy objects under natural illumination fields. Each image depicted an object with the illumination field as the background, rendered a resolution of 1080x1080 pixels. Object meshes were the Stanford Bunny and Dragon from the Stanford 3D Scanning Repository (Stanford University Computer Graphics Laboratory; http://graphics.stanford.edu/data/3Dscanrep/). Wavefront .obj files of these models were imported into the open-source modelling software Blender (v2.79) and rendered using the Cycles render engine, which is an unbiased, physically based path-tracing engine.

Stimuli were illuminated by the “kitchen” and “campus” light fields from the Debevec Light Probe Image Gallery[61]. Interactions between light and surfaces were modelled using the Principled BSDF shader, which is based on Disney’s principled model known as the “PBR” shader[62]. The Principled BSDF shader approximates physical interactions between illuminants and surfaces with a diffuse and specular component (for dielectric materials), and a microroughness parameter that controls the amount of specular scatter. An advantage of the Principled BSDF shader over other models like the Ward model is that it accounts for the Fresnel effect. The microfacet distribution used is Multiple-scattering GGX, which takes multiple bounce (scattering) events between microfacets into account, giving energy conserving results. Although there are many parameters to adjust in the Principled BSDF shader, we manipulated only the following (details can be found at https://docs.blender.org/manual/en/dev/render/shader_nodes/shader/principled.html):

- *Base Color:* Proportion of light reflected by the R, G, and B diffuse components.
- *Specular:* Amount of light reflected by the specular component. The normalised range of this parameter is remapped linearly to the incident specular range 0-8%, which encompasses

most common dielectric materials. Values above 1 (i.e., above 8% specular reflectance) are also possible. A value of 5 (the maximum value used) translates to 40% specular reflectance.

- *Specular tint:* Tints the facing specular reflection using the base colour, while glancing reflection remains white.
- *Roughness:* Specifies microfacet roughness of the surface, controlling the amount of specular scatter.
- *Anisotropic:* Amount of anisotropy for specular reflections. Higher values give elongated highlights along a surface tangent direction given by a tangent field. This field is obtained per object using the spherical mapping functionality of Blender. Internally, it works in two steps. First, a radial pattern of directions is defined on a 3D sphere enclosing the object, by assigning the direction of the meridian going through each spherical point. Such a pattern has singularities at sphere poles, where all directions converge. Second, each point **p** on the object is mapped to a point **q** on the enclosing sphere by tracing a ray from the object center through **p**. The direction at **q** is then projected on the local tangent plane at **p** and renormalized to yield the local tangent. Note that singularities remain singularities after projection.
- *Anisotropic rotation:* Rotates the direction of anisotropy in the local tangent plane, with a value of 1.0 indicating a rotation of 360°.

Rendering parameters: Experiment 1 (free-naming). A total of 270 stimuli were generated for the free-naming experiment. Stimuli were generated by rendering combinations of the parameters for a dragon model embedded in the kitchen light field:

- *Base Color:* Six diffuse base colours based on two levels of saturation (greyscale, saturated yellow) and three levels of value (dark, medium, and light). For the greyscale stimuli, RGB values were equal for each lightness level (RGB=0.01, 0.1, and 0.3 for dark, medium, and light stimuli, respectively). The RGB values for saturated stimuli were [0.01, 0.007, 0.001] for dark stimuli, [0.1, 0.074, 0.01] for medium stimuli, and [0.3, 0.221, 0.03] for light stimuli.
- *Specular:* Three specular levels: 0.1, 0.3, and 1, which correspond to 0.8%, 2.4%, and 8% specular reflectance, respectively.
- *Specular tint:* Two specular tint levels for saturated stimuli: 0 (no specular tint) and 1 (full specular tint).
- *Roughness:* Four roughness levels: 0, 0.1, 0.3, and 0.6.
- *Anisotropic:* Two anisotropic levels for stimuli with roughness levels greater than zero: 0 and 0.9.
- *Anisotropic rotation:* Two anisotropic rotations for anisotropic stimuli: 0 (no rotation) and 0.25 (90° rotation).

Rendering parameters: Experiment 2 (18-AFC) and Experiment 3 (gloss ratings). A total of 924 stimuli were generated for the 18-AFC and gloss rating experiments. Stimuli were generated by rendering combinations of the parameters for two shapes (dragon, bunny) and two light fields (kitchen, campus):

- *Base Color:* Six diffuse base colours based on two levels of saturation (greyscale, saturated yellow) and three levels of value (dark, medium, and light). For the greyscale stimuli, RGB values were equal for each lightness level (RGB=0.01, 0.03, and 0.3 for dark, medium, and light stimuli, respectively). The RGB values for saturated stimuli were [0.01, 0.007, 0.001] for dark stimuli, [0.03, 0.022, 0.003] for medium stimuli, and [0.3, 0.221, 0.03] for light stimuli.
- *Specular:* Four specular levels for the darkest four diffuse shading levels: 0.1, 0.3, 1, and 5, which corresponds to 0.8%, 2.4%, 8%, and 40% specular reflectance, respectively. The light-coloured stimuli were only rendered with the first three specular levels.
- *Specular tint:* Two specular tint levels for saturated stimuli: 0 (no specular tint) and 1 (full specular tint).
- *Roughness:* Three roughness levels: 0, 0.3, and 0.6.
- *Anisotropic:* Two anisotropic levels for stimuli with roughness levels greater than zero: 0 and 0.9.
- *Anisotropic rotation:* Two anisotropic rotations for anisotropic stimuli: 0 and 0.25.

Stimulus generation: Experiment 4 (feature manipulations). We generated the feature-manipulated stimulus set with an approach similar to compositing techniques used in visual effects. For our manipulation we need several image components of an input scene: a full rendered image, an image of the object rendered with diffuse shading only (diffuse component), a binary mask image, and an image of the surface normals. All of these components were rendered with the Cycles engine in Blender: A manipulated image is obtained via a sequence of steps that are depicted in Figure 7a: a specular component is first obtained by subtracting the diffuse component from the full rendered image; the sharpness of this specular component is reduced via a normal-based blur operator; the diffuse and specular components are then optionally multiplied by a colour; finally, the intensity and saturation of the resulting specular and diffuse components are adjusted and the final image is obtained by addition of these components (specular+diffuse). These steps are explained in more detail below.

The normal-based blur operator is implemented with a cross bilateral filter (Paris et al., 2009) on the specular component and takes into account the orientation of the surface normals. The cross bilateral filter consists in an image convolution with a modified blur kernel: it not only weights the contribution of neighbour pixel colours given their distance (the spatial weight), but also based on additional pixel-wise information (the range weight). The two weights are combined multiplicatively, and the blur kernel is systematically normalized so that the sum of combined weights is one. In our case, the space weight is given by a box kernel of a large static size (160x160 pixels), which merely serves to limit the set of pixels that are considered for blurring. The range weight is given by an exponentiated normal dot product: $(\mathbf{n} \cdot \mathbf{n_0})^s$, where $\mathbf{n_0}$ is the normal at the centre pixel in the kernel, $\mathbf{n}$ is the normal at a neighbour pixel, and s is a shininess parameter. This simple range weighting formula is directly inspired by Phong shading: large shininess values result in a sharp angular filter, whereas small shininess values result in a broad angular filter, which has the effect of blurring specular reflections. In practice, we use a blur parameter b in the [0..1] range, which is empirically remapped to shininess using $s=1+1/((b/5)^4*100+\varepsilon)$, where ε is a small constant used to avoid division by 0.

Some particular categories of materials such as gold or chocolate require a colorized diffuse or specular component. This is obtained by simply multiplying the image component by a colour: *sCol* for specular, *dCol* for diffuse. In some cases (gold and silver), the diffuse component is entirely discarded, which is equivalent to having *dCol*=(0,0,0) as indicated in Figure 7a. We have made the choice of using bright and saturated colours in all other cases in order to make the parameters of the next step more easily comparable.

For the last intensity and colour adjustment step, each image component is first converted to the HSV colour space. The following manipulations apply either to the specular or diffuse component, with the corresponding parameters either prefixed by 's' or 'd'. The value channel is manipulated by a combination of gamma exponentiation (controlled by *dGamma* and *sGamma*) and multiplication (by *sBoost* or *dBoost*) using the following simple formula: $Boost*v^{Gamma}$, where *v* stands for value in HSV colour space. The saturation channel is also multiplied (by *sSat* or *dSat*). Both components are converted back to RGB and added together to yield the final image.

We use an additional filter for the specific case of velvet/satin. It is applied to the specular component right after normal-based blurring. The specular component is multiplied by a first mask m_1 that is computed via a non-linear mapping of the luminance *l* of the specular component. We use the following formula: $m_1=f(0, l_m/2, l)$ if $l<l_m/2$, and $m_1=1-f(l_m/2, l_m, l)$ otherwise, with l_m the maximum luminance in the specular component image, and $f(a,b,x)$ the "smooth step" function defined in the OpenGL Shading Language that maps $a\leq x\leq b$ to the $[0,1]$ range (See https://www.khronos.org/registry/OpenGL-Refpages/gl4/html/smoothstep.xhtml for details). The effect of multiplication by m_1 is to darken the core of the brightest highlights, hence producing elongated highlight loops as seen in Supplementary Figures 16 and 17. It is similar in spirit to the sinusoidal modulation of Sawayama and Nishida[63] (Figure 11 in their paper), except that it is only applied to the specular component and with a different non-linear remapping. We have also found it necessary to multiply the specular component by a second mask m_2 to slightly attenuate some of the elongated highlights. This mask is computed using a simple normal dot product: $m_2=(\mathbf{n}\cdot\mathbf{n_m})$, where **n** is the normal at a pixel and $\mathbf{n_m}$ is a manually-set direction that is chosen per scene, pointing to the direction of the main highlight. This has the effect of mimicking the unconventional darkening observed in images of velvet materials at locations where the main highlight should occur.

Manipulation parameters: Experiment 4 (feature manipulations). For each of the four scenes showing either of two objects (bunny and dragon) in either of two lighting environments (campus and kitchen), we have taken as input a material configuration previously classified as 'Glazed Ceramic', from which we have produced 12 manipulations for each of the other material categories. There were also two manipulation conditions (see below), yielding a total of 104 stimuli for the feature manipulation experiment.

The input material was systematically chosen to have a grayscale *Base Color* of 0.1, a *Specular* level of 0.3, a *Specular Tint* of 0, a *Roughness* of 0, and an *Anisotropic* level of 0.

The manipulation parameters are listed in Supplementary Tables 2 and 3. In the "simple" manipulation condition (Supplementary Table 2), all the *Gamma* (i.e., non-linear) parameters were set to 1 and the special velvet filter was discarded. In the "complex" manipulation condition (Supplementary Table 3), setting some of the *sGamma* and *dGamma* parameters away from 1 had

an impact on the intensity of either the specular or diffuse component; as a result, we also had to adjust the *sBoost* and *dBoost* parameters. We applied the velvet filter only for the velvet manipulation in this condition, using the following $\mathbf{n_m}$ directions: (0.57,-0.53,0.63) for Bunny/Campus, (-0.22,-0.48,0.85) for Bunny/Kitchen, (-0.62,0.7,0.35) for Dragon/Campus, and (-0.72,0.57,0.39) for Dragon/Kitchen.

Procedure

Stimulus presentation. Stimulus presentation and data collection were controlled by a MATLAB script (release 2018b, Mathworks, Natick, MA) using the Psychophysics Toolbox version 3[64]. In the free naming task, stimuli were projected onto a white wall in a classroom. Thick black cloth was used to block light from the windows, so that the only source of light came from the projected image. For all other experiments the stimuli were presented on a Sony OLED monitor running at a refresh rate of 120 Hz with a resolution of 1920x1080 pixels controlled by a Dell computer running Windows 10. Stimuli were viewed in a dark room at a viewing distance of approximately 60cm. The only source of light was the monitor that displayed the stimuli.

Task: Experiment 1 (free-naming). Fifteen participants gathered in a classroom and completed the task at the same time. Two authors A.S.C. and K.D. were present in the room with the participants. This was a qualitative data collection session, and no specific hypothesis was considered at this stage. They viewed stimuli one at a time and were asked to classify the material of each stimulus, with no restrictions. They were provided with sheets of paper with space for each trial number to write down their answers. The experimenter controlled stimulus presentation with a keyboard press. Each trial was presented for as long as it took for all participants to finish writing down their responses. A blank screen was shown between each trial for one second. The experiment took approximately three hours to complete, including breaks.

The instructions, written on a sheet of paper in both English and German, were as follows:

> You will be shown 270 images, and your task is to write down your impressions of the material of each object, i.e. what does it look like it is made of? Below are some suggestions of materials that might help prompt you. Your answers might not include, nor are they limited to, the suggestions below. There are no right or wrong answers and you can respond however you like. You can be as specific (e.g. aluminium foil, polyethylene), or general (e.g. metal, plastic) as you like. You can also write down more than one material (e.g. “looks most like glazed ceramic but could also be plastic”), or even say that it doesn’t look like anything you know. If you can’t remember the name of a material, you can write e.g. “the stuff that X is made out of”.

The following examples were provided below the instructions:

- Metal, e.g. silver, gold, steel, iron, aluminium, chrome, foil

- Textiles, e.g. velvet, silk, leather
- Plastic, e.g. PVC, nylon, acrylic, polyethylene, Styrofoam
- Ceramic, e.g. porcelain, china
- Minerals, e.g. stone, concrete, rock
- Coatings, e.g. glazed, painted, enamel, PTFE/Teflon
- Other: soap, wax, chalk, pearl, composite materials

Task: Experiment 2 (18-AFC). The 924 stimuli were randomly split into 4 sessions, so that each participant categorised a quarter of the stimuli (20 participants per stimulus). The experiment was self-paced with no time constraints, and for most participants the experiment lasted approximately 1-1.5 hours. Only experimenter and participant were present during the experiment.

On each trial, observers were presented with an object in the centre of the screen (29° visual angle) with 18 categories displayed along the sides of the stimulus (Figure 4a). They were asked to choose the category that best applied to that stimulus. If they were unhappy with their choice, they could change the confidence rating at the bottom right of the screen. Before the experiment, participants were asked to read carefully through the list of materials and were shown several examples of the stimuli to be presented. The purpose of this was so observers got a sense of the range of materials in the experiment. Observers were restricted to choosing only one category for each stimulus, and were given the following instructions verbally by the experimenter:

> Use the mouse to click on the category that best describes the material of each object. When you are satisfied with your choice press the space bar to proceed to the next trial. In the case of categories with multiple items (e.g. velvet/silk/fabric), the perceived material only needs to apply to one, not all, the categories. There are no right or wrong answers, as the experiment is about the perception of materials. Not all categories will have an equal number of stimuli – you may choose one category more or less than others (or not at all). If you are not satisfied or confident with your choice, change the confidence rating at the bottom of the screen. This should be used in the case you feel that none of the available categories fit; if you think more than one category applies then just choose the most suitable option.

Task: Experiment 3 (gloss ratings). Twenty-two native-level German and English speakers participated in the gloss ratings experiment. One participant was excluded because they did not understand the task instructions, resulting in 21 participants. The experiment was self-paced with no time constraints, and for most participants the experiment lasted approximately 1 hour. Only experimenter and participant were present during the experiment.

Before the experiment, participants were shown real world examples of glossy objects (some of which are shown in Figure 1a). As they were shown these images, they were given the following instructions:

> Many objects and materials in our environment are glossy or shiny. Glossy things have specular reflections, and different materials look glossier/shinier

than others. You will be shown different objects and asked to rate how glossy each object looks. This experiment is about visual perception so there is no right or wrong answer – just rate how glossy each object looks to you.

Participants were shown many example trials before starting the experiment so that they could experience the full range of stimuli and calibrate their ratings to the range of gloss levels in the experiment. On each trial, observers were presented with an object in the centre of the screen (29° visual angle). The task was to rate how glossy the object was by moving the mouse vertically to adjust the level of a bar on the right side of the screen. They were told to look at many points on the object before making their decision. The starting level of the rating bar was randomly set on each trial.

Task: Experiment 4 (feature manipulations). Twenty-two native-level German speakers participated in the feature manipulation experiment. Each participant categorised all 104 stimuli, and the task was identical to the 18-AFC task in Experiment 2. The experiment was self-paced with no time constraints, and for most participants the experiment lasted approximately 45 minutes. Only experimenter and participant were present during the experiment.

Analyses

Reduced set of category terms: Experiment 1 (free-naming). A student assistant went through participants' responses of the free-naming task and extracted all category terms (nouns) and descriptors (e.g., adjectives describing fillings, coatings, finishes, and states), translating them to English. Each participant often used multiple terms per stimulus, all of which were recorded as separate entries. Similar terms (like foil/metal foil/aluminium foil, or pearl/pearlescent/mother of pearl/bath pearl) were combined into a single category. Some terms like "Kunststoff" and "Plastik" were considered duplicates because they translated to a single term (plastic) in English.

Reduced set of category terms: Experiment 2 (18-AFC). We decided to retain the category terms from the free-naming task that at least 5 out of 15 participants used. This was an arbitrary cut off with the aim of being quite inclusive while at the same time maintaining some consensus among participants. Supplementary Figure 2 shows the 29 category terms that survived this cutoff. Subsequently, two of the authors (A.C.S. and K.D.) worked together to reduce this set of category terms. For example, materials that were visually or semantically similar were combined (specifically: fabrics, including velvet and silk; unglazed porcelain, stone, and chalk; latex and rubber; wax and soap; brass, bronze, and copper; iron and chrome; and silver and aluminium). To have some numerical guidance for this reduction we computed the correlation between each pair of categories, calculated from the number of participants that used each term for each stimulus (Supplementary Figure 4). That is, for each pair of categories, correlated vectors were indexed by stimulus number with the values being number of observer responses. The superordinate category "metal" was not included separately. In the free-naming task, non-category terms were often required to distinguish particular coatings (glazed/varnished), finishes (polished), or states (liquid/melting; see Supplementary Figure 2). For the 18-AFC task, we separated categories where these terms applied; for example, we included both

glazed porcelain and unglazed porcelain, both liquid and solid chocolate, and also included a "covered in wet paint" category. Thus, the final set of 18 category terms (Supplementary Figure 5) that participants could choose from in the 18-AFC task was as inclusive as possible while not being too large to overwhelm participants.

Factor analysis: Experiment 2 (18-AFC). The correlations between different categories in terms of their stimulus profiles (Supplementary Figure 6) indicate that some category terms were not independent. This suggests the existence of underlying common dimensions; that is, participants used the same underlying criteria for different category terms. An exploratory factor analysis was performed on the stimulus profiles from Figure 3b, which allowed us to explain some of this covariation and reveal the underlying category space of our stimulus set. Figure 4a (red squares) shows that there is a steady increase in the common variance explained by each additional factor. The amount of additional variance explained by each factor did not drop off after a certain number of factors, indicated by the absence of a plateaux in this plot. Therefore, we extracted 12 factors (the upper limit based on degrees of freedom), which where interpretable and explained over 80% of the common variance between stimulus profiles. This is similar to or greater than the amount of variance explained by the factors/components retained in other material perception studies that have used factor analysis or principal components analysis[65]. The factors were interpretable (Figure 4B) and were labelled based on the original categories that loaded most strongly onto each factor (Figure 4c). Factor scores were calculated using a weighted least-squares estimate (also known as the "Bartlett" method). For comparison, Supplementary Figure 7 shows the results of a principal components analysis that retained all dimensions in addition to the results of other factor solutions, whose dimensions overlap with those here.

Visual features: Measurements for Experiment 2 (18-AFC) and Experiment 3 (gloss ratings). Calculations for most of the visual features relied on specular highlight coverage maps. For glossy surfaces with uniform reflectance properties (like the stimuli used in the present study), specular reflections cover the entire surface. However, for low-gloss objects we often only see specular highlights, or "bright" reflections, which are usually reflections of direct light sources like the sun, or a lamp. For very shiny surfaces (like metal) and in some lighting conditions we also see lowlights, or "dark" reflections[25], which are reflections of indirect light coming from other surfaces in the scene. We chose to define *coverage* as the amount of the surface covered in specular highlights, excluding lowlights, which is consistent with previous literature[36,66]. Marlow and colleagues measured *coverage*, *contrast*, and *sharpness* of specular highlights using perceptual judgments from participants, due to the difficulty in segmenting the specular component of an image from other sources of image structure (such as diffuse reflectance and shading). An acknowledged concern of this approach is that it uses one perceptual output (perceived coverage, contrast, and sharpness) to match another (perceived gloss). It is unclear what image structure observers use to judge each feature, and participants might conflate their judgments of the visual features with each other and with perceived gloss[32] (see also van Assen et al.[67] for a similar use of this method). Therefore, we wanted to develop objective measures of specular reflection features. Currently there is no established method for segmenting specular reflections and diffuse shading from a single image that is robust across different contexts (e.g., changes in surface albedo, shape, and illumination

conditions). To help with this segmentation we rendered additional images that isolated specular and diffuse components.

Visual features: Specular reflection segmentation. For each stimulus, a purely specular image was rendered, which had the same specular reflectance as the original (full rendered) image but with diffuse component turned off. Two purely diffuse images were rendered, which we call "diffuse image 1" (used to calculate *coverage*) and "diffuse image 2" (used to calculate *sharpness* and *contrast*; first column in Supplementary Figure 9). Kim et al.[25] separated specular highlights and lowlights by subtracting an image of a rendered glossy surface (with a specular and diffuse component) from an "equivalent" fully diffuse image with the same total reflectance as the glossy surface (i.e., the surfaces reflected the same amount of light in total; for the diffuse surface light was scattered equally in all directions, and for the glossy surface some light was reflected specularly with the rest scattered diffusely). Kim et al. used the Ward model[68] where the total reflectance could be easily matched between glossy and diffuse renderings because specular reflectivity is constant at all viewing angles. Since the Principled BSDF simulates the Fresnel effect (whereby specular reflectance increases at grazing viewing angles, depending on the index of refraction), the diffuse (1) renderings were matched to the facing (along normal) reflectivity of the purely specular renderings. The second diffuse image (diffuse image 2) was created by rendering only the diffuse component of the original (full rendered) stimulus, with the specular component turned off.

Specular reflections were segmented into specular highlights and lowlights by subtracting the diffuse image 1 from the purely specular image, which resulted in a subtracted image. This subtracted image was thresholded to serve as a coverage mask (second column in Supplementary Figure 9). For best results for our stimulus set, the threshold was set to one third of the maximum diffuse shading for each stimulus. Pixels above this threshold were considered highlights, and the remaining pixels were considered lowlights. For calculations of sharpness and contrast, specular reflections were segmented from diffuse shading by subtracting the diffuse image 2 from the original "full" stimulus (second column in Supplementary Figure 9).

Visual features: Definition of cues. Visual features (cues) were calculated using RGB images. *Gloss cues: Coverage* was defined as the proportion of object pixels that were calculated to be specular highlights (excluding lowlights); *contrast* was the sum of root-mean-squared (RMS) contrast of extracted specular reflections at different spatial frequency bandpasses; *sharpness* of extracted specular reflections was calculated for each pixel within the highlight regions using a measure of local phase coherence (we used the MATLAB implementation of the method by Hassen et al.[69], available here: https://ece.uwaterloo.ca/~z70wang/research/lpcsi/), then these values were averaged. *Colour cues: Highlight saturation* and *lowlight saturation* were calculated as the average colour saturation of pixels within the highlight region, and outside of the highlight region (which we call the lowlight region), respectively, as measured by the rgb2hsv function in MATLAB (release 2018b, Mathworks, Natick, MA), which converts the red, green, and blue values of an RGB image to hue, saturation, and value (HSV) values of an HSV image (https://www.mathworks.com/help/matlab/ref/rgb2hsv.html); *lowlight value* was calculated as the average value of pixels in the lowlight region (also using the rgb2hsv MATLAB function). A root

transformation on sharpness and contrast was applied to linearize the relationship of these cues to gloss ratings.

Visual features: Measurements for Experiment 4 (feature manipulation). The same visual feature measurements were used for the rendered stimuli (from Experiments 2 and 3) and the manipulated images (from Experiment 4), so that the measurements would be comparable (see Supplemental Figure 14). However, slight modifications had to be made. For the rendered stimuli (Experiment 2), the segmentation between highlights and lowlights (described in the previous section) relied on two additional rendered images: a rendered specular image and a rendered diffuse image (1) with the same total reflectance (Supplementary Figure 9). For Experiment 4, the image manipulations are all applied to the same "Glazed Ceramic" material (see Methods). We reused the corresponding rendered specular and diffuse images from the stimulus set from Experiment 2, with one additional step: the specular image was further modified by the normal-based blur filter to account for the change in highlight coverage induced by a change in sharpness. Apart from this, the rest of the visual feature measurement routines remained unchanged.

Linear discriminant analysis (LDA): Experiment 2 (18-AFC). Materials were classified based on linear combinations of visual features using multi-class LDA in MATLAB (release 2018b, Mathworks, Natick, MA), which finds a set of linear combinations that maximises the ratio of between-class scattering to the within-class scattering (https://www.mathworks.com/help/stats/discriminant-analysis.html). The MATLAB function *fitdiscr* was used to fit a discriminant analysis classifier to the training data (regularised linear discriminant analysis where all classes have the same covariance matrix). To train a classifier, the fitting function estimates the parameters of a Gaussian distribution for each class (https://www.mathworks.com/help/stats/creating-discriminant-analysis-model.html). Test data labels were predicted with the MATLAB function *predict*, where the trained classifier finds the class with the smallest misclassification cost (https://www.mathworks.com/help/stats/prediction-using-discriminant-analysis-models.html).

Correlations: Experiment 3 (gloss ratings). Intersubject correlations for gloss ratings were calculated using Pearson correlation, and the median was taken. Pearson correlations used for analyses in Figure 5d and f were Fisher-transformed.

Linear regression: Experiment 3 (gloss ratings). A Linear regression was performed on 920 stimuli predicting mean gloss ratings from the three gloss cues (coverage, sharpness, contrast). Only stimuli that were allocated a dimension in the factor analysis (from Experiment 2) were included in this analysis (four stimuli loaded negatively onto dimensions, but not enough to make a new dimension for our stimulus set).

DATA AVAILABILITY

Psychophysics data and stimuli used in the experiments are available on Zendo: DOI: 10.5281/zenodo.5080227. The 3D meshes of the bunny and dragon objects were obtained from the Stanford 3D Scanning Repository and can be found under the following link: http://graphics.stanford.edu/data/3Dscanrep/.

CODE AVAILABILITY

Code for analyses (including image analyses) are available on Zendo: DOI: 10.5281/zenodo.5080227.

ACKNOWLEDGEMENTS

A.C.S. and K.D. are supported by a Sofja Kovalevskaja Award endowed by the German Federal Ministry of Education, awarded to K.D. K.D. was also supported by The Adaptive Mind, a research cluster funded by the Hessian Ministry of Higher Education, Research, Science and the Arts, Germany. A.C.S. was also supported by a Walter Benjamin Fellowship funded by the German Research Foundation (DFG). P.B. is supported by ANR project VIDA (ANR-17-CE23-0017). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Marios Panayi, Chris Baker, Filipp Schmidt, and Karl Gegenfurtner for helpful feedback on earlier versions of this manuscript.

AUTHOR CONTRIBUTIONS STATEMENT

A.C.S. & K.D. contributed to the conception and design of the work; the acquisition, analysis, and interpretation of data; and wrote the manuscript and revised it. P.B. contributed to the conception and design of the work; the interpretation of data; to the writing the manuscript and its revisions. P.B. also designed and implemented the image manipulation techniques.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

FIGURE LEGENDS/CAPTIONS

Figure 1. The relationship between reflectance properties, image structure, and material appearance. a. Most objects that we see every day are made from materials whose specular reflectance properties produce characteristic image structure. The appearance of these reflections is determined (and constrained) by the way in which these materials scatter light, in addition to other generative processes (see Figure 2). **b.** However, the extent and mechanism by which specular structure contributes to categorisation remains unknown. A feedforward view of neural processing **(i)** assumes that categories are defined by combinations of estimated mid-level properties like gloss, colour, and apparent shape, etc., which the visual system tries to "recover" from the image[31]. In contrast, the simultaneous hypothesis **(ii)** assumes that the visual system naturally learns about statistical variations (regularities) in image structure, from which the identity or category of a material can be "read out" simultaneously with surface qualities like gloss[29]. **c.** Some aspects of a surface's appearance (like perceived gloss) tend to correlate with physical properties (like specular reflectance), all else being equal: Systematically manipulating specular reflectance properties of otherwise identical objects causes salient visual differences in the appearance of highlights, which affects perceived gloss in predictable ways (see Figure 2). Specifically, increasing specular strength (bottom to top) increases the contrast of specular highlights, causing the dragon to appear glossier; increasing specular roughness (from left to right) decreases the clarity (or sharpness) of the specular highlights, causing the dragon to appear less glossy. These manipulations also affect our *qualitative* (i.e., categorical) impressions: the surfaces resemble different materials like glazed ceramic, glossy plastic, dull plastic, rubber, polished metal and brushed metal. Since shape, surface colour, and illumination conditions are held constant, all visual differences are caused by differences in specular reflectance properties, suggesting that specular structure may directly contain diagnostic information about material class, simultaneously with surface gloss (note that for some materials like gold, colour information also contributes to perceptual classification). **d.** Reflectance parameters that were manipulated in the experiments, and examples of how this affected the visual appearance of surfaces.

Figure 2. Visual features linked with specular image structure. a-c. The surface appearance of glossy objects is constrained by generative processes, which determine the coverage, sharpness, and contrast of specular highlights. Specifically, image structure is constrained by the way that surface reflectance properties interact with 3D surface geometry and the light field[32]. *Microroughness:* Whereas matte surfaces scatter light in all directions, glossy surfaces reflect light directionally, preserving the structure of the illumination field. Highly smooth surfaces produce narrow specular lobes[33] that increase the contrast and clarity (sharpness) of the reflected image relative to microscopically rougher surfaces, which produce broader specular lobes that blur this structure. Surfaces with higher microroughness also cause more spread-out specular highlights (higher coverage) due to the larger range of surface normal orientations reflecting bright light sources towards the observer. *Specular strength:* While dielectric materials like plastic reflect only a proportion of light specularly, metals reflect all light specularly, increasing the contrast of the reflected image. *Illumination sources:* Because of the structure-preserving properties of glossy objects, the contrast, sharpness, and coverage of specular highlights depend on the incident light that is being reflected, i.e., the intensity and structure of the illumination field. *3D shape:* Furthermore, 3D shape distorts this illumination structure. Specular reflections cling to points of high surface curvature and are elongated along (but slide rapidly across) directions of minimal surface curvature[23,24,34,35]. This means that 3D shape and observer viewpoint affect the location and distortion of specular reflections and thus proximal stimulus properties of coverage, sharpness, and contrast of the highlights. **d-f.** The brightness and colour saturation inside and outside of the highlight region is also determined by interactions between a surface's absorption/reflectance properties, 3D shape, and the illumination field. For example, specular reflections from dielectric materials like plastic preserve the spectral content of incident light; hence they look tinted only if the prevailing illumination is itself coloured. In contrast, coloured metals affect the tint of reflected light. We refer to (a-c) as *gloss cues* and (d-f) as *colour cues*.

Figure 3. Multiple semantic terms can describe the same qualitative visual appearance of a material. a. The object in the first panel has a dark brown, lumpy surface with medium-clarity specular reflections and could be labelled “melted chocolate” or “mud”. This surface has a different visual quality to the object in the middle panel, which has low-clarity, dim reflections that make it look like “latex” or “rubber”. The object in the third panel has very rough, anisotropic specular reflections that rapidly change in brightness at the boundaries between highlights and lowlights, giving it the visual characteristics of “velvet”, “silk”, or “fabric”. **b.** Sum of confidence ratings from the 18-AFC experiment (Experiment 2) for each stimulus and each category for the first 15 stimuli (out of 924). *Category profiles* are the distribution of category responses for each stimulus and reveal the extent to which multiple category terms apply to the same stimulus. *Stimulus profiles* are the distribution of stimuli that were allocated to each category and reveal the extent to which category terms are correlated with one another. Stimulus profiles were used to reveal the latent perceptual space of materials for our stimuli (factor analysis; see Figure 4), and category profiles were used to calculate material dissimilarity scores between each pair of stimuli (representational similarity analysis; see Supplementary Analysis B & Supplementary Figure 15).

Figure 4. Factor analysis performed on stimulus profiles obtained from the 18-AFC task. a. Black circles: cumulative total variance explained by each component from an initial principal components analysis. Red squares: Cumulative shared variance explained by each factor from the 12-factor solution, which accounted for 80% of the shared variance between stimulus profiles. **b.** Example stimuli from the 13 emergent material dimensions from the factor analysis. Supplementary Figure 8 shows more example stimuli for each shape and light field. **c.** Heat plot of the factor loadings for each category from the 18-AFC task, with blue and red cells showing positive and negative loadings, respectively. The highlighted cells show the factor onto which each category most strongly loaded. Note that the 13th dimension emerged from the negative loadings onto Factor 4. The emergent dimensions were highly interpretable from the category loadings onto the factors; for example, uncoloured metals like steel and silver all loaded strongly onto factor 1, forming a single dimension. We gave material category labels to each dimension (shown in **b**); however, note that these are arbitrary and are only included for interpretative convenience. Supplementary Figure 7 shows the results of a principal components analysis that retained all dimensions in addition to the results of other factor solutions, whose dimensions overlap with those here.

Figure 5. Material class is not determined by but may constrain gloss perception. a. Results of a linear regression model predicting perceived gloss from coverage, sharpness, and contrast (gloss cues). The model accounts for 76% of the variance in gloss ratings between stimuli and is comparable to a model that predicts participant’s gloss ratings from each other (leave one out; average R^2=.72). **b.** Histograms show that stimuli from the same material class exhibited a wide distribution of gloss levels, which persisted even when only the strongest loading stimuli on each dimension were considered (Supplementary Figure 10). Although some dimensions have a narrower range of gloss levels than others, stimuli from very visually distinct material dimensions like ceramics and (gold and uncoloured) metals have completely overlapping distributions of gloss ratings. Subjacent scatter plots show correlations between perceived gloss and material score, for each material class. Red coefficients indicate statistically significant correlations (uncoloured metals: r=0.58, p<0.001; ceramics: r=0.55, p<0.001; rubber-like: r=-0.31, p=0.001; gold metals: r=0.47, p<0.001; glazed porcelain: r=0.61, p<0.001; plastic: r=0.31 p=0.004; melted chocolate r=0.49, p<0.001). The results hold for other factor solutions (Supplementary Figure 11). **c.** The variance in gloss accounted for by the gloss cues differed within each material class (R^2, black bars), and the strength and direction of each cue’s correlation with gloss ratings / material score differed across materials (Pearson correlation colour coded according to

strength and direction; asterisked correlations indicate $p<0.05$; see Supplementary Figure 13 for correlation between cues). **d.** The correlations between perceived gloss and material score (from b) are confounded with the material scores correlating with the gloss cues themselves. In (d) z stands for Fisher-transformed correlation coefficients (Pearson correlation). Datapoints in (a) and (d) are colour coded by material category (see legend).

Figure 6. Linear discriminant analysis (LDA) predicting material category from visual features. a. The radial violin plots show the distribution of measured visual features for each material. For each feature, solid lines correspond to the 50^{th} percentile and dashed lines to the 25^{th} and 75^{th} percentile of the distribution (n=36 stimuli per plot). See text for details. **b.** The results of a linear classifier with leave one condition out validation procedure. The red dotted line indicates chance level (1/13). **c.** Classification accuracy generalises across different illumination and shape conditions. See Supplementary Figure 14 for similar results with other dimension reduction solutions. **d.** The stimuli are plotted in linear discriminant space (LD1 and LD2 stand for linear discriminant 1 and 2, respectively). Points are colour coded by either category (left), or visual features (right). **e.** The same stimuli are plotted for different linear discriminants (top: LD1 vs. 4; bottom: LD 2 vs. 3. These plots illustrate that the discriminations made by the model are perceptually intuitive: For example, a combination of average saturation and brightness within the lowlight region can be used to discriminate materials like porcelain (bright, uncoloured body) and waxy materials (bright, coloured body) from other materials with darker body colours; saturation within the highlight region is useful for discriminating brown metals (coloured highlights) from solid chocolate (uncoloured highlights). See Supplementary Table 1 for LDA weights. Supplementary Analysis B and Supplementary Figure 15 further support the idea that visual features predict material changes better than gloss perception.

Figure 7. Manipulating visual features transforms perceived material category. a. The manipulations were performed on specular and diffuse images separately before being recombined: as with the original feature measurements, the specular component was obtained by subtraction of the diffuse component from the full rendering (see Supplementary Figure 9). The sharpness of reflections was first modified by blurring pixels of the specular component that have sufficiently similar normals (see Methods). The subsequent filters adjusted the colour and intensity of each component. For instance, for materials like gold (top manipulated image), the specular component is first multiplied by a colour (the diffuse component is multiplied by 0); then its intensity and saturation are adjusted. For materials like pearl (bottom manipulated image), the intensity of the diffuse component is also adjusted. See the Methods section for a detailed description of each filter, along with the special filter used for velvet, based on a non-monotonic remapping of the specular component. **b.** Example stimuli after simple manipulations (top) and complex manipulations (bottom). White bars show the stimulus factor scores after participant judgments (see main text). After applying simple feature manipulations, materials sometimes resembled unintended categories (red arrows). The perceived materials align much better with the intended category (green arrows) after the complex feature manipulations. **c.** and **d.** Heat maps showing average factor scores for stimuli from each intended category after simple and complex feature manipulations, respectively. The highlighted cells show the factor onto which each category most strongly loaded

REFERENCES

1. Foster, D. H. Color Constancy. *Vision Res.* **51**, 674–700 (2011).
2. Anderson, B. L. Visual perception of materials and surfaces. *Curr. Biol.* **21**, R978–R983 (2011).
3. Anderson, B. L. Mid-level vision. *Curr. Biol.* **30**, R105–R109 (2020).
4. Chadwick, A. C. & Kentridge, R. W. The perception of gloss: A review. *Vision Res.* **109**, 221–235 (2015).
5. Fleming, R. W. Visual perception of materials and their properties. *Vision Res.* **94**, 62–75 (2014).
6. Fleming, R. W. Material Perception. *Annu. Rev. Vis. Sci.* **3**, 365–388 (2017).
7. Balas, B. Children's use of visual summary statistics for material categorization. *J. Vis.* **17**, 22 (2017).
8. Baumgartner, E., Wiebel, C. B. & Gegenfurtner, K. R. Visual and Haptic Representations of Material Qualities. *Multisens. Res.* **26**, 429–455 (2013).
9. Nagai, T., Hosaka, Y., Sato, T. & Kuriki, I. Relative contributions of low- and high-luminance components to material perception. *J. Vis.* **18**, 1–19 (2018).
10. Fleming, R. W., Wiebel, C. B. & Gegenfurtner, K. R. Perceptual qualities and material classes. *J. Vis.* **13**, 1–20 (2013).
11. Lagunas, M., Serrano, A., Gutierrez, D. & Masia, B. The joint role of geometry and illumination on material recognition. *J. Vis.* **21**, 1–18 (2021).
12. Nagai, T. *et al.* Temporal properties of material categorization and material rating: visual vs non-visual material features. *Vision Res.* **115**, 259–270 (2015).
13. Norman, J. F., Todd, J. T. & Phillips, F. Effects of illumination on the categorization of shiny materials. *J. Vis.* **20**, 1-16 (2020).
14. Sharan, L., Rosenholtz, R. & Adelson, E. H. Accuracy and speed of material categorization in real-world images. *J. Vis.* **14**, 1–24 (2014).
15. Tamura, H., Higashi, H. & Nakauchi, S. Dynamic Visual Cues for Differentiating Mirror and Glass. *Sci. Rep.* 1–12 (2018) doi:10.1038/s41598-018-26720-x.
16. Todd, J. T. & Norman, J. F. The visual perception of metal. *J. Vis.* **18**, 1-17 (2018).
17. Wiebel, C. B., Valsecchi, M. & Gegenfurtner, K. R. The speed and accuracy of material recognition in natural images. *Atten. Percept. Psychophys.* **75**, 954–66 (2013).
18. Beck, J. & Prazdny, K. Highlights and the perception of glossiness. *Percept. Psychophys.* **30**, 407–410 (1981).
19. Blake, A. & Bülthoff, H. H. Does the brain know the physics of specular reflection? *Nature* **343**, 165–168 (1990).
20. Wendt, G., Faul, F. & Mausfeld, R. Highlight disparity contributes to the authenticity and strength of perceived glossiness. *J. Vis.* **8**, 1–10 (2008).
21. Todd, J. T., Norman, J. F. & Mingolla, E. Lightness Constancy in the Presence of Specular Highlights. *Psychol. Sci.* **15**, 33–39 (2004).
22. Anderson, B. L. & Kim, J. Image statistics do not explain the perception of gloss and lightness. *J. Vis.* **9**, 1–17 (2009).
23. Kim, J., Marlow, P. J. & Anderson, B. L. The perception of gloss depends on highlight congruence with surface shading. *J. Vis.* **11**, 1–19 (2011).
24. Marlow, P. J., Kim, J. & Anderson, B. L. The role of brightness and orientation congruence in the perception of surface gloss. *J. Vis.* **11**, 1–12 (2011).
25. Kim, J., Marlow, P. J. & Anderson, B. L. The dark side of gloss. *Nat. Neurosci.* **15**, 1590–1595 (2012).
26. Komatsu, H. & Goda, N. Neural mechanisms of material perception: quest on Shitsukan. *Neurosci.* **392**, 329-347 (2018).
27. Schwartz, G. & Nishino, K. Recognizing material properties from images. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 1981–1995 (2020).

28. Tanaka, M. & Horiuchi, T. Investigating perceptual qualities of static surface appearance using real materials and displayed images. *Vision Res.* **115**, 246–258 (2015).
29. Fleming, R. W. & Storrs, K. R. Learning to see stuff. *Curr. Opin. Behav. Sci.* **30**, 100–108 (2019).
30. Storrs, K. R. & Fleming, R. W. Learning About the World by Learning About Images. *Curr. Dir. Psychol. Sci.* **30**, 95-192 (2021).
31. Barrow, H. G. & Tenenbaum, J. M. Recovering intrinsic scene characteristics from images. Computer Vision Systems. *Comput. Vis. Syst.* **2**, 3-26 (1978).
32. Marlow, P. J. & Anderson, B. L. Generative constraints on image cues for perceived gloss. *J. Vis.* **13**, 1–23 (2013).
33. Klinker, G. J., Shafer, S. A. & Kanade, T. The measurement of highlights in color images. *Int. J. Comput. Vis.* **2**, 7–32 (1988).
34. Fleming, R. W., Torralba, A. & Adelson, E. H. Specular reflections and the perception of shape. *J. Vis.* **4**, 798–820 (2004).
35. Koenderink, J. J. & van Doorn, A. J. Photometric invariants related to solid shape. *Opt. Acta Int. J. Opt.* **27**, 981–996 (1980).
36. Marlow, P. J., Kim, J. & Anderson, B. L. The Perception and Misperception of Specular Surface Reflectance. *Curr. Biol.* **22**, 1909–1913 (2012).
37. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis – connecting the branches of systems neuroscience. Frontiers in Systems. **2**, 1–28 (2008).
38. Pellacini, F., Ferwerda, J. A. & Greenberg, D. P. Toward a psychophysically-based light reflection model for image synthesis. *Proc. SIGGRAPH* 55-64 (2000).
39. Vangorp, P., Barla, P. & Fleming, R. W. The perception of hazy gloss. *J. Vis.* **17**, 19 (2017).
40. Toscani, M., Guarnera, D., Guarnera, G. C., Hardberg, J. Y. & Gegenfurtner, K. R. Three perceptual dimensions for specular and diffuse reflection. *ACM Trans. Appl. Percept.* **1**, 1–27 (2020).
41. Hunter, R. S. Methods of determining gloss. *J. Res. Natl. Bur. Stand.* **18**, 19-41 (1937).
42. Xiao, B. *et al.* Looking against the light: How perception of translucency depends on lighting direction. *J. Vis.* **14**, 1–22 (2014).
43. Irawan, P. & Marschner, S. Specular reflection from woven cloth. *ACM Trans. Graph.* **31**, (2012).
44. Fleming, R. W. Human Perception: Visual Heuristics in the Perception of Glossiness. *Curr. Biol.* **22**, R865–R866 (2012).
45. Purves, D., Morgenstern, Y. & Wojtach, W. T. Will understanding vision require a wholly empirical paradigm? *Front. Psychol.* **6**, 1–6 (2015).
46. Okazawa, G., Koida, K. & Komatsu, H. Categorical properties of the color term 'GOLD'. *J. Vis.* **11**, 1–19 (2011).
47. Pasupathy, A., Kim, T. & Popovkina, D. V. Object shape and surface properties are jointly encoded in mid-level ventral visual cortex. *Curr. Opin. Neurobiol.* **58**, 199–208 (2019).
48. Schmid, A. C. & Anderson, B. L. Perceptual dimensions underlying lightness perception in homogeneous center-surround displays. *J. Vis.* **17**, 1-20 (2017).
49. Mooney, S. W. J. & Anderson, B. L. Specular Image Structure Modulates the Perception of Three-Dimensional Shape. *Curr. Biol.* **24**, 2737–2742 (2014).
50. Wijntjes, M. W. A., Doerschner, K., Kucukoglu, G. & Pont, S. C. Relative flattening between velvet and matte 3D shapes: Evidence for similar shape-from-shading computations. *J. Vis.* **12**, 1–11 (2012).
51. Storrs, K. R., Anderson, B. L. & Fleming, R. W. Unsupervised learning predicts human perception and misperception of gloss. *Nat. Hum. Behav.* **5**, 1402-1417 (2021).
52. Nishio, A., Goda, N. & Komatsu, H. Neural Selectivity and Representation of Gloss in the Monkey Inferior Temporal Cortex. *J. Neurosci.* **32**, 10780–10793 (2012).
53. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).

54. Long, B., Yu, C.-P. & Konkle, T. Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc. Natl. Acad. Sci.* **115**, E9015–E9024 (2018).
55. Bracci, S., Ritchie, J. B. & de Beeck, H. O. On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia* **105**, 153–164 (2017).
56. Kaiser, D., Azzalini, D. C. & Peelen, X. M. V. Shape-independent object category responses revealed by MEG and fMRI decoding. J Neurophysiol. **115**, 2246-2250 (2016).
57. Zeman, A., Ritchie, J. B., Bracci, S. & Beeck, H. Op de. Orthogonal representations of object shape and category in deep convolutional neural networks and human visual cortex. Sci. Rep. **10**, 2453 (2019).
58. Schmid, A. C. & Doerschner, K. Representing stuff in the human brain. *COBEHA* **30**, 178–185 (2019).
59. Malcolm, G. L., Groen, I. I. A. & Baker, C. I. Making Sense of Real-World Scenes. *Trends Cogn. Sci.* **20**, 843–856 (2016).
60. Groen, I. I. A., Silson, E. H. & Baker, C. I. Contributions of low-and high-level properties to neural processing of visual scenes in the human brain. Phil. Trans. R. Soc. B **372**, 20160102 (2017).
61. Debevec, P. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proc. SIGGRAPH* 189-198 (1998).
62. Burley, B. Physically-based shading at Disney. In ACM SIGGRAPH 2012 Course: Practical Physically-based Shading in Film and Game Production. SIGGRAPH'12. <https://disney-animation.s3.amazonaws.com/library/s2012_pbs_disney_brdf_notes_v2.pdf> (2012).
63. Sawayama, M. & Nishida, S. Material and shape perception based on two types of intensity gradient information. *PLoS Comput. Biol.* **14**, 1-40 (2018).
64. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
65. Schmid, A. C. & Doerschner, K. Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials. *J. Vis.* **18**, 1-32 (2018).
66. Di Cicco, F., Wijntjes, M. W. A. & Pont, S. C. Understanding gloss perception through the lens of art: Combining perception, image analysis, and painting recipes of 17th century painted grapes. *J. Vis.* **19**, 1-15 (2019).
67. van Assen, J. J. R., Barla, P. & Fleming, R. W. Visual Features in the Perception of Liquids. *Curr. Biol.* **28**, 1-7 (2018).
68. Ward, G. J. The RADIANCE lighting simulation and rendering system. *Proc. 21st Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH* **28**, 459–472 (1994).
69. Hassen, R., Wang, Z. & Salama, M. M. A. Image sharpness assessment based on local phase coherence. *IEEE Trans. Image Process.* **22**, 2798–2810 (2013).

b

i. FEEDFORWARD HYPOTHESIS

Image structure → **Property estimation** → **Categorisation**

(process "low-level" features like contrast and sharpness of image gradients)

(recover "mid-level" properties like shape, gloss, surface colour, translucency, etc.)

(apply "high-level" conceptual identity, e.g. porcelain, pearl plastic, metal, etc.)

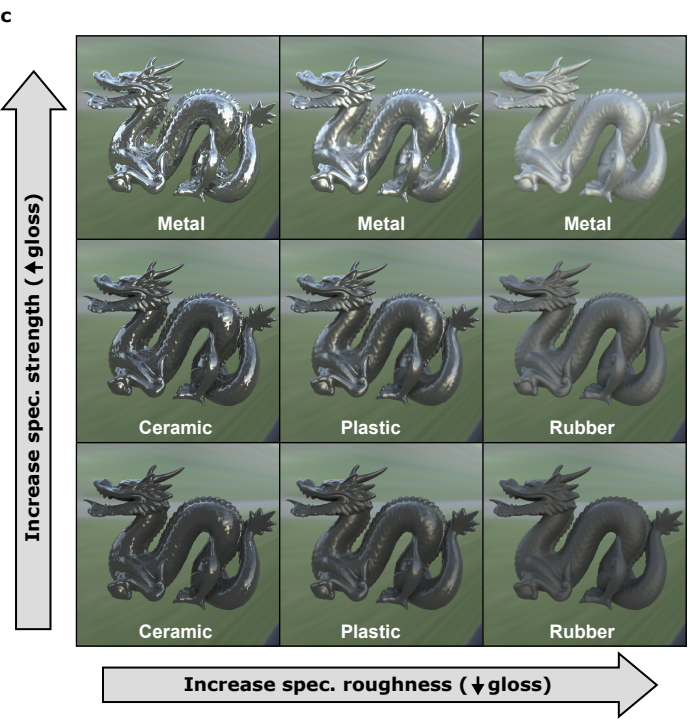
ii. SIMULTANEOUS HYPOTHESIS

Image structure → **"Mid-level" appearance** → **Property estimation** *(e.g. glossy, light-coloured, curved surface)*

"Mid-level" appearance → **Categorisation** *(e.g. porcelain)*

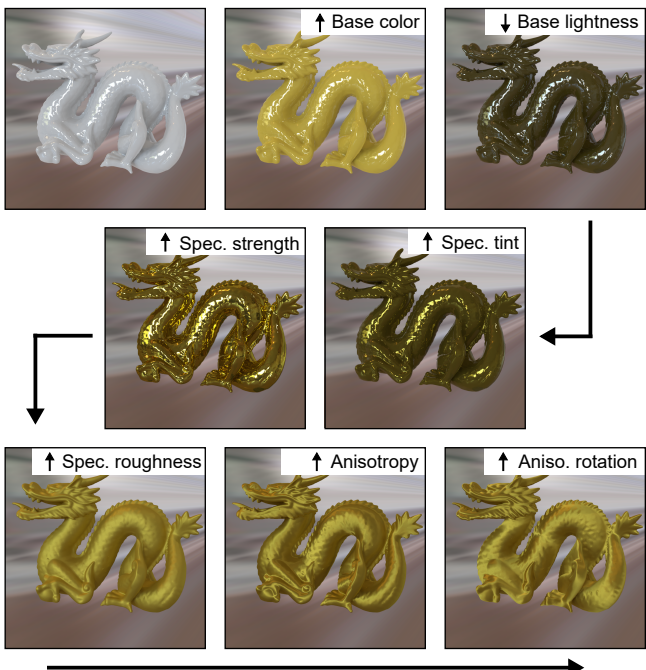
(sharpness, contrast etc. of image gradients)

(from cue combinations e.g. shallow diffuse shading gradients with sharp, bright highlights)



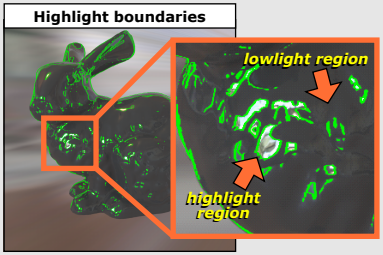
d

Effect of manipulating parameters

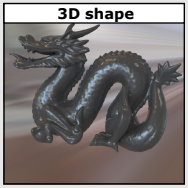


a Coverage

Area of surface covered by highlights



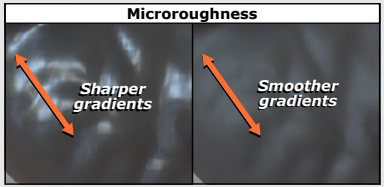
Affected by physical sources such as...

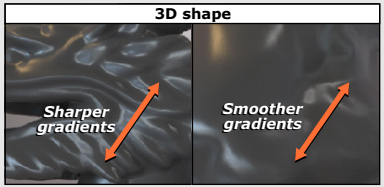


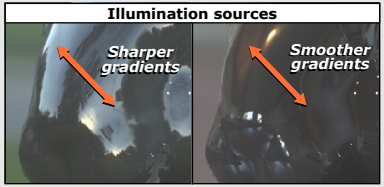
b Sharpness

Rapidness of brightness changes

Affected by physical sources such as...

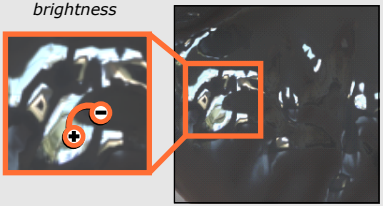




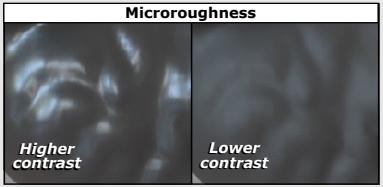


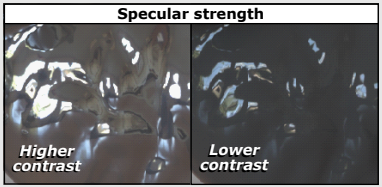
c Contrast

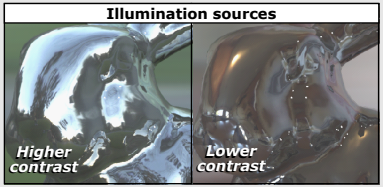
Variation in brightness



Affected by physical sources such as...







d Lowlight value

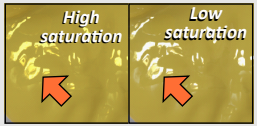
Brightness in the lowlight region

e Lowlight saturation

Colour saturation in the lowlight region

f Highlight saturation

Colour saturation in the highlight region



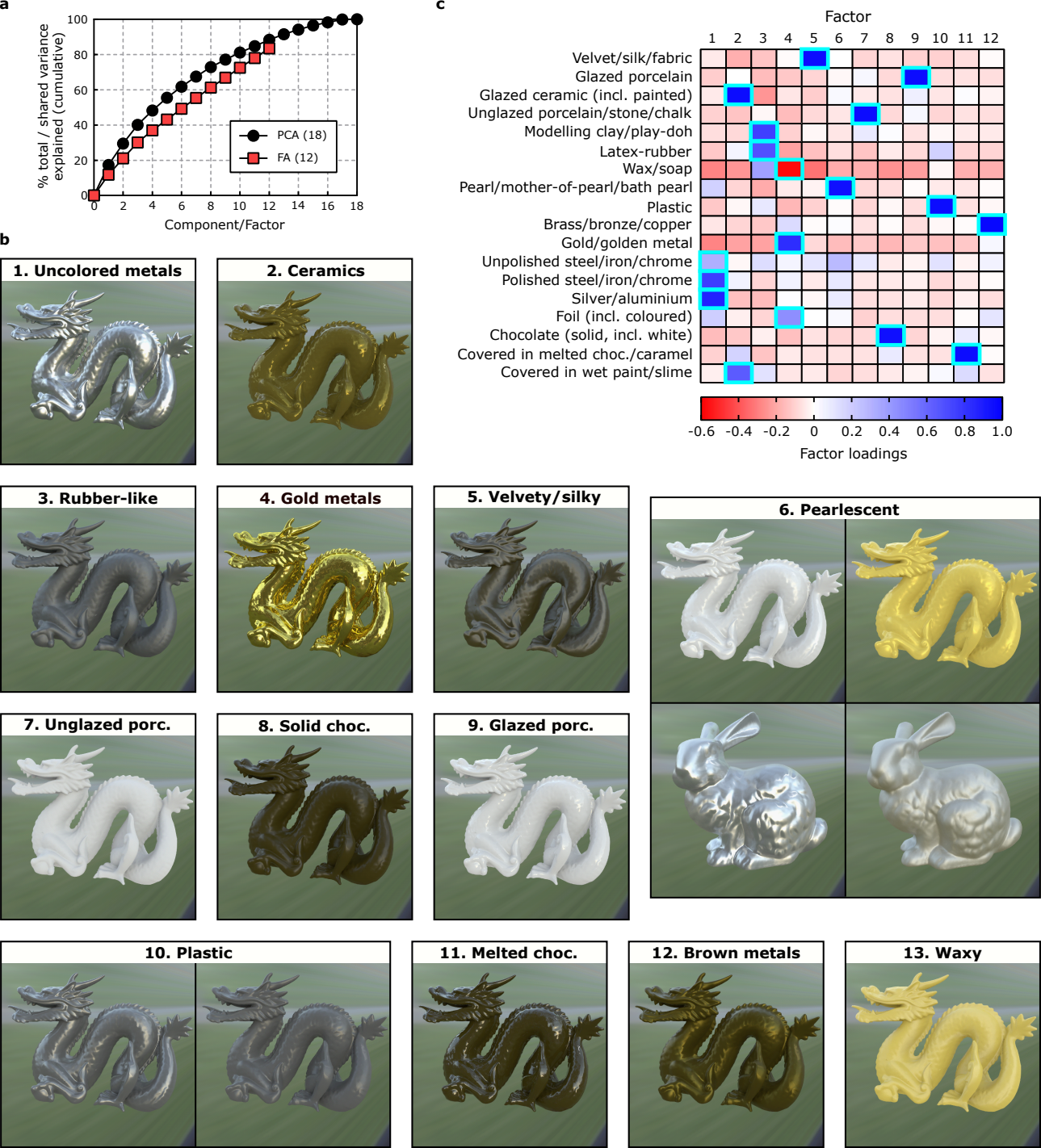
a

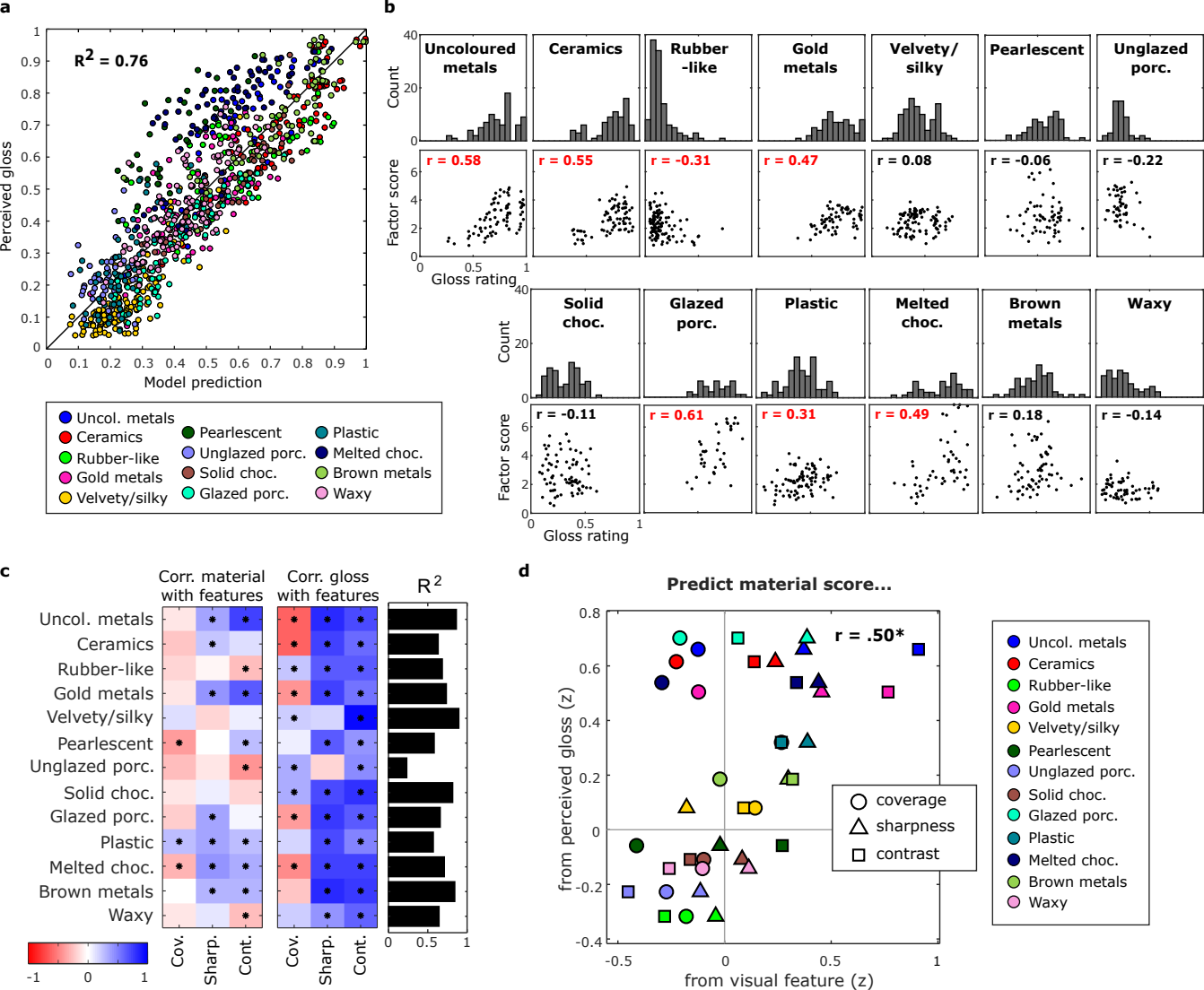
b

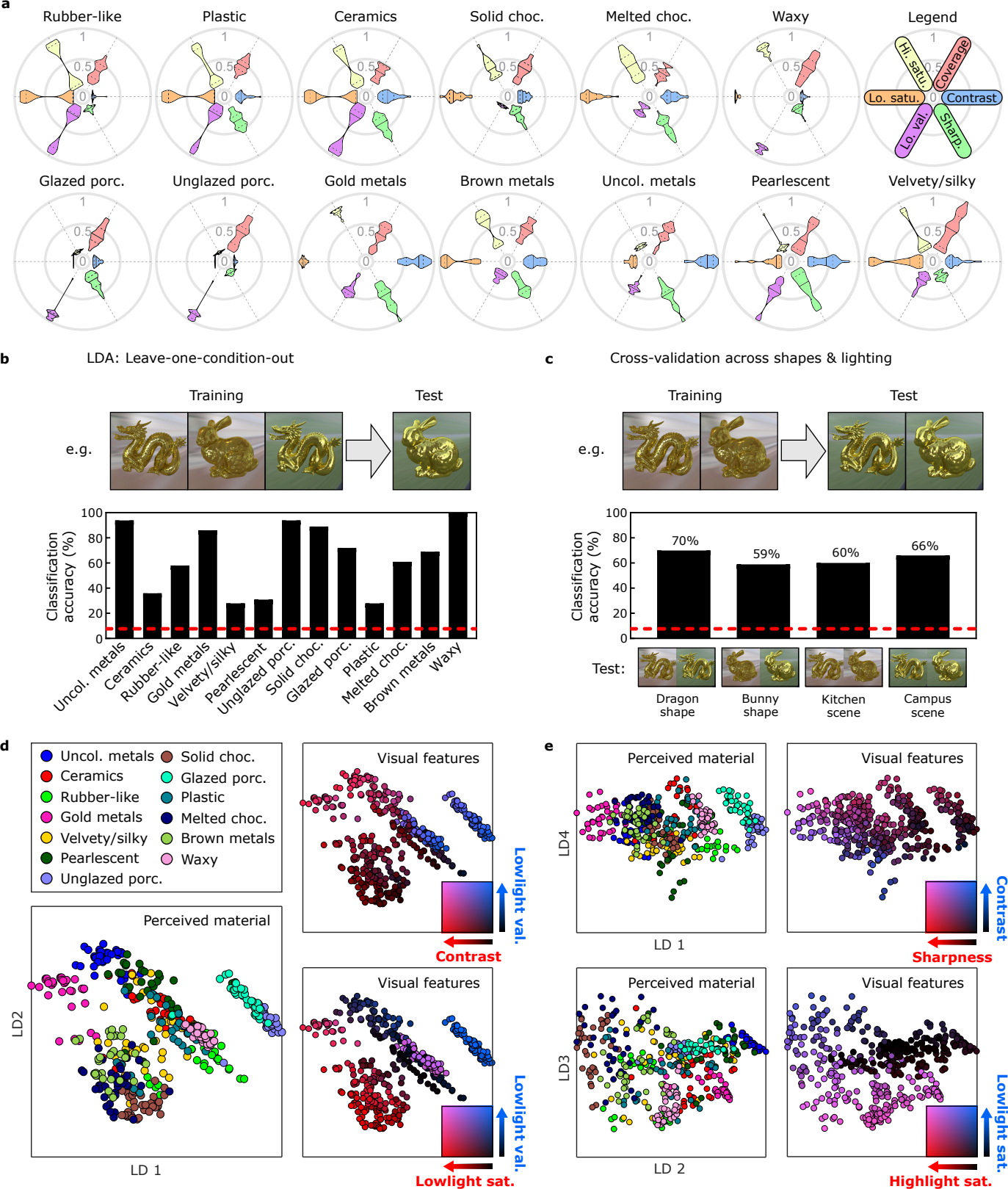
	Stim. 1	Stim. 2	Stim. 3	Stim. 4	Stim. 5	Stim. 6	Stim. 7	Stim. 8	Stim. 9	Stim. 10	Stim. 11	Stim. 12	Stim. 13	Stim. 14	Stim. 15	...	Stim. *n*
Velvet / silk / fabric												3		3			
Glazed porcelain	3			3													
Glazed ceramic	21			17				18			3		3		15		
Unglazed porcelain / stone / chalk			9		3	3	9							3			
Modelling clay / play-doh		9	15		15		5			3							
Latex / rubber	3	9	20	21	24	12	21		3		3	9	3	9			
Wax / soap			3	6	5	9	9		3	3	3	3	3				
Pearl / mother-of-pearl / bath pearl																	
Plastic	2	20	6	8	3	27	6		6	4	6						
Brass / bronze / copper								3	3	6			3				
Gold / golden metal																	
Unpolished steel / iron / chrome		6			3												
Polished steel / iron / chrome							3										
Silver / aluminium																	
Foil					3					3							
Solid chocolate		5		3					36	30	27	44	41	43			
Covered in melted chocolate / caramel		3						36	9	6	18		3		24		
Covered in web paint / slime	30	6	6		3	6	6	3		3			3		21		

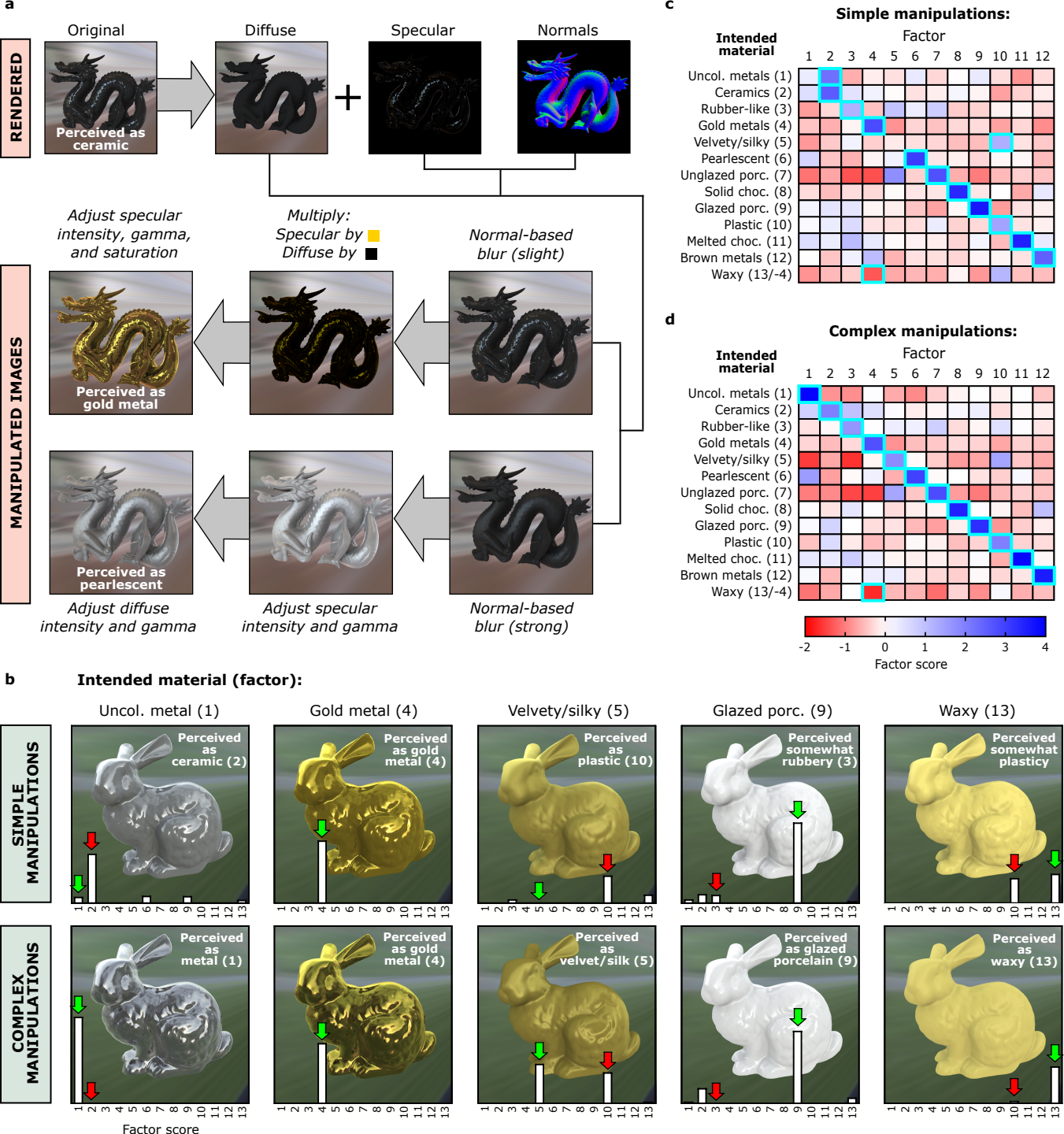
Stimulus profile for glazed ceramic

Category profile for stimulus 5


a
% total / shared variance explained (cumulative)
Component/Factor
PCA (18)
FA (12)
b
1. Uncolored metals
2. Ceramics
3. Rubber-like
4. Gold metals
5. Velvety/silky
6. Pearlescent
7. Unglazed porc.
8. Solid choc.
9. Glazed porc.
10. Plastic
11. Melted choc.
12. Brown metals
13. Waxy
c
Factor
Velvet/silk/fabric
Glazed porcelain
Glazed ceramic (incl. painted)
Unglazed porcelain/stone/chalk
Modelling clay/play-doh
Latex-rubber
Wax/soap
Pearl/mother-of-pearl/bath pearl
Plastic
Brass/bronze/copper
Gold/golden metal
Unpolished steel/iron/chrome
Polished steel/iron/chrome
Silver/aluminium
Foil (incl. coloured)
Chocolate (solid, incl. white)
Covered in melted choc./caramel
Covered in wet paint/slime
Factor loadings


a
R^2 = 0.76
Perceived gloss
Model prediction
Uncol. metals
Ceramics
Rubber-like
Gold metals
Velvety/silky
Pearlescent
Unglazed porc.
Solid choc.
Glazed porc.
Plastic
Melted choc.
Brown metals
Waxy
b
Uncoloured metals
Ceramics
Rubber -like
Gold metals
Velvety/ silky
Pearlescent
Unglazed porc.
Solid choc.
Glazed porc.
Plastic
Melted choc.
Brown metals
Waxy
Count
Factor score
Gloss rating
r = 0.58
r = 0.55
r = -0.31
r = 0.47
r = 0.08
r = -0.06
r = -0.22
r = -0.11
r = 0.61
r = 0.31
r = 0.49
r = 0.18
r = -0.14
c
Corr. material with features
Corr. gloss with features
R^2
Uncol. metals
Ceramics
Rubber-like
Gold metals
Velvety/silky
Pearlescent
Unglazed porc.
Solid choc.
Glazed porc.
Plastic
Melted choc.
Brown metals
Waxy
Cov.
Sharp.
Cont.
d
Predict material score...
r = .50*
from perceived gloss (z)
from visual feature (z)
coverage
sharpness
contrast
Uncol. metals
Ceramics
Rubber-like
Gold metals
Velvety/silky
Pearlescent
Unglazed porc.
Solid choc.
Glazed porc.
Plastic
Melted choc.
Brown metals
Waxy


a
Rubber-like
Plastic
Ceramics
Solid choc.
Melted choc.
Waxy
Legend
Hi. satu.
Coverage
Lo. satu.
Contrast
Lo. val.
Sharp.
Glazed porc.
Unglazed porc.
Gold metals
Brown metals
Uncol. metals
Pearlescent
Velvety/silky
b
LDA: Leave-one-condition-out
Training
Test
e.g.
Classification accuracy (%)
Uncol. metals
Ceramics
Rubber-like
Gold metals
Velvety/silky
Pearlescent
Unglazed porc.
Solid choc.
Glazed porc.
Plastic
Melted choc.
Brown metals
Waxy
c
Cross-validation across shapes & lighting
Training
Test
e.g.
Classification accuracy (%)
70%
59%
60%
66%
Test:
Dragon shape
Bunny shape
Kitchen scene
Campus scene
d
Uncol. metals
Ceramics
Rubber-like
Gold metals
Velvety/silky
Pearlescent
Unglazed porc.
Solid choc.
Glazed porc.
Plastic
Melted choc.
Brown metals
Waxy
Perceived material
LD1
LD2
Visual features
Lowlight val.
Contrast
Visual features
Lowlight val.
Lowlight sat.
e
Perceived material
LD4
LD 1
Visual features
Contrast
Sharpness
Perceived material
LD3
LD 2
Visual features
Lowlight sat.
Highlight sat.


a
RENDERED
Original
Diffuse
Specular
Normals
Perceived as ceramic
MANIPULATED IMAGES
Adjust specular intensity, gamma, and saturation
Multiply: Specular by Diffuse by
Normal-based blur (slight)
Perceived as gold metal
Perceived as pearlescent
Adjust diffuse intensity and gamma
Adjust specular intensity and gamma
Normal-based blur (strong)
c
Simple manipulations:
Intended material
Factor
1 2 3 4 5 6 7 8 9 10 11 12
Uncol. metals (1)
Ceramics (2)
Rubber-like (3)
Gold metals (4)
Velvety/silky (5)
Pearlescent (6)
Unglazed porc. (7)
Solid choc. (8)
Glazed porc. (9)
Plastic (10)
Melted choc. (11)
Brown metals (12)
Waxy (13/-4)
d
Complex manipulations:
Intended material
Factor
1 2 3 4 5 6 7 8 9 10 11 12
Uncol. metals (1)
Ceramics (2)
Rubber-like (3)
Gold metals (4)
Velvety/silky (5)
Pearlescent (6)
Unglazed porc. (7)
Solid choc. (8)
Glazed porc. (9)
Plastic (10)
Melted choc. (11)
Brown metals (12)
Waxy (13/-4)
-2 -1 0 1 2 3 4
Factor score
b
Intended material (factor):
Uncol. metal (1)
Gold metal (4)
Velvety/silky (5)
Glazed porc. (9)
Waxy (13)
SIMPLE MANIPULATIONS
Perceived as ceramic (2)
Perceived as gold metal (4)
Perceived as plastic (10)
Perceived somewhat rubbery (3)
Perceived somewhat plasticy
COMPLEX MANIPULATIONS
Perceived as metal (1)
Perceived as gold metal (4)
Perceived as velvet/silk (5)
Perceived as glazed porcelain (9)
Perceived as waxy (13)
Factor score