



Task-dependent extraction of information from videos of iridescent and glossy samples

LI SHIWEN,^{1,*}  TAKUMA MORIMOTO,^{1,2}  JULIE M. HARRIS,³ AND HANNAH E. SMITHSON¹ 

¹Department of Experimental Psychology, University of Oxford, Woodstock Road, OX2 6GG, Oxford, UK

²Department of General Psychology, Justus-Liebig-Universität Gießen, Gießen, Germany

³School of Psychology and Neuroscience, University of St Andrews, St Andrews, Fife, KY16 9JP, UK

*shiwen.li@psy.ox.ac.uk

Received 7 November 2022; revised 23 January 2023; accepted 24 January 2023; posted 25 January 2023; published 21 February 2023

We present an exploratory study on iridescence that revealed systematic differences in the perceptual clustering of glossy and iridescent samples that was driven by instructions to focus on either the material or the color properties of the samples. Participants' similarity ratings of pairs of video stimuli, showing the samples from multiple views, were analyzed using multidimensional scaling (MDS), and differences between the MDS solutions for the two tasks were consistent with flexible weighting of information from different views of the samples. These findings point to ecological implications for how viewers perceive and interact with the color-changing properties of iridescent objects.

Published by Optica Publishing Group under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

<https://doi.org/10.1364/JOSAA.479795>

1. INTRODUCTION

Assigning a color to an object seems like a simple task that we often perform to recognize and sort objects. The use of color-terms as adjectives that help define specific instances of an object is commonplace, such as the “red pen” or the “blue mug.” For color arising from absorption pigments, modification of the spectral content of incoming light is bound to the surface of the material itself and is invariant to the angle from which the surface is viewed. However, some color effects arise not from pigments but from nanoscale structures that generate optical interference that modifies the relative intensities of different wavelengths of light. This structural color depends heavily on the viewing angle and gives rise to the percept of iridescence. While a pigmented object is associated with a fixed color property, an iridescent object is not. In this paper, we report perceptual judgments of color samples coated with interference paint that produces viewing-angle-dependent color effects.

The perceptual constancy literature has grappled with the relationship between fixed properties of an object (such as its pigmentation) and the varied properties of the proximal image (such as chromaticity) associated with the object under specific viewing conditions. Surface colors that are generated through pigmentation are characterized by their surface reflectance function (the proportion of light reflected as a function of wavelength), and color constancy has been defined as the association of a fixed color appearance associated with surface reflectance, despite changes in the conditions of observing that modify the spectral content of light reaching the eye (see [1] for review).

For matte objects, the light reflected can be modeled as the wavelength-by-wavelength multiplication of the incident light and the surface reflectance, known as the diffuse reflection. Since the spectrum of the diffuse reflection depends on the illuminant spectrum, color constancy (of surface color appearance) requires some “discounting” of this influence.

For dielectric glossy objects, the reflected light is modeled as the sum of two components [2]—the diffuse component (as above) and the specular component (which carries the unmodified spectrum of the illuminant). Since the diffuse and specular components have different imaging geometries, regions of the proximal image differ in the relative contributions of the two. For example, specular highlights in the image are dominated by the specular component, and their location and extent in the proximal image depend on viewing geometry.

Perceiving constant object color from proximal images—“discounting the illuminant” and understanding the effect of glossiness on object color perception—has been well documented and investigated empirically [3–8]. The human perception of iridescent objects has received less attention (but see [9–15] for work in visual ecology and [16, 17] for rendering iridescence). While viewing geometry modifies the proximal image for both glossy and iridescent objects, for iridescent objects there is no fixed color property of the object; in these cases, the visual characteristic that identifies the material is the color-changing behavior. Iridescent samples prompt us to consider signatures of material perception that depend on dynamic sampling of objects as they are manipulated within

the lighting environment and with respect to the viewing angle. The focus of this study is to consider how participants extract information about materials from a sequence of frames of a video during which the surface normal of the sample rotates. Although prompted by the study of iridescence perception in the present paper, the dynamic signatures in the proximal images also provide information to separate specular and diffuse reflection [18], so they may represent a more widely applicable mode of viewing.

A. Our Study

The present study used material samples that differed in their darkness (opacity of a super-imposed black layer) and saturation (pigment-based), and their surface coating (containing only transparent gloss layers or transparent gloss layers plus a layer of interference paint). We used multidimensional scaling (MDS) to understand similarity judgments between the samples, judged in response to two different task instructions: (i) according to their “average color” and (ii) depending on “how they’re made.” Both instructions required participants to attribute the properties of the proximal stimuli to properties of objects in the world. In theory, two objects can be very similar in color but made from different materials and vice versa. Therefore, having two separate MDS questions allowed us to disentangle the effects of color and material percepts on perceived similarity. To allow full experimental control of the proximal stimuli that participants accessed, samples were presented as videos of real samples at a pre-determined set of orientations in the environment. Samples were flat patches of the material, and sample orientation was characterized as the tilt of the sample. We tested whether observers’ patterns of responses were consistent with modeled hypothetical observers that used summary statistics of the proximal images. In particular, we compared models that differed in the way they sampled and combined information from frames of the dynamic stimuli. Participants’ responses were consistent with different weightings of the dynamic information, depending on the task instructions. The first part of the paper presents the MDS methods and results, laying the foundation for the second part of the paper, which focuses on modeling these empirical findings.

2. METHODS

A. Samples

Stimuli were derived from photographs of real glossy and iridescent samples, constructed first by printing a blue base layer of color and then adding surface coatings. We opted to print the base layer, rather than applying paint manually, to maximize uniformity in base color and texture between samples. The parameters of the printed colors were defined in Adobe Photoshop CC 2019. The photoshop specification was multi-layered and comprised a uniform blue background, noise to introduce spatial inhomogeneities and texture, and a super-imposed black layer that had varying levels of opacity (see Fig. 1(A) for specific settings). For both glossy and iridescent samples, the base color was printed on HP Premium Photopaper using an HP Envy 5540 Inkjet printer. The printed samples then received the surface coatings summarized in Figs. 1(B) and 1(C). First, transparent nail varnish (off-the-shelf top-coat nail varnish) was applied to both glossy and iridescent samples since it produced a better surface for even application of the interference paint. For the iridescent samples, a layer of DecoArt Media Gold Acrylic Media paint (translucent from some angles and gold from others) was used to create the iridescent effect (angle-dependent color change from blue to gold). A second layer of transparent nail varnish was added to both types of samples since without this the iridescent samples looked visibly more matte. Finally, the finished samples were trimmed to 1×1 cm squares. Since the process of manual application produced slightly variable results, multiple copies of each sample were made, and the most consistent ones were selected by eye for use in the study.

The glossy and iridescent samples differed along two physical dimensions: (1) opacity of the black layer (0%; 40%; 60%; or 80%), with higher opacity resulting in a perceptually darker sample, and (2) surface coating (with and without interference paint), producing eight different samples in combination.

The presence of the color-changing effect is dependent on the viewer being able to view the samples from different angles. To convey the color-changing effect to participants, we photographed the samples from a fixed location at various tilts with respect to the lighting environment and combined the photographs into videos. The samples were illuminated using a TaoTronics LED Desk Lamp (set to 5500 k in “Reading mode” at maximum brightness) and photographed with a Canon 700D Digital SLR Camera (1/60 exposure, 7.1 f, ISO 800) in CR2 format. To achieve consistent control of tilt for each sample, the samples were glued to a magnetic dish that attached to a tiltable platform. The tilt was monitored using a smartphone

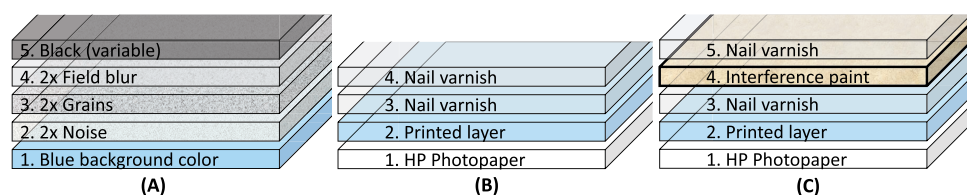


Fig. 1. Schematic description of the making of the samples, showing the structuring of layers from bottom (layer 1) to top. (A) Photoshop layers for printing: 1. Uniform blue background color (RGB = [54, 192, 255]). 2. Noise set to 96.61% amount, 8% size, 56% roughness, 0% color, and 100% highlights. 3. Grains set to 2% light bokeh, 0% bokeh color, light range between 191 and 255. 4. Field blur set to 15 pixels. 5. Varying opacity of a super-imposed black layer (0%, 40%, 60%, or 80%). Layers 2–4 in Photoshop were double layers. (B) and (C) show the order in which surface coatings were applied for (B) glossy and (C) iridescent samples.

app (Angles Pro) with a precision of $\pm 0.5^\circ$ and photographed at 5° steps, from -20° to 20° . A schematic diagram of the setup is shown in Fig. 2(A).

The resulting nine photographs of the samples formed the sequence of frames used to create videos of the samples tilting from left to right at 25 fps, and back to left, as illustrated in Fig. 3. The range of visual signals available to an observer is usually difficult to fully characterize as it is dependent on the geometry between the object, viewer, and light source. In our study, the range of image statistics is constrained to the range of pixels presented in the videos. For each tilt, the pixels corresponding to the sample were extracted (see Fig. 2(C)), and the intensities of the RGB channels of each pixel were transformed to CIELAB assuming the sRGB (standard RGB, as defined by Microsoft and HP) model. The summary CIELAB coordinate for each sample for each tilt (see Fig. 4) was derived by averaging spatially across the pixels corresponding to the sample. Toward the “middle” tilts (4–7), there is greater separation between the glossy and iridescent samples along the L^* and b^* axes than at the “extreme” tilts (1, 2, 8, 9).

B. Participants

The study was run online and was approved by the Medical Sciences Inter-Divisional Research Ethics Committee at University of Oxford in agreement with the Declaration of Helsinki (RE77333/RE001). In addition to 14 participants recruited via word of mouth in the early stages of the study, 40 additional participants participated via Prolific.org. Consent was obtained prior to the study from all participants, and they were provided with detailed instructions with the aim of standardizing the stimulus conditions as closely as possible. This included instructions to use macOS or Windows10 laptops only, disconnect any external monitors and use the laptop screen, connect the laptop to power, and complete the study in a dark environment.

C. Display Characteristics and Exclusion Criteria

Participants were asked to adjust display settings for: (1) brightness, (2) colorimetric control, and (3) gamma. For the brightness and colorimetric settings, participants were provided with instructions to disable automatic display adjustments on macOS and Windows OS systems. For macOS users, they were also asked to change their gamma settings to 2.2. Participants were asked to indicate whether each step was completed.

To check for display distortions that fall outside acceptable deviation from a standard display after the adjustments, all participants also completed a perceptual brightness matching task [19] in which the brightness of a uniform gray disk was matched to a binary pattern of black and white pixels presented in a surrounding annulus. In separate trials, the proportion of black pixels was set to $n/9$, where n could take values from 1 to 8. A linear model was used fit to each participant’s brightness matches (three repeats per gray-level), and those that fell below an R^2 of 0.8 ($N = 19$) were excluded from all other analyses in this paper. Second, participants whose data did not produce valid MDS solutions for *both* tasks ($N = 1$) were also excluded from the analyses. The lack of MDS conversion could be caused by large inconsistencies in their responses, which may be due to inattention or misunderstanding of task instructions. In total, 34 participants’ datasets were included in subsequent analyses. As the task in this study was a supra-threshold relative judgment of appearance (i.e., similarity rating), we expect distortions caused by non-optimal viewing conditions or display characteristics would have only minor effects on the results.

D. Multidimensional Scaling Tasks

To understand the perceptual differences and relationships between samples, we asked participants to rate their similarity, so that the similarity ratings could be analyzed using MDS. The 28 unique pairs of eight samples were randomized and repeated 5 times, resulting in 140 trials per task. The trials consisted of

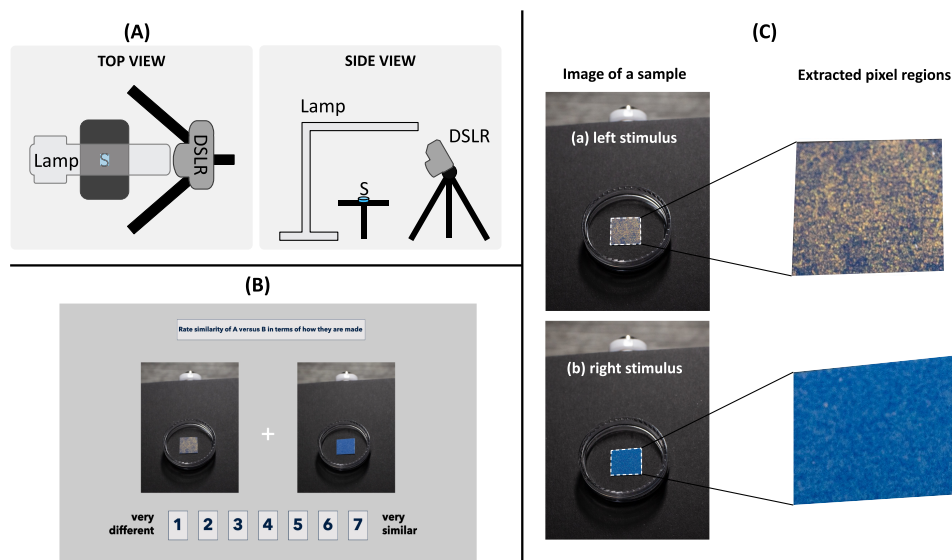


Fig. 2. (A) Simplified schematic of the photographic setup viewed from top-down and the side (approximate dimensions); “S” represents physical sample. (B) A screenshot of an example trial. (C) Regions of interest for modeling were manually chosen to contain only pixels that correspond to the image of the sample (shown in example trial in panel B).

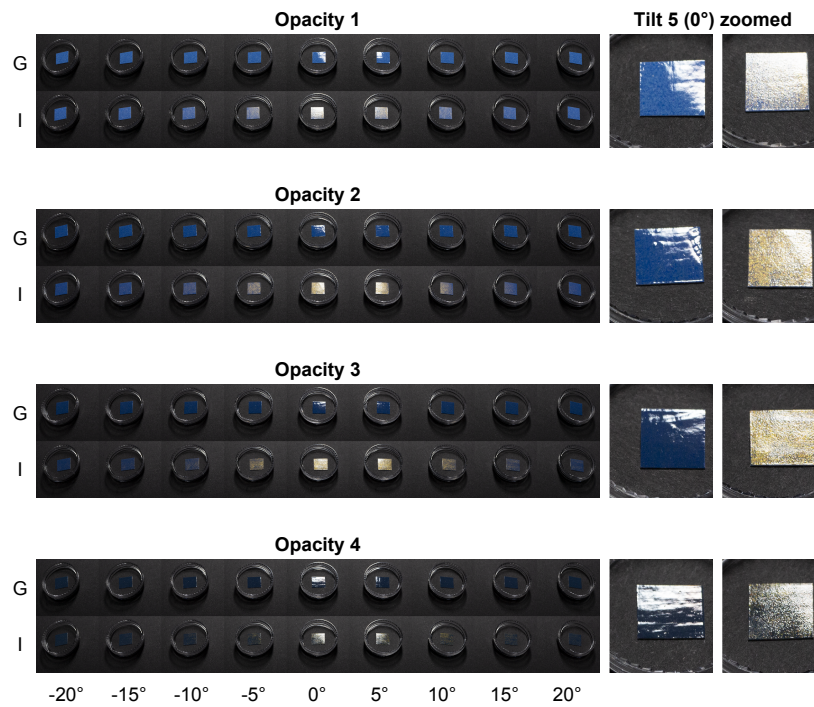


Fig. 3. Photographs that formed the frames of the videos of samples tilting from left to right. Each opacity level of the black layer (1 = 0%, 2 = 40%, 3 = 60%, 4 = 80%) of the printed squares was used to make a pair of glossy (top rows) and an iridescent samples (bottom rows). Zoomed-in images at tilt 5 are shown for glossy (left) and iridescent (right) samples for comparison.

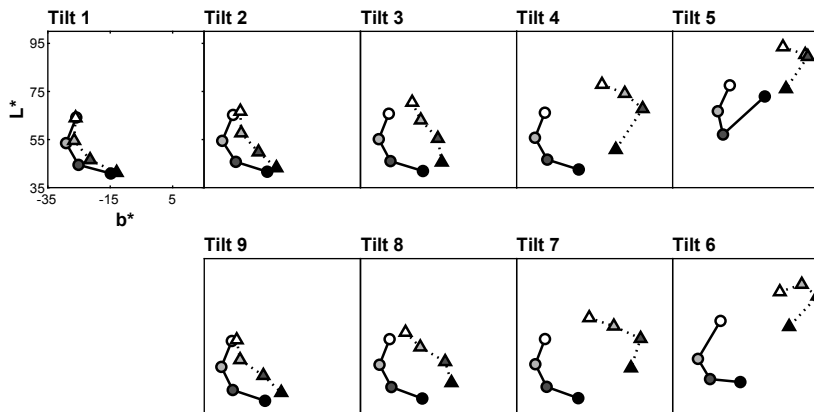


Fig. 4. Average $L^*a^*b^*$ of glossy (circles) and iridescent (triangles) samples at each tilt. Marker colors represent physical opacity level (from lightest to darkest).

pairs of video stimuli (see screenshot in Fig. 2(B)) presented on each side of the fixation cross. Each video occupied 50% of window height (at full screen) and had a width of 75% of stimulus height, which ensures the video aspect ratio (0.75:1) independent of screen or video size. Although this means that the stimulus size will be relative to screen height, we expect this variation to be relatively minimal as participants were constrained to using laptops only and completed the study in full screen mode.

Participants were asked to rate the similarity between each pair of samples on a scale of 1 (very different) to 7 (very similar), on the basis of “How similar are the two objects in terms of *how they’re made*” for Task 1 (Material Task), and “[...] in terms of their *average color*” for Task 2 (color Task). For the color Task, they were given an additional example of “average color” of a

black and white checkerboard (darker average gray) and a gray and white checkerboard (lighter average gray). The order in which participants completed each task was counterbalanced, with 16 completing the material task first and 18 the color task first.

E. MDS Analysis

The data analysis procedure was identical for both tasks. The median similarity ratings across the five repeats for each pair were converted to dissimilarities and analyzed with the MATLAB 2022b built-in *mdscale* function, which maps the Euclidean distance between the samples in a p -dimensional space to the dissimilarity values.

3. RESULTS

A. MDS Solutions

The MDS solutions contain the dimensions needed to explain how the samples vary perceptually, for each task, for each participant. The number of dimensions required can be determined by considering a scree plot of stress values for increasing numbers of dimensions. For all participants and all tasks, inclusion of the second dimension gave a marked reduction in stress; for 17/34 and 15/34 participants, on the color and material tasks, respectively, a third dimension was needed; and for only 4/34 and 5/34 on the two tasks did a fourth dimension produce further reduction. For ease of comparison across participants, and in-line with the dimensionality of color perception (though not necessarily material perception) and for ease of modeling, we chose to work with the three dimensional solution in each case.

We show examples of how perceptual judgments of the samples map on three dimensions in Fig. 5. The distance between a pair of samples in a space defined by the MDS dimensions represents the perceptual dissimilarity between them. Observer 54 seems to be employing the same strategy to make perceptual judgments in both tasks, with dimension 1 (D1) potentially corresponding to the differences in the samples produced by the opacity parameter in their construction since they seem to be sorted from level from 1 to 4. The sample mappings are noticeably different for Observers 21 and 32 as the glossy and iridescent samples cluster into two separate groups for both participants. Even though the results for their material tasks are similar, the mappings on their color tasks are different. These results illustrate that the location of the samples on each dimension is unique to each participant, suggesting that the dimensions may correspond to different percepts.

An intuitive analysis is to test whether the difference within the four samples of a surface coating—due to the differences in the opacity parameter, for example—is smaller than the difference between surface coatings. The mean coordinates of the four samples of each surface coating type (represented by pink triangles and circles for iridescent and glossy samples in

Fig. 5) represent the centers of where the two groups of samples cluster. We can see visually in Fig. 5 that in some cases the two local means are well separated, and in others cases the local means have near identical coordinates. In the latter case, the local means are also close to the global mean (the center of all samples, shown as the pink crosses), suggesting that the participant was not discriminating between surface coatings and was making similarity judgments based on other characteristics. The closer in Euclidean distance the samples are to their respective local mean, the more similar they are within material type. Here we are interested in the average distance to the local means compared to the average distance to the global mean and how this differs between the two tasks. A paired-sample t-test shows that the difference between local and global Euclidean distance was significantly higher for the material task ($M = 2.14$, $SD = 1.76$) compared to the color task ($M = 0.83$, $SD = 1.23$), $t(1, 33) = 3.97$, $p < 0.001$. To further understand these differences in participants' responses between tasks, we conducted additional modeling using image statistics of the stimuli they were shown.

4. SPECIFIC TEMPORALLY SELECTIVE AND NON-SELECTIVE MODELS

A key feature of our samples is the angle dependency of the statistics of the proximal image. However, it is likely that not every frame of the video contributes equally to a perceptual judgment. In this section, we present a series of models that compares the perceptual MDS solutions to predicted similarities between samples based on summary image statistics derived from videos. Comparisons between perceptual MDS results and the models were made using Procrustes transformation to find the best mapping of MDS values onto modeled image statistics. The image statistics here refer to averaged CIELAB values, similar to those presented in Fig. 4, which were derived by averaging the channel intensities, such that each tilt contributed equally to the average. Here, we parameterize the weighting of each tilt to allow an imbalance in contribution. In other words, the

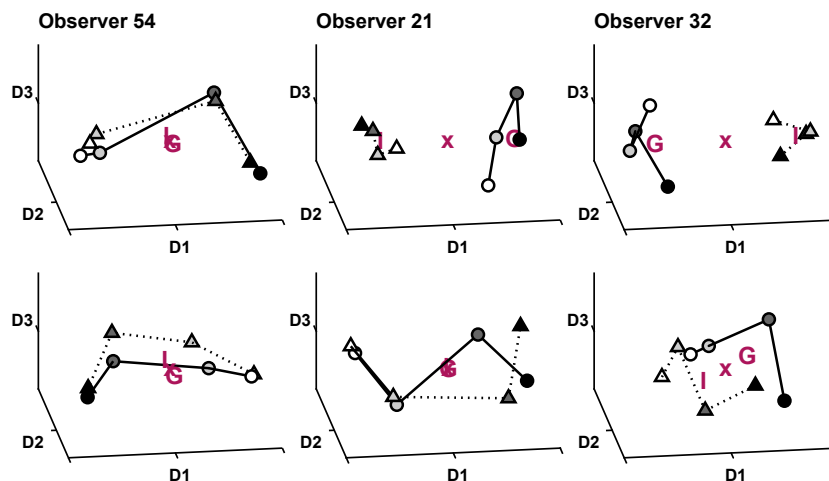


Fig. 5. Example MDS results: three-dimensional solutions are shown for three observers (54, 21, and 32) in the material task (top row) and the color task (bottom row). The axes represent dimensions (D), with the first dimension accounting for most of the variability in participants' similarity ratings between pairs of samples. Pink circles and triangles represent local means, which are the centers of the four samples for each surface coating, and the pink letters represent the local glossy (letter G), iridescent (letter I), and global means (letter x, located in the middle of all samples).

information from some tilts (such as the yellowness from tilt 5) may be minimized or maximized in its contribution to the summary statistic.

The coordinates of each stimulus in CIELAB space are meaningful indications of where a stimulus locates on the perceptual dimensions of lightness, red–greenness, and blue–yellowness. They can be considered as the MDS solutions for a hypothetical observer whose similarity ratings were determined entirely by the Euclidean distance between samples in this three-dimensional space. MDS analysis of the perceptual data—performed separately for each participant and task—also produces a triplet of Cartesian coordinates. Each coordinate shows where a stimulus locates on an undefined dimension, and the further apart a pair of stimuli are on a given dimension, the lower the subjective similarity ratings were for that pair. To measure model fit, we used Procrustes transformation to find the transformation matrix that would minimize the distance between the MDS coordinates and the CIELAB coordinates from the weight-adjusted image statistics. Through this process, we can quantify how well each participant’s MDS data align with the structure of similarities predicted for a hypothetical observer who uses only the calculated image statistics. The methods to vary tilt-weight can be split into (1) experimenter-selected and (2) data-driven approaches. Both approaches reveal systematic differences in the weightings of the tilts that generated the least residual error. We discuss each approach and the findings in detail below.

A. Experimenter-Selected Modeling

The simplest comparison of information across frames is to ask how well each single frame accounts for the MDS data. Mean model errors from single frames distribute differently for the two tasks [Fig. 6(A)]. Next, we created two hypothetical observers based on image statistics of the video stimuli. The two hypothetical observers differ in which tilts were excluded when deriving the average CIELAB values. The *non-selective hypothetical* observer considers all tilts equally ($L^*a^*b^*$ averaged across all tilts) and the *selective hypothetical observer* excludes all tilts apart from 1 and 9. The mean model error of each model across participants for both tasks is shown in Figs. 6(B) and 6(C) (models 1 and 2). A two-way ANOVA shows a significant interaction

between task and model [$F(1, 132) = 13.45, p < 0.001$], significant main effect of task [$F(1, 132) = 12.6, p < 0.001$], and no main effect of the model [$F(1, 132) = 0.7, p = 0.406$]. LSD post-hoc comparisons show that there were significant differences in model residual error for both tasks. For the material task, the non-selective model ($M = 0.38, SD = 0.14$) was a better fit than the selective model ($M = 0.55, SD = 0.30$). The opposite was true for the average color task, such that the selective model ($M = 0.39, SD = 0.14$) was a better fit compared to the non-selective model ($M = 0.29, SD = 0.22$).

B. Data-Driven Modeling: Principal Component Analysis

1. Introduction to PCA Steps

The experimenter-selected modeling suggests that participants flexibly use information from different views of the samples according to the judgment they are making. However, these models pre-determine the weightings (0 or 1) given to information from each tilt, and through this type of analysis it is not possible to exhaustively test all possible combinations of information from different views. We verify the findings from above through data-driven modeling.

We used principal component analysis (PCA) to reduce the dimensionality of the dataset, preserving only the principal components (PCs) of variation across tilts that explain the highest proportion of variance between samples (defined by their $L^*a^*b^*$ coordinates). For each PCA analysis, we classified the nine tilts as the variables and the eight stimuli as observations. PCA allows us to meaningfully reduce the variables (currently a nine-dimensional space) into minimal number of PCs required to account for most of the variance across observations (samples). PCA reduction was done separately for each $L^*a^*b^*$ coordinate, and for each coordinate we retain each PC that explains at least approximately 10% of the variation.

2. Interpretation of $L^*a^*b^*$ Principal Components

For L^* and b^* , we retain the first and second components, explaining 85.0% and 13.0% of the variance for L^* and 89.5% and 9.6% for b^* , respectively. The coefficients of the retained PCs across the nine tilts are shown in the top of Fig. 7. The

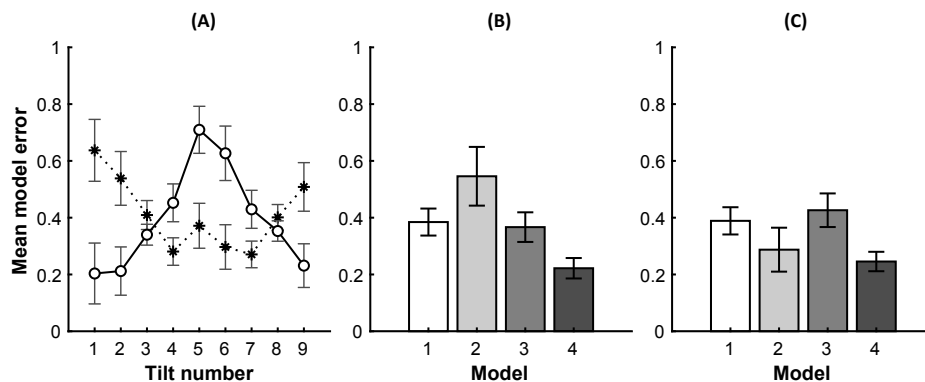


Fig. 6. Mean model errors following Procrustes transformations of individual MDS solutions. (A) Single-tilt model: fits based on $L^*a^*b^*$ values of each individual tilt for material (asterisks) and color (circles) tasks, where error bars represent standard errors. (B) and (C) Fits to four models: 1. Non-selective. 2. Selective. 3. PC1-weighted $L^*a^*b^*$. 4. PC1 and optimal PC2 scaled for (B) material and (C) color tasks, where error bars show 95% confidence intervals.

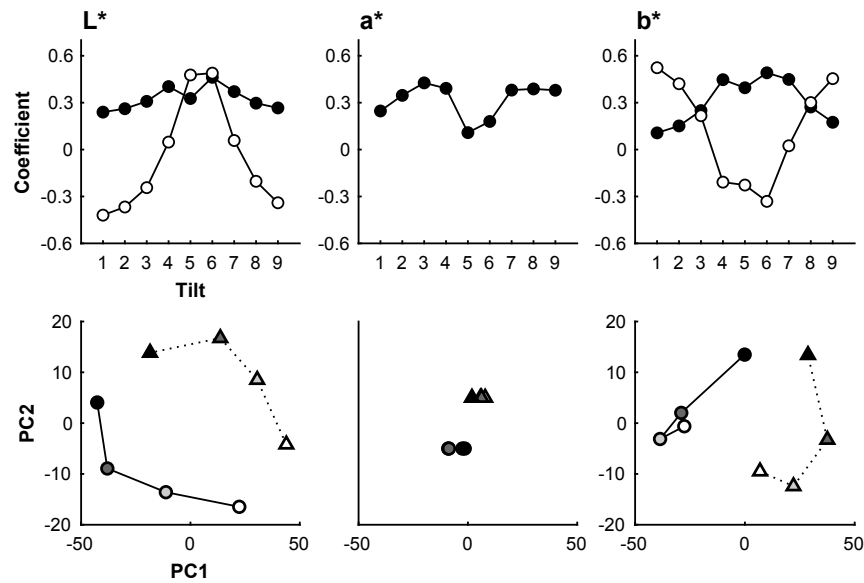


Fig. 7. Top row, the coefficients of principal components (PCs) that explain 10% or more of the variability between samples across tilts, for each of L^* , a^* , b^* coordinates. The components (black, PC1; white, PC2) weight each tilt differently. Bottom row, the PCs weighted average $L^*a^*b^*$ across nine tilts for each of the eight samples (glossy, circles; iridescent, triangles; opacity parameter, gray level), showing how the samples plot in the principle component space. The y axis of a^* (middle column) is arbitrary as only one component is retained, and data points for glossy and iridescent samples are displaced on this axis for clarity.

coefficients of the PCs indicate the required weighting of each tilt to best capture the differences between samples.

Figure 7 also shows the way the eight samples plot in the principle component space for each color coordinate ($L^*a^*b^*$). According to the number of retained components, the space is two-dimensional for L^* and b^* and one-dimensional for a^* . For L^* , PC1 with some small contribution from PC2 captures the variation in sample statistics that is produced by the opacity manipulation, with relatively uniform contribution from all tilts. PC2 almost perfectly separates samples according to surface coating, with some overlap between the darkest glossy samples and the lightest iridescent samples. PC2 heavily down-weights the contribution from the extreme tilts and up-weights the middle tilts at which the two surface coatings produce the most difference in L^* .

For b^* , PC1 separates the samples by material coating, drawing most heavily on the middle tilts where iridescent samples appear yellower (more golden) than the glossy samples. PC2 partially separates the samples according to the opacity manipulation, especially for the darkest samples compared to the lighter ones, and this information is drawn from the extreme tilts where neither the gold interference color nor the specular reflection is present.

3. Mapping MDS Coordinates to PC-Weighted Coordinates

One instructive analysis would be to test whether the PC1-weighted chromatic statistics are good predictors of participants' MDS solutions. This is similar to taking the “non-selective” average (our simplest summary of the chromatic data), but here the average is weighted to maximize the difference between samples in the way their CIELAB coordinates vary with tilt. We report this as model 3 in Fig. 6.

To test the hypothesis that observers base their similarity judgments on image statistics collected flexibly across tilts differently for each task, we further used optimization techniques to scale the relative contributions of PC1 and PC2 to best fit the data. We first extracted the $L^*a^*b^*$ values weighted by only the PC1s of L^* , a^* , and b^* , and we used Procrustes transformation to fit each participant's MDS solution to these PC1-weighted values. We then optimized the relative contribution of PC2 in a weighted sum (PC1 + weighted PC2) to minimize the residual error following Procrustes transformation. In the first Procrustes transformation, we included reflection, but once the MDS solution had been correctly oriented to PC1, the transformation in the optimization was limited not to include further reflection and the weights constrained to be between 0.001 and 10.

A higher optimal weighting of PC2 suggests that relatively more PC2 contribution was required in addition to PC1 to generate a better fit. Optimal weightings, particularly for the L^* values, tended to cluster as either requiring a contribution of from PC2 or not. We, therefore, tested the proportion of participants' who require a non-zero weighting (i.e., scaling over 0.01) in each task for L^* and b^* .

A Chi-square test of independence confirms that there were significant differences between the two tasks in the proportion of participants requiring PC2 for L^* , such that for the material task 20 out of 34 of participants required PC2 contribution and for the color task this was only eight participants, [$X^2(1, 34) = 8.74, p = 0.003$]. The opposite was found for b^* , such that for the material task 14 participants required PC2 contribution compared to only three participants for the color task [$X^2(1, 34) = 9.49, p = 0.002$].

5. DISCUSSION

The focus of this study is the dynamic sampling of visual signals from objects as they are manipulated within the lighting environment and with respect to viewing angle. A strong conjecture is that perception of a material is informed by the way it behaves when manipulated, rather than from single static images (see parallel considerations of dynamic information by [20] who consider perception of liquids and their viscosity from image motion). Visual signals from an object may be used to derive information about the identity of an object or about its nature, and different information may be available under different viewing conditions. These considerations led us to investigate the effect of task instructions on similarity judgments between samples that varied in the way they were constructed. Two sets of physical samples were made, each starting from the same base layer of printed mixtures of blue and black (whose contribution was manipulated with an opacity parameter). These base layers were coated with either two layers of varnish (the “glossy” samples) or two layers of varnish with an interleaved layer of interference paint (the “iridescent” samples). Similarity judgments, elicited in response to two separate questions, were analyzed with MDS to reveal the perceptual dimensions that separated the samples. The organization of the stimulus samples in perceptual space depended on the question. Task instructions have previously been shown to have strong effects on color matching, depending on whether participants are asked about their color or material percepts [21]. An analysis of the Euclidean distances between the average location of samples grouped by surface coating confirms that, across participants, there is a reliable systematic difference between the color task and the material task.

In this study, samples were photographed from a fixed viewpoint as they were positioned at specific tilts with respect to the lighting environment. Photographs were combined into videos presented to participants via an online experiment. While not permitting the participants to manipulate the samples themselves, this mode of presentation did provide precise control over the image statistics available to participants from the images of the samples. In the second part of the paper, we used extensive modeling to understand how MDS solutions from subjective similarity ratings might map onto models derived from objective image statistics. These models functioned as “hypothetical observers” that weighted information from the tilts differently.

The modeling approach was to derive a summary of the sample “color” by extracting from the images composite L^* , a^* , and b^* coordinates that could be used to locate the sample in a perceptual (color) space and to look for similarities between representations of the samples in this space and the structure of the MDS solutions. The single-tilt models assumed, for example, that the summary extracted from the videos was simply the average L^* , a^* , and b^* coordinates of pixels corresponding to the sample in one single tilt, ignoring all other tilts. In contrast, the non-selective model equally weighted L^* , a^* , and b^* coordinates from all tilts, while the selective model used only the most extreme tilts (1 and 9). In such models, the researcher chooses the rule for combining information from frames of the video, and particularly for the selective model the choice is informed

by knowledge of the stimulus properties and task. We, therefore, additionally sought a data-driven analysis to verify the results.

One option would be to optimize the weights on statistics from each frame of the video to produce composite L^* , a^* , and b^* coordinates that best fit the MDS solution for each participant. However, the information contained in successive frames is highly correlated, and such an optimization is highly under-constrained. We, therefore, used PCA to reduce the dimensionality of the stimulus set. Importantly, the PCA was performed across tilts (and not, for example, to reduce the dimensionality of the color-space description), and separate PCA solutions were found for each of the three color dimensions, L^* , a^* , and b^* . The extracted PCs, therefore, characterize the way in which the eight material samples are distributed in a nine-dimensional space where each dimension represents the relevant color coordinate (L^* , a^* , or b^*) for one of the nine tilts. The coefficients for each successive PC indicate the relative weightings of information from each tilt required to best capture the residual variation between samples. Beyond this study, such analyses may provide useful characterizations of the signatures of particular material types as they are manipulated in a lighting environment.

Most of the variation between samples in this study was in the L^* and b^* coordinates. For each of these coordinates, the first PC explained a high proportion of the variance between samples in the way they changed with tilt, but in both cases there was additional variance ($\sim 10\%$) explained by the second PC. Models that optimized the relative weighting of PC1 and PC2 in deriving composite L^* , a^* , and b^* coordinates produced reliably better fits to the MDS solutions than a model that used only PC1. Importantly, the numbers of participants whose best fitting models included PC2 was significantly different for the color task and the material task, for both L^* and b^* coordinates. This result provides direct evidence that participants utilize information from a sequence of views of an object differently depending on the type of judgment they are making about the object.

For L^* the color task results were, for most participants, well explained by PC1. Interestingly, this PC has a small suppression of tilt “5”, which is the tilt most affected by specular reflection. For the material task, an additional contribution from PC2 was required for significantly more participants. PC2 down-weights the extreme tilts and up-weights the middle tilts (“5” and “6”), capturing the increased lightness associated with the iridescent samples at these angles. For b^* , PC2 contributed strongly to the model fits for the majority of participants for both tasks, but for significantly more participants for the color task. For this coordinate, PC2 down-weights the middle tilts and up-weights the extreme tilts, suggesting that when judging “average color” participants suppress the golden hue that appears at selective tilts from the interference paint. Our average color task purposefully gave no instruction as to how to treat iridescent color percepts. In this study, it appears that most participants focused on the color of the base layer of the material samples in their average color similarity judgments. Further empirical studies would be required to investigate this in detail, but it draws attention to what participants understand by terms such as “surface” and “surface color,” and whether iridescent percepts are categorically different from absorption pigment-based color percepts (such

that, for example, they contribute minimally to the “average color” of the sample).

6. CONCLUSION

We conclude that our findings align with flexible weighting of information from videos of tilting material samples, depending on task instructions. This opens exciting avenues for future work on iridescence—particularly on how information is extracted from the proximal image with real objects that participants can manipulate themselves. The colors produced by iridescence, and the colors accessible to the observer, depend on the geometry of the object, observer, and light source. Therefore, dynamic weighting of information might also drive how we physically handle and perceive iridescent objects.

Acknowledgment. We thank Tsvetomira Dumbalska for providing code for the perceptual brightness matching task and Allie Hexley for code on sRGB transformations. We would also like to thank two anonymous reviewers for providing helpful feedback on an earlier version of the manuscript.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

- H. E. Smithson, “Sensory, computational and cognitive components of human colour constancy,” *Philos. Trans. R. Soc. London B* **360**, 1329–1346 (2005).
- G. J. Ward, “Measuring and modeling anisotropic reflection,” in *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques* (1992), pp. 265–272.
- A. C. Chadwick and R. Kentridge, “The perception of gloss: a review,” *Vis. Res.* **109**, 221–235 (2015).
- J. Kim, P. J. Marlow, and B. L. Anderson, “The dark side of gloss,” *Nat. Neurosci.* **15**, 1590–1595 (2012).
- J. Kim, P. Marlow, and B. L. Anderson, “The perception of gloss depends on highlight congruence with surface shading,” *J. Vis.* **11**(9):4 (2011).
- J. Berzhanskaya, G. Swaminathan, J. Beck, and E. Mingolla, “Remote effects of highlights on gloss perception,” *Perception* **34**, 565–575 (2005).
- R. W. Fleming, R. O. Dror, and E. H. Adelson, “Real-world illumination and the perception of surface reflectance properties,” *J. Vis.* **3**(5):3 347–368 (2003).
- R. S. Hunter, “Methods of determining gloss,” NBS Research paper RP 958 (1937).
- M. G. Meadows, M. W. Butler, N. I. Morehouse, L. A. Taylor, M. B. Toomey, K. J. McGraw, and R. L. Rutowski, “Iridescence: views from many angles,” *J. R. Soc. Interface* **6**, S107–S113 (2009).
- L. Ng, L. Ospina-Rozo, J. E. Garcia, A. G. Dyer, and D. Stuart-Fox, “Iridescence untwined: honey bees can separate hue variations in space and time,” *Behav. Ecol.* **33**, 884–891 (2022).
- D. Stuart-Fox, L. Ospina-Rozo, L. Ng, and A. M. Franklin, “The paradox of iridescent signals,” *Trends Ecol. Evol.* **36**, 187–195 (2021).
- K. Kjærnsmo, H. M. Whitney, N. E. Scott-Samuel, J. R. Hall, H. Knowles, L. Talas, and I. C. Cuthill, “Iridescence as camouflage,” *Curr. Biol.* **30**, 551–555 (2020).
- K. Kjærnsmo, J. R. Hall, C. Doyle, N. Khuzayim, I. C. Cuthill, N. E. Scott-Samuel, and H. M. Whitney, “Iridescence impairs object recognition in bumblebees,” *Sci. Rep.* **8**, 8095 (2018).
- S. A. Fabricant, A. Exnerová, D. Ježová, and P. Štys, “Scared by shiny? The value of iridescence in aposematic signalling of the hibiscus harlequin bug,” *Anim. Behav.* **90**, 315–325 (2014).
- O. C. Sacken, “Iridescent colours,” *Nature* **47**, 102 (1892).
- L. Belcour and P. Barla, “A practical extension to microfacet theory for the modeling of varying iridescence,” *ACM Trans. Graph.* **36**, 1–14 (2017).
- S. Werner, Z. Velinov, W. Jakob, and M. B. Hullin, “Scratch iridescence: wave-optical rendering of diffractive surface structure,” *ACM Trans. Graph.* **36**, 1–14 (2017).
- R. J. Lee and H. E. Smithson, “Motion of glossy objects does not promote separation of lighting and surface colour,” *R. Soc. Open Sci.* **4**, 171290 (2017).
- K. Xiao, C. Fu, D. Karatzas, and S. Wuerger, “Visual gamma correction for LCD displays,” *Displays* **32**, 17–23 (2011).
- T. Kawabe, K. Maruya, R. W. Fleming, and S. Nishida, “Seeing liquids from visual motion,” *Vis. Res.* **109**, 125–138 (2015).
- L. Arend and A. Reeves, “Simultaneous color constancy,” *J. Opt. Soc. Am. A* **3**, 1743–1751 (1986).