

# Lightness Constancy in Surface Visualization

Danielle Albers Szafir, *Student Member, IEEE*, Alper Sarikaya, *Student Member, IEEE*,  
and Michael Gleicher, *Member, IEEE*

**Abstract**—Color is a common channel for displaying data in surface visualization, but is affected by the shadows and shading used to convey surface depth and shape. Understanding encoded data in the context of surface structure is critical for effective analysis in a variety of domains, such as in molecular biology. In the physical world, lightness constancy allows people to accurately perceive shadowed colors; however, its effectiveness in complex synthetic environments such as surface visualizations is not well understood. We report a series of crowdsourced and laboratory studies that confirm the existence of lightness constancy effects for molecular surface visualizations using ambient occlusion. We provide empirical evidence of how common visualization design decisions can impact viewers’ abilities to accurately identify encoded surface colors. These findings suggest that lightness constancy aids in understanding color encodings in surface visualization and reveal a correlation between visualization techniques that improve color interpretation in shadow and those that enhance perceptions of surface depth. These results collectively suggest that understanding constancy in practice can inform effective visualization design.

**Index Terms**—I.3.7 Three-Dimensional Graphics and Realism - Color, shading, shadowing, and texture; Lightness Constancy; Molecular Visualization; Surface Visualization; Visual Perception.



## 1 INTRODUCTION

VISUALIZATIONS often use color to encode scalar field data on three-dimensional surfaces. Color is a powerful cue for surface visualization as it can intuitively represent data within the context of a surface. Visualizing data in context is especially critical for surfaces such as molecules, where functional and structural features provide a meaningful scaffold for understanding charge, binding sites, protein-protein interfaces, and other data.

However, shading models used to render surfaces directly impact color encodings: shadows and shading manipulate color to convey depth, resulting in a conflict between representations of shape and data. Surface features, like pockets and loops, often hold interesting areas for exploration, but tend to be the most deeply shadowed. Misinterpreting color encodings in these regions adversely impacts a visualization’s effectiveness, but removing surface shading impairs perceptions of surface depth and shape. By understanding how visualization design influences how accurately viewers can read colors from shaded regions, we can design surface visualizations that better support both shape and data comprehension.

Reading color-encoded scalar data from the surface of a molecule requires matching colors against a legend or key. Because the *image color* of data on the surface depends on shadows and shading, the *apparent color* of the data may not be the same as the unshaded *key color* in the legend. In this work, we explore how visualization design affects viewers’ abilities to match shadowed image colors to the corresponding unshaded color in a key. In the real world, this task would be enabled by lightness constancy—the ability of the visual system to use various visual cues to disentangle color and shadow. Lightness constancy is well studied in perceptual science, and a number of theories and models exist explaining how different visual cues contribute

to this ability. However, these models focus on explaining constancy in real world or simple computer-generated scenes (see Kingdom [1] for examples). They provide little guidance for how mechanisms for interpreting surface colors may be affected by the stylized or simplified techniques used to render interactive complex surfaces in visualization. Lightness constancy is also sensitive to a variety of visual factors: in studies, simply moving from the real world to a virtual image has significantly impaired constancy [2]. Techniques commonly used in visualization, such as ambient occlusion lighting [3], may remove many visual cues that theoretical work indicates are used for lightness constancy, such as lighting direction [4].

In this work, we derive inspiration from prior research on lightness constancy to understand how visualization design can support the accurate interpretation of encoded data in surface visualization. In a series of experiments, we measure color-matching performance for molecular surface visualizations rendered using ambient occlusion. We confirm that viewers can read color encodings in shadow with some accuracy for the simplified rendering methods often used in visualization, and that how the surface is visualized directly influences the strength of this ability. Specifically, the visualization techniques used to render a surface can significantly improve or inhibit viewers’ ability to correctly interpret shadowed colors. Our results point to a correlation between techniques that enhance depth perceptions and improved performance in interpreting shadowed colors. These results can guide designers in creating surface visualizations that more accurately depict shadowed data. They also illustrate trade-offs for designing surface visualizations using color. Given the complex and unfamiliar structures of molecular surfaces, we anticipate that these results could be applied to visualizing surfaces in other domains.

**Contributions:** We present six experiments that explore how visualization design influences viewers’ abilities to accurately interpret color-encoded data for molecular surface visualizations using ambient occlusion. Our results confirm that viewers can

• D.A. Szafir, A. Sarikaya, and M. Gleicher are with the Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, 53706.

E-mail: [dalbers, sarikaya, gleicher]@cs.wisc.edu

Manuscript received April 19, 2005; revised September 17, 2014.

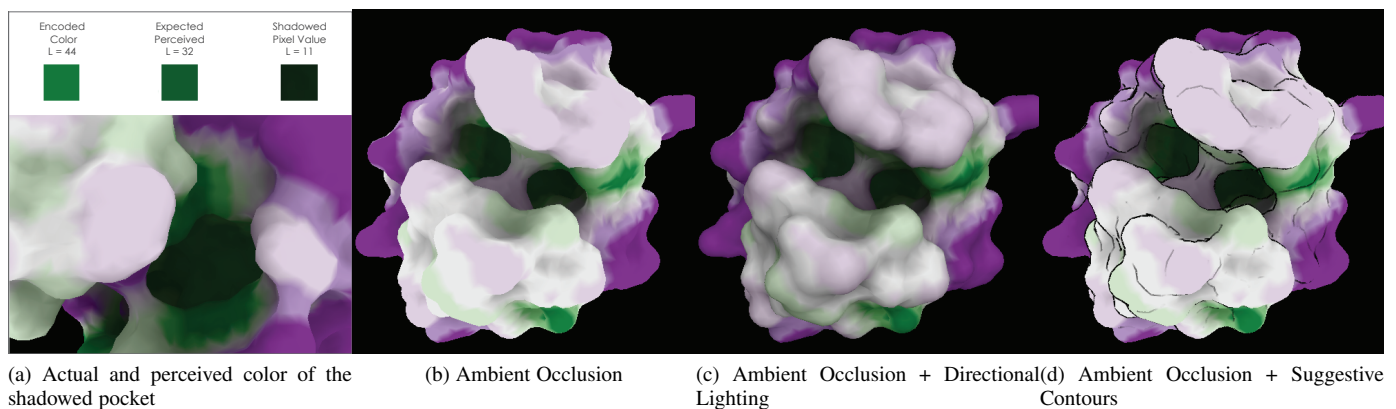


Fig. 1: Our findings, exemplified by hydrophobicity data in the shadowed regions above, show that visualization design significantly impacts viewers’ abilities to read data encoded on a surface. (a, b) Ambient occlusion surfaces support viewers in reading shadowed data, which is improved by (c) directional shading. Conversely, (d) stylized shape cues may hinder this ability.

resolve color on visualized surfaces with some degree of accuracy despite approximations in the illumination model. We present further evidence that different surface visualization techniques, when used in conjunction with ambient occlusion, can affect this accuracy. We specifically find that adding directional lighting and stereo viewing improve viewers’ abilities to read shadowed colors, whereas suggestive contours hinder performance. These results provide preliminary evidence of a correlation between techniques that enhance depth perceptions and accurate interpretations of shadowed data. A summary of results is presented in Figure 1.

## 1.1 Background

Visualization allows analysts to explore data in the context of a surface by mapping visual representations of data on a rendering of the surface itself. The resulting image combines a number of different visual factors to support data analysis. The visual system has several different constancy mechanisms that account for variation in visual factors. Color constancy in particular allows us to resolve colors under different lighting conditions and has three principal elements [5], [6]: lightness constancy, hue constancy, and saturation constancy. All three components can be used to encode data along a surface [7]. Supporting their constancy allows visualization designers to use these channels effectively. In this work, we focus on lightness constancy as it allows the visual system to account for luminance variations underlying the shadows and shading that convey surface structure.

Perceptual psychology has established models to explain how lightness constancy functions account for changes in illumination in the real world [8], [9], [10], [11], [12], [13]. Existing theories hypothesize that properties such as contrast ratios between light and shadow [14], [15], shadow intensity [6], lighting intensity [16], lighting direction [4], object colors and reflectance [15], [17] and spatial cues [18], [19], [20], [21] may all contribute to the brain’s ability to disentangle an object’s color from the lighting used to illuminate it. For example, the visual system may identify a luminance value in a scene as an “anchoring point,” such as the lightest or average luminance value, and adjust perceptions of all residual colors accordingly [1]. The brain may also adapt to lightness differences in smaller spatial regions of a scene and adjust perceptions to maximize these local contrasts, essentially increasing the perceived dynamic range of the scene [16].

This prior work focuses on perceptual mechanisms, quantifying constancy as a function of low-level visual features under highly controlled conditions for both artificial and naturally-occurring scenes. Studies of constancy in digital environments generally use simple stimuli, such as two-dimensional images (e.g. flat square planes or collections of randomly sized and colored rectangles that form “Mondrians”) or checkerboards overlaid on simple three-dimensional shapes (e.g. cubes [22], [23], [24] or creased rectangular planes [25]). It is unclear how these findings translate to surface visualizations, where complex surface structures are often illuminated using approximated and stylized lighting models (e.g. ambient occlusion).

Surface visualizations commonly use ambient occlusion [3] to approximate global illumination. Several properties of this illumination model may inhibit or even remove visual cues that are hypothesized to facilitate lightness constancy. For example, many theories suggest that lightness constancy relies largely on backcomputing color changes in a scene based on overall lighting and reflectance properties [5]. This idea of “estimating the illuminant” depends on the existence of measurable lighting contributions, including direction and relative intensity. However, ambient occlusion synthesizes equal light from all directions—the resulting illumination is directionless and of uniform intensity. This might inhibit lightness constancy and reduce viewers’ abilities to interpret shadowed colors.

Additionally, most of what we know of constancy is based on identifying grey-scale colors under differing levels of illumination viewed under controlled conditions (see Kingdom [26] for a survey of experiments considering color). In surface visualization, color ramps are not grey-scale, but often communicate data values through variation in hue, lightness, and saturation, and viewed under a variety of conditions.

The measures generated by perceptual models are not focused on providing feedback for designers; they model mechanisms of the visual system operating over specific visual features. These limitations make it difficult to apply these models to visualization: it is unclear how they inform whether different designs will sufficiently increase the effectiveness of a visualization. We do not aim to model lightness constancy for surface visualization, but instead we consider the effects of lightness constancy as a measure of visualization effectiveness. We seek to understand

how common visualization techniques influence how accurately viewers interpret surface colors in shadow.

## 2 RELATED WORK

In this work, we focus on viewers' abilities to distinguish colors applied to molecular surfaces. While other encodings, such as glyphs [27], [28] or textures [29], are sometimes used to convey surface data, color is one of the most common encodings for displaying data on these surfaces.

Molecules can be visualized in many different ways: as atomic representations [30], stylized moieties [31], or functional surfaces [32], [33]. In this work, we focus on solvent-excluded surface models, which are commonly used in conjunction with color encodings to display molecular data in a spatial context (see [29], [34], [35] for examples). Data such as charge, binding affinity, and machine learning results are projected across these surfaces to increase the functional and spatial understanding of the dataset. Color is often used to visualize this data in popular systems like VMD [36], Pymol [34], and BioBlender [37]. While the structural features of a solvent-excluded surface present visually interesting aspects of the surface for such investigations, surface shadows (which use grayscale color to convey depth) may be problematic when using color to encode data. Recent efforts have explored alternative techniques for visualizing heavily shadowed regions of surfaces, such as opacity reduction [38] and volume segmentation [39]; however, these methods focus only on deep pockets and reduce the visual quality of the overall surface to emphasize these pockets. Interactive techniques for exploring surfaces are also problematic for shadowed data—viewers may incorrectly interpret values obscured by shadow, making it difficult to accurately identify interesting regions to explore.

Although shape and shadow complicate color encodings, they are important for communicating spatial properties of a surface (Fig. 2). Recent research in volume rendering has explored how different shading models impact viewers' depth perceptions in visualization. Although they focus on volume visualization, the studies provide useful general insight into surface perception. For example, Lindemann and Ropinski [40] evaluated seven lighting models to derive design suggestions for effective depth-based rendering. More recently, Grosset *et al.* [41] demonstrated how subtle changes to a depth cue (depth of field) can significantly influence perceptions of a volume. Such research empirically evaluates common design decisions to confirm how different design choices impact perceptions of structural features in a 3D visualization, but do not consider how these choices influence perceptions of color or other visual encodings.

Ambient occlusion is commonly used to convey depth in molecular surface visualizations in both research [27], [42] and production tools [34], [37]. Ambient occlusion approximates shadows on a surface by assuming a constant light emitted from all directions, measuring the percentage of possible lighting directions visible from a given surface point, and attenuating the surface color at that point accordingly [3]. This provides a pre-computed approximation of shadow that conveys depth comparable to directional lighting models [43]. Yet, it often fails to convey subtle shape variations and is therefore often supplemented with other shape and depth cueing techniques such as diffuse illumination [42], contours [27], [42], and haloing [42] in molecular surface visualization.

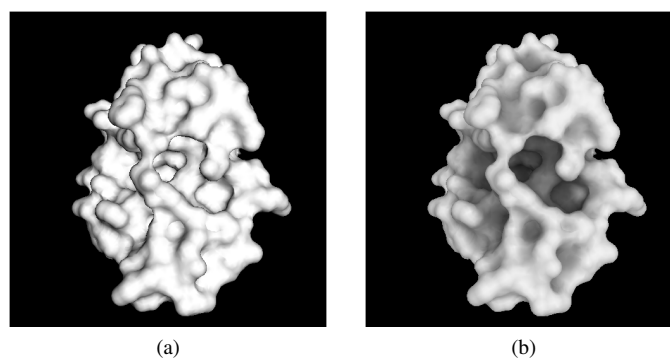


Fig. 2: Depth perception of a surface using (a) local illumination can be greatly enhanced by (b) adding ambient occlusion shading, which emphasizes the shape of structural features such as pockets.

In this work, we explore lightness constancy for molecular visualizations that leverage color encodings. We specifically focus on how visualization design influences constancy to support accurate performance on a color matching task. While the ensuing studies measure viewer performance on solvent-excluded molecular surfaces, we anticipate that the findings of this study, summarized in Figure 3, are extensible to visualizing of more general classes of surfaces.

## 3 MOTIVATION AND OVERVIEW

The way we represent data directly influences how accurately viewers interpret visualized data. For example, the rendering methods used to create a volume visualization impact perceptions of surface depth [40]. In visualization design, there is often a trade-off between how closely a visualization reflects the real world and how efficiently it can be rendered. We may choose to make this trade-off for many reasons, such as supporting interactivity, rendering on devices with different computational resources, or emphasizing certain properties of an object. By understanding how different visualization design choices influence how accurately visualized data is perceived, designers can systematically reason about these trade-offs to create visualizations that support specific tasks.

In this paper, we explore how different design techniques for visualizing surface data influence how accurately viewers interpret shadowed data. We focus on visualizations rendered with ambient occlusion as it is commonly used to convey surface depth and shape without the computational overhead of more complex shadow rendering techniques. Ambient occlusion computes the shading values for a surface once, and those values remain constant regardless of viewpoint. It exchanges many aspects of real world lighting captured by more complex global illumination models (e.g. interreflection in radiosity, lighting direction from cast shadows), which must be recomputed whenever the light or surface are moved, for computational tractability. This trade-off can improve performance for interactive visualizations.

Ambient occlusion supports perceptions of surface depth using shading to simulate shadows. When a data value is encoded as color on an ambient occlusion surface, shading makes the pixel value of the image color on the surface darker than the original encoded color. For shadowed objects in the real world, lightness constancy enables viewers to disentangle colors from

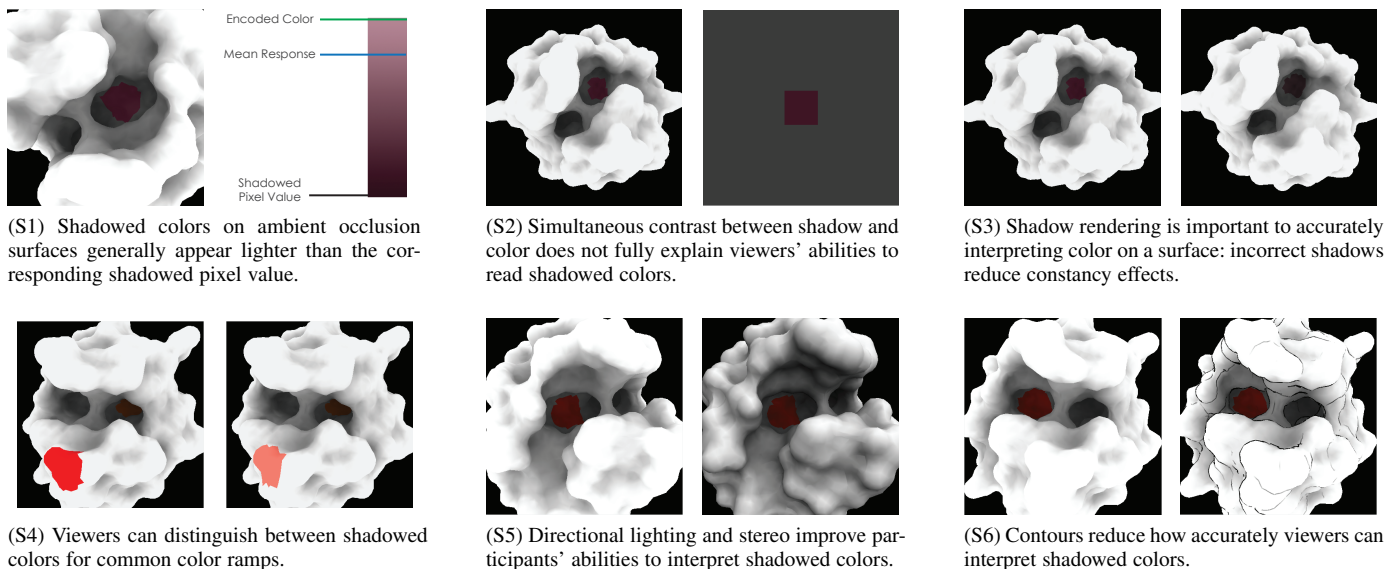


Fig. 3: We explore how visualization design influences viewers' abilities to accurately read shadowed colors in surface visualization. We first verify that viewers can interpret shadowed colors on ambient occlusion surfaces and that surface shading and structure supports this ability. We then explore how different surface visualization techniques might improve or impair performance. These results can help inform the design of effective surface visualizations.

shadow. Many properties of more complex global illumination models are known to contribute to these constancy effects, such as directional lighting, cast shadows, or interreflectance of light along the surface. Given the prevalence of ambient occlusion in surface visualization, we want to understand if this lighting can support accurate color interpretation in shadow and what aspects of visualization design influence these effects for surfaces rendered with ambient occlusion.

The studies presented here represent first steps in this exploration. The goals of this work are to (a) establish that viewers can accurately read shadowed surface colors from ambient occlusion surfaces, and (b) reason about the trade-offs in this ability for common surface visualization designs. We address these goals by answering six specific research questions (addressed in studies **S1** through **S6**). Like Anderson and Winawer's approach to analyzing the causes of constancy [44], we consider how different *design layers*—visualization design decisions that influence the presentation of a surface—may influence performance on a color matching task. We begin by verifying that ambient occlusion surfaces can support lightness constancy effects (**S1**) and that these effects are a function of visualization design rather than contrast between data and shadow (**S2**). This verification suggests that viewers can interpret shadowed data with some accuracy and that the design of a visualized surface may influence this ability.

Once these properties are verified, we then explore how the shadow computation itself influences performance (**S3**). Specifically, we show that some simplifications of shadow models may hinder the interpretation of shadowed colors.

The ability to correctly estimate the luminance of shadowed surface colors is imperfect: viewers still make more errors in interpreting shadowed surface colors than unshadowed colors. However, we find that performance in color matching tasks is still higher when luminance cues are included in the color mapping than for hue and saturation alone in common practice color ramps (**S4**).

Surface visualizations often combine other visualization techniques with ambient occlusion to enhance perceptions of the depth and shape of a surface. In order to understand the trade-offs involved in these design decisions, we compare viewers' abilities to interpret shadowed colors when adding directional lighting, stereo cues, and suggestive contours. Our results suggest that designs correlated with depth cueing (directional lighting and stereo) allow viewers to more accurately identify shadowed surface colors (**S5**), while stylized contours, which enhance shape percepts at the expense of shadow percepts [45], reduce performance (**S6**).

These studies, discussed in detail in the ensuing sections, collectively suggest that visualization design influences how well viewers' can read surface colors, and that there is a correlation between designs that support perceptions of surface depth and those that effectively convey shadowed data. A summary of results is provided in Table 1. This work serves as first steps in mapping a design space for understanding lightness constancy in visualization.

## 4 GENERAL METHODOLOGY

We evaluated the relationship between color matching performance and surface visualization design through a series of color matching experiments. Each experiment required participants to match a data value encoded as color on a surface to its original ramp color in a provided key. All experiments followed the same general procedure. Any variation from this methodology is discussed in detail for each experiment. The manipulated design component was treated as a between-participants factor in all but one experiment (§7.2).

Participants were first screened for color vision deficiencies using digital renderings of Ishihara plates [46]. Only participants who passed this screening were allowed to proceed, and a post-hoc questionnaire was used to further verify normal color vision. Participants were instructed that they would see a series of images

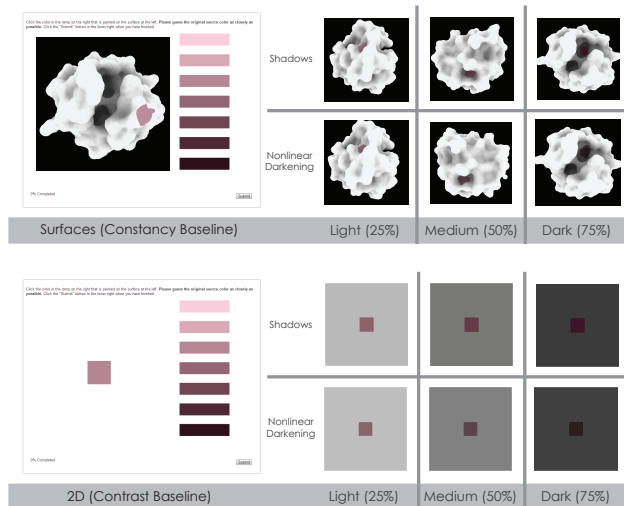


Fig. 4: We mapped colored patches to three levels of shadow. Colored patches applied to molecular surfaces rendered using ambient occlusion gauged performance for molecular surfaces (top), whereas 2D squares (bottom) measured effects due to contrast with the surrounding shadow.

with colored patches placed under different levels of shadow and would be asked to match the original color displayed in the image to a provided key. A pair of example problems were provided to help illustrate the task. Participants were then shown 32  $500 \times 500$  pixel stimulus images presented in a random order with the corresponding seven-step color ramp key immediately to the right of the stimulus (Fig. 4). Participants recorded the color in each ramp they felt best corresponded to the shadowed patch by clicking on a color in the key and clicking a “Submit” button to move to the next stimulus. To mitigate adaptation effects, participants saw a gray screen for three seconds between respective stimuli (duration was selected through pretesting). Participants were given unlimited time for each response.

#### 4.1 Stimulus Generation

Unless otherwise stated, stimuli consisted of static images of solvent-excluded surfaces rendered as a white surface on a black background. Surfaces were shaded using ambient occlusion plus a 10% constant ambient term. Shadows were generated by reducing surface luminance using ambient occlusion computed using the methods described in Landis [3] by attenuating the luminance component of the surface assuming a white light and a gamma of 2.2 [47].

Surfaces were derived from four different proteins from the Protein Data Bank [48] (PDB IDs: 1BBH (bacterial), 1B7V (bacterial, Fig. 1, 4, 9, and 3), 1DB4 (human, Fig. 2), 3CLN (mammalian, Fig. 11) and generated via MSMS [49] with each surface visualized entirely within the image. Since the experiments were conducted in the browser, all images were prerendered with sRGB embedded color profiles.

A single colored patch was mapped to a unique position on each surface for each shadow level. Patches were of roughly equivalent size on each surface—some variation was caused by the curvature of the surfaces—and never directly bordered the black background. For each experiment, patches were generated for three levels of shadow: light ( $25\% \pm 2\%$  shadow), medium

( $50\% \pm 2\%$  shadow), and dark ( $75\% \pm 2\%$  shadow), with all shadow levels measured after the 10% ambient lighting was applied. Patches were placed in shaded regions where no part of the region was lighter than the assigned shading level and at least 94% of the region was within the assigned shading level.

Each participant saw colors from two seven-step color ramps. To control the number of stimuli viewed by each participant, we selected three colors from each ramp as test colors to be displayed in shadow. Tested colors were selected such that each color was at least one just noticeable difference (JND) apart even in the darkest shadow condition. Throughout this work, we use the JND measure defined using crowdsourced metrics in [50] to help account for anticipated display variability.

Except in the stereo pilot (§7.2), each participant saw 32 stimuli total: six stimuli at each of the three shadow levels (three colors per ramp) and 14 stimuli with patches placed in an unshadowed position (one for each level of each ramp) for validation and to prevent biased responses from the reduced set of colors in the shadow conditions. The primary manipulation was a between-participants factor for all but the stereo study. Images were selected randomly from each of the four surface models, and the stimuli were presented in a random order to minimize adaptation to a given color or shadow level. The use of validation stimuli with “obvious” correct answers (in this case, the exact pixel match to the surface patch) is commonly used to gauge honest responses in crowdsourced studies, where participants sometimes “click-through” questions using random answers to complete the study as quickly as possible [51]. Participants responding two or more ramp units away from the correct answer on multiple validation stimuli were excluded from our analyses.

#### 4.2 Participant Selection

Participants were selected from two separate pools: in-person (20 participants total) and crowdsourced using Amazon’s Mechanical Turk (322 participants total). Mechanical Turk is known to be a generally reliable participant pool for graphical perception studies [51], [52] and also allows us to measure performance for viewers under a spectrum of real-world viewing conditions. This approach may introduce variability in viewing conditions and devices, which prevents us from making precise claims about visual perception. We hypothesize that this variability may be beneficial for measuring factors significantly influencing performance under realistic conditions, but leave this verification to future work. To ensure the quality of our results, we followed known best practices for ensuring honest responses [53], only recruited participants with at least a 95% overall “approval” rating, and used explicit validation questions. We also tracked both worker identification number and IP address across all experiments to ensure that each participant completed only one experiment.

We recruited 16 in-person participants (10 female, 6 male) between the ages of 21 and 31 ( $\mu = 25.75$ ,  $\sigma = 2.47$ ) to address **S1** and **S2** under controlled conditions. We then recreated these experiments using crowdsourced participants on Amazon’s Mechanical Turk. We found consistent results between in-person and crowdsourced participants, confirming that crowdsourcing is a sufficiently reliable method for recruiting participants for our color matching task. We addressed **S1** through **S6** using a cumulative total of 322 crowdsourced participants (174 male, 147 female, 1 declined to report) between the ages of 18 and 65 ( $\mu = 31.25$ ,  $\sigma = 9.66$ ).

Certain visualization conditions are not amenable to crowdsourcing, such as stereo viewing (S5), which requires specialized displays. For our stereo experiment, we also required participants to have prior stereo experience due to the nuances of proper stereo viewing with the available technologies. We recruited 4 participants with prior experience with stereo displays for pilot study S5. We present preliminary findings from this study, but consider it a pilot as we were only able to recruit a limited number of participants, all with some familiarity with our task, due to our qualification restrictions.

We analyzed our data for each experiment (except S1) for main and first-order interaction effects at the level of shadow  $\times$  color ramp  $\times$  primary independent variables using ANCOVAs (Analyses of Covariance). Question order was treated as a covariate to account for interparticipant variation from repeated measures across conditions. In all cases, results from each participant pool were analyzed independently, and an equal number of participants were considered for each condition within each experiment. If exclusions caused an imbalance between conditions within an experiment, participants were excluded at random until both conditions were balanced. Across all studies, only data from participants who reported normal or corrected-to-normal vision and no color vision deficiencies was considered. We observed no significant performance effects due to age.

## 5 VERIFYING LIGHTNESS CONSTANCY FOR SURFACES

### 5.1 Do we see constancy effects for ambient occlusion surfaces? (S1)

Before exploring color matching performance as a function of design, our first experiment aims to verify that participants can match image colors to a key when the colors are darkened by ambient occlusion shadows. If these visualizations support lightness constancy, we would anticipate that viewers would match colors closer to the original, unshadowed color than to the darkened pixel color in the image.

**Methods:** The procedure and stimuli for this experiment are outlined in §4. We carefully engineered two luminance-varying ramps such that, for each tested color, both the correct key color and the pixel value of the shadowed image color could be mapped to within one crowdsourced JND of a ramp color. These ramps allowed us to verify that the participants were able to employ lightness constancy in order to disambiguate between the pixel value of the shaded patch and its corresponding ramp value. Ramp luminance was varied in the CIELAB color space from  $L^* = 9$  to  $L^* = 87$ , with each step separated by 13 units and  $L^* = 35$ ,  $L^* = 61$ , and  $L^* = 87$  used as test colors (Fig. 4). We centered the ramps around blue and red such that all colors remained within the monitor gamut and consecutive colors were sufficiently different. Each participant saw 16 stimuli from each ramp, resulting in 32 total responses per participant.

Participants were drawn from two pools: 8 in-person participants to measure constancy effects under controlled conditions and 17 crowdsourced participants from Mechanical Turk to measure effects under the diverse array of conditions experienced in visualization applications. Two crowdsourced participants were excluded from the analysis for poor performance on validation stimuli, resulting in 15 participants total for our analysis. Participants completed the in-person study using an Asus G51J Series

Shadowed Pixel Color versus Participant Responses

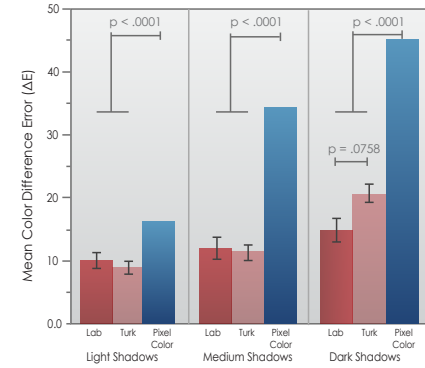


Fig. 5: Mean difference between the correct patch color and participant responses in S1. Both in-lab and crowdsourced participants mapped shadowed colors significantly closer overall to the original key color than to the shadowed pixel value in the image. All error bars encode standard error.

Laptop with an Nvidia GeForce GTX 260M graphics card in full screen using Google Chrome. Room lights were dimmed to control ambient illumination.

**Results:** Performance was measured as the difference between the correct key color and the color reported by participants. We use this metric rather than absolute correctness because constancy is an approximate phenomena—even in real scenes, constancy mechanisms cannot always exactly compute the correct color [54]—“right” or “wrong” measures do not adequately capture performance. Figure 5 summarizes our results.

A repeated measures Multivariate Analysis of Variance (MANOVA) on each set of participants revealed evidence of significant constancy effects. There was a significant difference between colors reported by participants and the actual shadowed surface colors ( $F_{in-person}(1,45) = 19.7818, p_{in-person} < .0001$ ;  $F_{Turk}(1,93) = 29.9981, p_{Turk} < .0001$ ). Participants mapped shadowed patches to significantly lighter colors than the surface pixel color. This result was consistent across all shadow levels ( $F_{in-person}(2,45) = 130.4005, p_{in-person} < .0001$ ;  $F_{Turk}(2,93) = 141.2638, p_{Turk} < .0001$ ). We did not find a significant difference between response errors at the three tested shadow levels in the in-person experiment ( $F(2,45) = 1.3754, p = .2566$ ) and between the light and medium shadows in the crowdsourced experiment ( $F(1,45) = 1.4694, p = .2264$ ). This lack of difference also indicates constancy effects—participants mapped these patches to roughly equivalent colors despite significant changes in shadow darkening. Overall, the performance of the crowdsourced participants was consistent with in-person participants. However, crowdsourced participants performed slightly worse on the darkest shadow conditions.

These results collectively suggest that participants can account for the effect of shadows on surface colors. Participants matched surface colors to colors significantly lighter than the pixel value of the shadowed image color, and darker shadows did not always influence the apparent color (Fig. 5). The consistency of these results across both in-person and crowdsourced conditions point to the robustness of this phenomena across viewing conditions and suggests its importance for visualization design. However, our results also suggest there may be room for improvement: there was still significant error in matching surface colors to the original

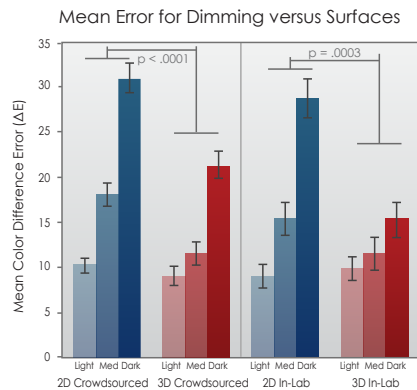


Fig. 6: Viewers identified colors more accurately on surfaces than on dimmed two-dimensional planes (S2), suggesting that visualized surface structure plays a role in identifying shadowed colors.

colors, and this error might be improved by different visualization techniques.

## 5.2 Do viewers use structural information to interpret surface colors? (S2)

Our results from S1 demonstrate that participants correctly identify shadowed image colors as lighter than the corresponding pixel color. However, these results do not confirm the cues the visual system uses to interpret these colors. It is possible that the darkness of the shadow surrounding a patch allows participants to interpret image colors rather than structural information from the visualization. Simultaneous color contrast between a stimulus and the surrounding shadows accounts for some aspects of lightness constancy in the real world [14], [16], although it is insufficient to explain all constancy effects in three-dimensional surfaces [21]. The visual system may normalize contrast for local windows around a patch at comparable depth plane [1], [55]. For visualizations, contrast between the patch color and the surrounding shadows may cause the local color patch to appear lighter than its actual pixel value—the dark shadow makes the patch appear lighter by contrast. In this experiment, we wanted to verify that contrast does not account for all of the effects reported in Section 5.1. If contrast sufficiently explains the S1 results, how we visualize a surface will not significantly influence perceptions of shadowed colors.

To test if the observed constancy effects were due to simultaneous contrast, we compared color matching on visualized surfaces to two-dimensional “shadowed” patches. If the effects from the surface color matching task (§5.1) are due to contrast, we would expect no significant difference between color perception for 2D shadows and 3D surfaces. This would imply that performance depends on the darkness of the surrounding shadow rather than the overall visualization design.

**Methods:** 2D stimuli consisted of 100 pixel-wide colored square patches centered in a 500 pixel-wide white square. Patch size was comparable to the 3D surface patches. Both the patch and background square were dimmed to the tested shadow level to mimic the local lighting on a molecular surface—the white background was dimmed to the grey of the surface shadow to indicate the lighting shift, and the colored square was dimmed to

the color of the shadowed patch (see Fig. 4, bottom). Participants were instructed that both patch and background were in shadow and that they were to identify the original, unshadowed color of the center patch. In order to simplify task instructions, 2D stimuli were not placed on a black background. As colored patches were generally placed far from the background in the 3D condition, we do not anticipate any confounds from this decision: contrast effects in constancy are speculated to operate over local windows within a visual scene [55].

Colors, shadow levels, general procedure, and stimulus distribution mirrored the 3D condition described in the previous section, including 14 unshadowed 2D validation stimuli (§5.1). Data was again collected from two participant pools: 8 in-person participants and 18 crowdsourced participants. Three crowdsourced participants were excluded for poor performance on validation stimuli, resulting in 15 participants for analysis. Both sets of participants were run simultaneously with those discussed in §5.1, with dimension treated as a between-participants factor.

**Results:** We ran a three-way ANCOVA (dimension, color ramp, and shadow level) on the difference between the original color and response color for the 2D planes and the 3D data from S1 for each participant pool. Participants matched colors on 3D surfaces significantly more accurately than on equally darkened 2D patches (Fig. 6,  $F_{in-person}(2,262) = 13.3018$ ,  $p_{in-person} = .0003$ ;  $F_{Turk}(1,532) = 23.9261$ ,  $p_{Turk} < .0001$ ). This accuracy varied significantly across shadow level ( $F_{in-person}(1,262) = 19.5687$ ,  $p_{in-person} < .0001$ ;  $F_{Turk}(2,532) = 74.6342$ ,  $p_{Turk} < .0001$ ), but not across color ramp ( $F_{lab}(1,262) = .2160$ ,  $p_{lab} = .2160$ ;  $F_{Turk}(1,532) = 0.1031$ ,  $p_{Turk} = .7483$ ).

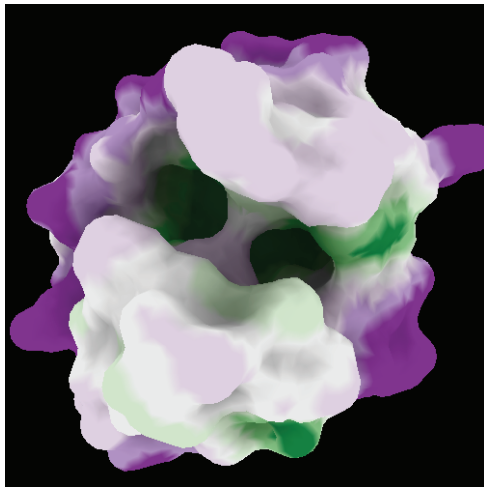
These results suggest that the effects measured in S1 (§5.1) are not entirely explained by simultaneous contrast: visualized surface structure accounts for a significant proportion of the reported color matching performance.

## 5.3 Does approximating shadow darkening affect color matching performance? (S3)

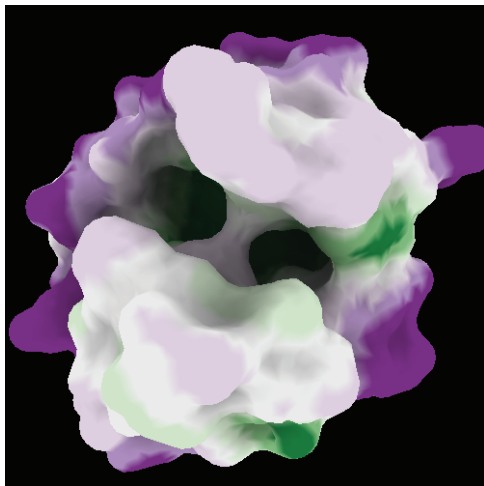
Correct ambient occlusion shadows render shading by attenuating the amount of light emitted by the display linearly with the amount of shadow. However, a designer may also approximate shading by darkening the surface using an image post-process, which may not account for non-linearities introduced by device gamma. The difference is subtle: because image processing occurs in device-dependent RGB, the shadows would be differently affected by gamma correction (Fig. 7). While the magnitudes of these changes are small, they distort the gradients of the resulting shadows, which may contain important information for constancy [56].

While modern visualization systems generally apply shadows correctly, the effects of subtle differences in shadow application provide evidence of the connection between perception theory and visualization practice. Namely, distorting these gradients may reduce performance on our color matching task. Given the subtle visual difference between the two conditions, a performance difference on our tested task would imply that perceptions of shadows and surface structure influence the apparent color of surface data.

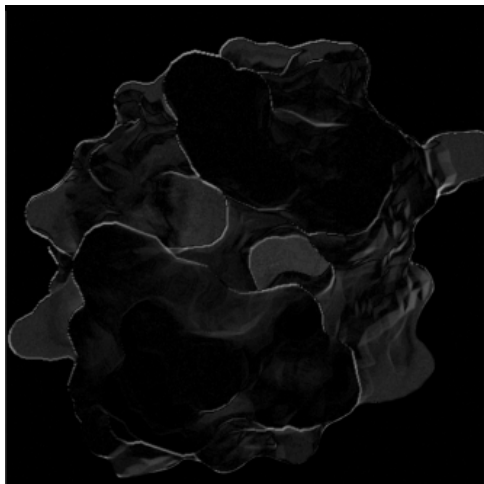
**Methods:** We replicated the previous experiments (§5.1 and §5.2) using stimuli that applied ambient occlusion attenuation to each channel of device-dependent RGB color ( $\gamma = 2.2$ ). The luminance of all corresponding correct and image-processing darkened colors were within one  $L^*$  JND measured under crowdsourced



(a) Hydrophobicity data with correct shadows



(b) Hydrophobicity data with image-processing shadows



(c) Color differences mapped to greyscale

Fig. 7: Differences in CIELAB  $\Delta E$  between correct and image-processing shadows for the surface visualized in Figure 1b. Color difference is encoded using linear greyscale, with black representing areas of no difference.

Mean Error for Dimension and Darkening by Shadow

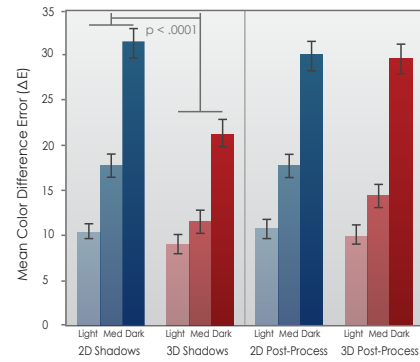


Fig. 8: No significant improvements were seen between dimmed planes and surfaces darkened using non-gamma corrected image-processing methods. This suggests that constancy mechanisms leverage shadow information when processing surface colors (S3) and small changes to those shadows can damage their effects.

conditions [50]. Dimension (2D versus 3D) was treated as a between-participants factor. The procedure otherwise mirrored that described in §4.

The study was run simultaneously with the crowdsourced studies discussed in Sections 5.1 and 5.2. We collected data from 34 participants on Mechanical Turk. Three participants were excluded from the 2D condition and one from 3D surfaces condition for performance on validation stimuli, resulting in 15 participants per condition for analysis.

**Results:** To address S3, we compared participant responses across all four crowdsourced conditions (dimmed 2D patches from Section 5.2, 3D surfaces from Section 5.1, image-processing darkened 2D patches, and image-processing darkened 3D surfaces). A four-way ANCOVA (dimension, shadow level, darkening type, and color ramp) was used to analyze participant responses. Participants identified colors significantly more accurately on 3D surfaces than 2D planes ( $F(1, 1061) = 18.4228, p < .0001$ ). Shadow level significantly influenced performance ( $F(2, 1061) = 160.0194, p < .0001$ ), but we found no significant effect of color ramp ( $F(1, 1061) = 1.2035, p = .2729$ ), and only a marginal main effect for darkening type ( $F(1, 1061) = 3.8011, p = .0515$ ). We also found a significant interaction effect of dimension and darkening type ( $F(2, 1061) = 5.5637, p = .0185$ , Fig. 8). A Tukey’s Test of Honest Significant Difference (HSD) found no significant difference between performance in both 2D conditions and the image-processing darkened surfaces, but revealed that correct shadowed surfaces outperformed all other conditions (at  $\alpha = .05$ ).

These results suggest that precise shadow information facilitates data interpretation along a surface. While we found no performance differences for 2D planes, incorrect shading significantly decreased color matching performance on molecular surfaces despite the subtlety of the visual differences between the displayed images (Fig. 9). Incorrect shadows may communicate surface structure, but may not be sufficient to support constancy mechanisms in interpreting encoded data—we found no evidence that incorrect shadows provide any performance gains beyond what 2D shadows provide. These findings also suggest that visualization design decisions that manipulate surface shading may influence viewers’ abilities to correctly interpret surface data in visualization.



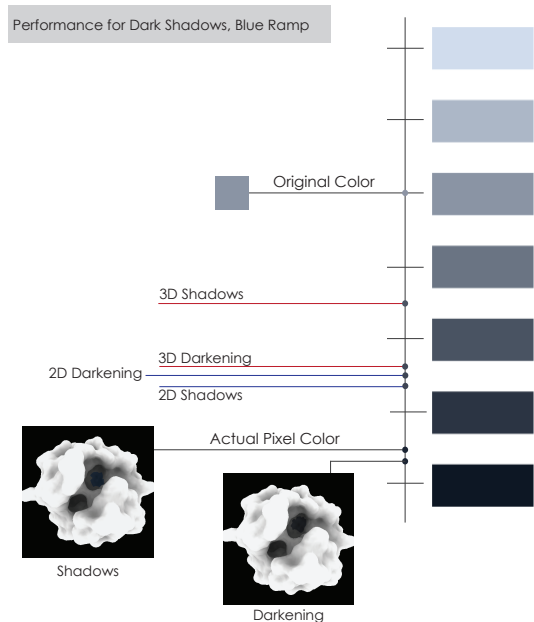


Fig. 9: Molecular visualizations using standard ambient occlusion demonstrated the best color identification performance. The lack of a significant difference between 2D dimmed patches and 3D surfaces using image-processing darkening suggests that the visual system actively uses shadow information to extract shadowed surface colors.

## 6 DO CONSTANCY EFFECTS PRESERVE PERFORMANCE GAINS FROM LUMINANCE VARIATION? (S4)

In practice, well-designed color ramps integrate luminance variation with other color cues (see Moreland [57] for a discussion of color ramp design considerations). Luminance is a strong cue for identifying colors in visualization; however, shadows compress luminance variation in surface visualization. Lightness constancy effects must sufficiently preserve luminance variations in shaded regions for luminance-varying ramps to retain their performance benefits over isoluminant ramps. While §5 provides empirical evidence that some of these cues can be preserved, it does so using carefully engineered ramps that strictly use luminance cues in order to gauge subtle effects. In this experiment, we compared three commonly-used ramps that integrate luminance variation with their isoluminant equivalents to determine if luminance cues are beneficial for color ramps used to encode data on ambient occlusion surfaces.

**Methods:** Stimuli were constructed as discussed in §4, with any colors outside sRGB gamut [47] clamped via chroma reduction. Color ramps consisted of a purple-white-green (PWG) and red-yellow-blue (RYB) diverging ramp from ColorBrewer [58] and a rainbow ramp (Ra) akin to that used in PyMol [34]. As opposed to the red and blue luminance-varying ramps from the previous experiments, these ramps represent common practice color choices for surface visualization—the ColorBrewer ramps represent general good practice for data encoding, while the rainbow ramp provides an example of extreme hue variation that is used in practice but suffers from several known limitations [59], [60].

The isoluminant variations of these ramps were computed by setting the CIELAB  $L^*$  values of the ramps to  $L^* = 65$ , near the

Mean Error for Color Ramp and Luminance Conditions

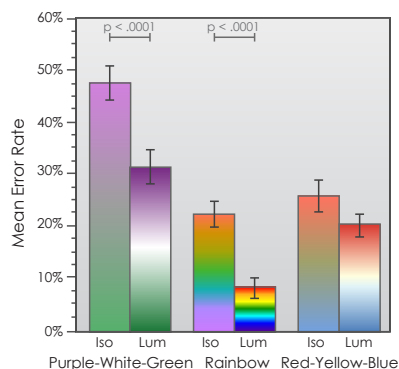


Fig. 10: Luminance-varying ramps supported significantly better performance than their isoluminant equivalents, suggesting that lightness constancy helps viewers interpret data encoded with well-designed color ramps (S4).

average luminance of all ramp colors. While the distance between colors is reduced in the isoluminant ramps, this reduction is only in lightness. As a result, we can gauge if lightness constancy effects are sufficient to preserve the luminance cues in the original ramps. To control for this compression, we verified that consecutive color steps within each ramp differed by at least three times our benchmark JND [50]. A pilot identified three sample values from each ramp as potentially misidentified colors to be used as test colors: dark purple, light purple, and mid-green for PWG; orange, red, and mid-blue for RYB; and orange, cyan, and purple for Ra.

The experimental procedure was identical to that described in §4, with luminance treated as a between-participants factor. Each participant saw colors from one isoluminant ramp and a different luminance-varying ramp (PWG with isoluminant RYB, Ra with isoluminant PWG, and RYB with isoluminant Ra). Data was collected from 92 participants on Mechanical Turk. One participant was excluded from each of the PWG/isoluminant RYB and the RYB/isoluminant Ra condition for performance on the unshadowed validation stimuli, resulting in 30 participants per condition for analysis.

**Results:** As the difference between consecutive colors varied between ramps, we used the number of ramp units between the original and response color as our primary measure and absolute correctness as a secondary measure (Fig. 10). Since a direct mapping exists between the isoluminant and luminance varying ramps, our primary measure uniformly quantifies performance differences between ramps despite the fact that this difference is not necessarily uniform in color space.

We analyzed the primary measure using a three-way ANCOVA (luminance variance, shadow, and ramp). Overall, luminance-varying ramps significantly outperformed isoluminant ramps ( $F(1,1664) = 23.5519, p < .0001$ ). Performance varied significantly across shadow level ( $F(2,1664) = 53.8479, p < .0001$ ), and color ramp ( $F(2,1664) = 47.0705, p < .0001$ ). Both PWG and Ra ramps significantly outperformed their isoluminant equivalents ( $F_{PWG}(1,1664) = 17.3171, p_{PWG} < .0001$  and  $F_{Rainbow}(1,1664) = 9.1601, p_{Rainbow} = .0025$ ). While RYB outperformed isoluminant RYB on average, the difference was not significant ( $F(1,1664) = 1.4980, p = .2211$ ).

Our results suggest that performance gains from luminance variation in well-designed ramps are preserved for ambient oc-

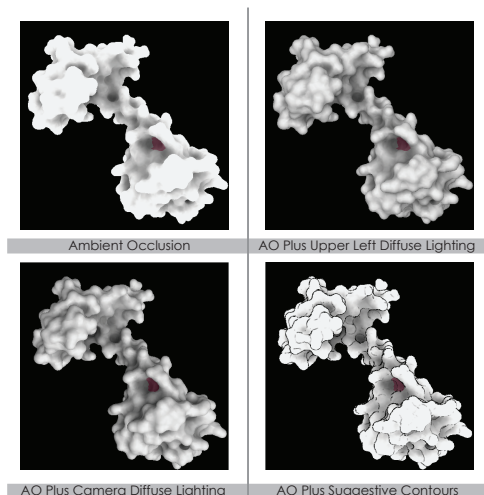


Fig. 11: We measured how design influences viewers’ abilities to interpret shadowed colors using three additions to ambient occlusion from the molecular visualization literature: diffuse local lighting (both sourced at the camera and in the upper left) and stereo viewing to enhance depth, and suggestive contours to enhance shape.

clusion surfaces. This performance gain implies that lightness constancy matters in practice for surface visualization—luminance is a strong color cue; if we can create designs that better support lightness constancy, we can improve visualization effectiveness.

## 7 AFFECTS OF DEPTH AND SHAPE CUES

Our results indicate that lightness constancy may enhance the apparent color of surface data on molecular surfaces rendered with ambient occlusion. Our first three studies (S1–S3, §5) suggest that the spatial cues created on a surface by ambient occlusion shading support participants in accurately matching surface colors to a key. Ambient occlusion provides both shape and depth cues. Both may enhance perceptions of surface structure, but previous work shows that these factors may influence performance on our color matching task in different ways [54].

Molecular surface visualizations often supplement ambient occlusion with other rendering techniques that provide additional structural cues. Adding depth cues to ambient occlusion surfaces can improve depth perceptions [61], but is unclear if these added cues will significantly increase color identification performance [56]. Strict shape cues can damage the abilities of viewers to infer shape from shading for a surface rendering [45], which our results from S3 suggest may, in turn, reduce performance. In this section, we compare three techniques commonly used in conjunction with ambient occlusion to enhance depth and shape cues in molecular visualization: directional lighting [62], stereo viewing [34], and suggestive contours [42] to test how depth and shape cueing affect participants’ abilities to interpret surface colors (Fig. 11).

### 7.1 How do added depth cues from directional lighting affect color matching performance? (S5)

Local directional lighting is commonly used to supplement ambient occlusion in molecular surface visualization. This provides an estimable lighting direction and increased depth cueing, both

Mean Error for Directional Lighting Conditions

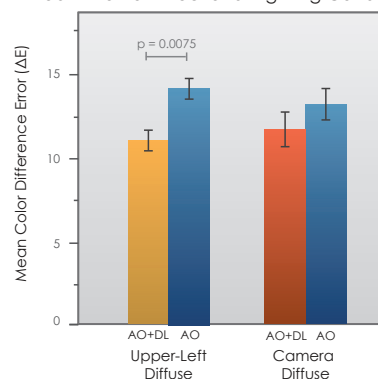


Fig. 12: Adding directional lighting to ambient occlusion significantly improved viewers’ ability to identify colors in shadow; however, this improvement appears to be correlated with the amount of depth cueing (S5) provided by the lighting direction.

of which may enhance constancy effects over ambient occlusion alone [4], [19]. We anticipate that adding local directional lighting may improve color matching performance for surface visualization. However, this improvement might depend on the position of the light source, which influences how significantly the added lighting improves depth perceptions [63]. We explored the relationship between local directional lighting and constancy effects using light sourced at two positions: from the upper left where it is strongly correlated with depth perceptions and from the camera where it provides substantially less depth cueing.

**Methods:** Stimulus images were generated as described in §4.1, with colors drawn from the seven-step red luminance-varying ramp (§5.1) and purple-white-green diverging ramp (§6). We generated two stimuli collections, each consisting of one set of visualizations using ambient occlusion alone and a corresponding set using ambient occlusion plus a directional light. The directional light was positioned at the camera in the first collection and to the upper left of the molecule in the second. Surface patches were slightly displaced from the previous experiments and between each collection to account for variation in shadow introduced by the directional shading, but patch placement was identical within each collection.

Supplementing ambient occlusion with directional lighting could cause surfaces to be significantly lighter than with ambient occlusion alone. This would potentially confound our experiment: superior performance of directional lighting may be caused by lighter shadows rather than structural cues introduced by directional lighting. To avoid this confound, we implemented directional light as diffuse shading and computed surface shading using the equation  $A = 0.5 \cdot ao + 0.5 \cdot ao \cdot (\hat{l} \cdot \hat{n})$ , where  $ao$  is the ambient occlusion value,  $\hat{l}$  is the unit vector from the center of the molecule towards the light source and  $\hat{n}$  is the unit normal. This model bounds the surface shading such that the directional plus ambient occlusion surface shading is never lighter than the raw ambient occlusion values.

The experimental procedure was otherwise identical to that described in §4. Each participant saw 32 stimuli from exactly one lighting condition. Data was collected from 108 participants on Mechanical Turk. Two participants were excluded from the upper left lighting condition and three from each condition in the

camera-sourced lighting collection for performance on validation stimuli, resulting in 25 participants per condition.

**Results:** We ran a three-way ANCOVA (shading model, shadow, and ramp) on the differences between the original color and participant responses for each stimulus collection (Fig. 12). We found significant main effects of shadow ( $F_{upperleft}(1,801) = 76.1552$ ,  $p_{upperleft} < .0001$ ;  $F_{camera}(1,676) = 59.7345$ ,  $p_{camera} < .0001$ ) and ramp ( $F_{upperleft}(1,801) = 10.4646$ ,  $p_{upperleft} = .0013$ ;  $F_{camera}(1,676) = 10.4646$ ,  $p = .0013$ ). Surface visualizations with directional lighting supported significantly better color judgments than ambient occlusion surfaces alone when the light was positioned to the upper left of the molecule ( $F_{upperleft}(1,801) = 7.1918$ ,  $p_{upperleft} = .0075$ ). Adding camera-sourced directional lighting also improved perceptions on average, but the difference was not significant ( $F(1,676) = 1.4369$ ,  $p = .2311$ ).

These results indicate that enhanced depth cues may be more important to interpreting color-coded data along a surface than lighting direction: lighting sourced to the upper left of a surface provides both better depth cueing [63] and greater color matching performance than lighting sourced at the camera. Artificially bounding shading in the directional lighting conditions makes shadows generally darker than in the baseline ambient occlusion condition. Therefore, we anticipate that effects seen in this experiment will likely increase in practice without this bound.

## 7.2 How do added depth cues from stereo viewing affect color matching performance? (S5)

The previous experiment provides evidence that enhancing depth cues may increase how accurately participants interpret color-coded information on a surface. Stereo viewing also increases depth cueing and is supported by many commercial molecular visualization packages [34]. Experiments in psychology have found that binocular stereo cues may improve constancy effects [64], [65]; however, these studies are based on simple stimuli and do not use commercial stereo devices. We anticipate that stereo depth cues might improve participants' abilities to correctly identify color on surfaces, but it is unclear if other tradeoffs made by stereo viewing, such as reduced color fidelity, will outweigh these effects in practice.

**Methods:** We tested stereo viewing using a within-subjects pilot study on a passive stereo display (Zalman Trimon ZM-M220W). Two stimulus sets were generated: one consisting of row interlaced stereo visualizations and another with the corresponding monocular images. All stimuli assumed a uniform interpupillary distance. Participants were initially screened for stereo blindness and then shown a sample stereo molecule and asked to adjust their position until the object appeared as a continuous, three-dimensional shape. The procedure was otherwise identical to that described in §5.1. Participants wore polarized stereo glasses through the entirety of the study in both the stereo and monocular conditions.

We compared stereo and monocular viewing in an in-person pilot with four participants. Because stereo viewing relies on proper display technologies and is highly sensitive to a number of parameters, we required participants to have prior experience with stereo viewing. This constraint limited the number of participants we were able to recruit. To help account for the limited number of participants in this study, we doubled the number of shadowed stimuli seen by each participant (each participant saw each shadow condition twice per tested color) and treated stereo

Mean Error for Binocular Depth Cueing

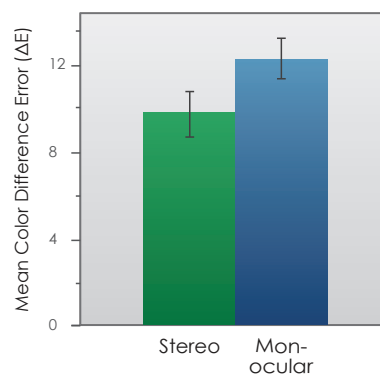


Fig. 13: Surface color perceptions improved when molecular surfaces were supplemented with binocular depth cues (S5).

Mean Error for Suggestive Contours

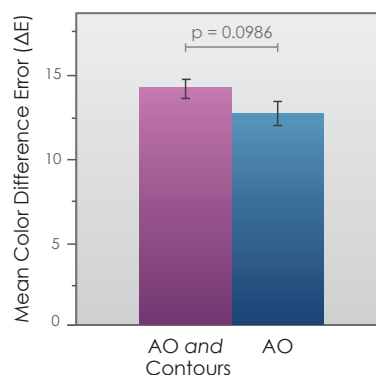


Fig. 14: Enhancing shape using contours resulted in marginally decreased performance over ambient occlusion alone (S6).

and monocular viewing as within-participants factors. Stereo and monocular viewing were blocked, with participants waiting at least 24 hours between each block to mitigate learning effects. While the size of this pilot limits its statistical power, the results provide evidence of the influence of stereo vision on constancy effects.

**Results:** We found evidence of a performance benefit of stereo viewing over monocular viewing ( $\mu_{stereogain} = 2.73\Delta E \pm 1.04$ , 95% confidence interval). While we found no evidence of a benefit for the darkest shadow level ( $\mu_{stereogain} = 0.03\Delta E \pm 1.75$ ), we did find a performance improvement for medium ( $\mu_{stereogain} = 6.88\Delta E \pm 2.01$ ) and light ( $\mu_{stereogain} = 1.28\Delta E \pm 0.66$ ) shadows.

Surfaces viewed in stereo supported more accurate color identification than in the monocular condition (Fig. 13). These findings support the observations about S5 in §7.1: binocular depth cues provided by stereo viewing may improve overall color identification in surface visualization. Further study is needed to verify the magnitude of this effect.

## 7.3 How do added shape cues affect performance? (S6)

Sections 7.1 and 7.2 together indicate a correlation between depth cues and performance on our color matching task. However, techniques like directional lighting also improve perceptions of surface shape. To reason about how shape perceptions might

influence performance, we also measured performance for ambient occlusion surfaces with suggestive contours. Suggestive contours [66] are used to in surface visualization to enhance representations of surface shape. Contours use lines instead of shading to emphasize high-level depth discontinuities along the surface, creating an image resembling a hand-drawn sketch. In previous studies [45], adding contours to a shaded surface inhibited shadow perceptions. Given the importance of shadow perception for constancy effects, as suggested by **S3** (§5.3), contours may consequently inhibit participants' abilities to accurately map surface colors to their corresponding original key color.

**Methods:** Two sets of stimuli were generated: one consisting of visualizations using ambient occlusion and a corresponding set using ambient occlusion plus suggestive contours. Contours were generated using the implementation provided by DeCarlo et al. in the TriMesh package [66] and layered on top of the original ambient occlusion surface. The procedure was otherwise identical to the previous experiments (§4).

Data was collected from 55 participants from Mechanical Turk. One participant was excluded for poor performance on validation questions, resulting in 27 participants per condition.

**Results:** We ran a three-way ANCOVA (contours, shadow, and ramp) on the difference between the key and response colors. We found a significant main effect of shadow  $F(1,278) = 57.0108$ ,  $p < .0001$ . Adding contours marginally decreased performance over ambient occlusion alone ( $F(1,278) = 2.7329$ ,  $p = .0986$ , Fig. 14). The marginal decrease in performance points to a potential trade-off between shape and color identification performance for non-photorealistic rendering techniques in surface visualization. Combined with previous results [45], this provides further evidence that accurate shadow perceptions improve the interpretation of color-coded data on visualized surfaces.

## 8 DISCUSSION AND DESIGN IMPLICATIONS

In surface visualization, viewers explore data in the context of surface structure. Surface structure is commonly conveyed through shadows and shading, which may obscure information encoded on the surface. Supporting lightness constancy in visualization can improve how well a visualization supports accurately reading encoded data in shadow. Our results, summarized in Table 1, demonstrate that despite several approximations made in surface visualizations rendered using ambient occlusion, viewers are able to interpret color-coded data along these surfaces. Performance for this task is directly influenced by visualization design. **S1** through **S3** isolate constancy effects and suggest that the visual system leverages information about synthetic shadows to disentangle encoded data from surface features, and **S4** suggests that these effects are sufficient to preserve performance gains from luminance variation for common color mappings. **S5** and **S6** inform how visualization design can influence color identification performance.

Visualization techniques generally represent trade-offs: they often improve performance for certain types of task at the expense of others. Surface visualizations have traditionally been concerned with supporting depth percepts to convey structure and color percepts to convey additional data about that structure. Our findings suggest that improving percepts of depth and of color may go hand-in-hand: techniques that enhance the apparent depth of a visualized surface may also improve how effectively the

visualization communicates encoded data. Designers can leverage this correlation to develop visualizations that effectively convey surface data in context. Supplementing ambient occlusion with other visualization techniques that enhance depth perceptions, such as directional lighting or stereo viewing, may improve how effectively viewers can interpret information encoded using color on a surface.

This coordination between depth and color may be part of a delicate balance. The visual system appears to be sensitive to the methods used to communicate surface structure. **S3** and **S6** collectively suggest that some design choices can hinder perceptions of surface colors. The visual system likely processes shadows generated by ambient occlusion as more than simply surface shading. Small variations that damage the physical basis of these shadows can significantly diminish viewers' abilities to correctly interpret color encodings. Further, that contours marginally degraded constancy effects suggests that simply enhancing the perception of surface shape is not enough to improve this ability. Such design decisions may represent a trade-off between perceptions of encoded data and of surface structure and could be used to inform task-driven design.

We anticipate that these findings will generalize to other types of surfaces beyond solvent-excluded molecular surfaces. While molecular surfaces represent a realistic use case where correctly inferring data in shadowed regions is often important, these surfaces are unfamiliar complex visual structures to our non-expert participants. The tested surfaces in the context of these studies therefore simply represent smooth, amorphous structures. As constancy effects are influenced by object familiarity [67], we would anticipate that for naïve observers our results provide a baseline measure for color identification performance in ambient occlusion surface visualization more generally. Although these structures do not represent all possible surface structures (they are continuous and have no sharp corners), we believe that our results generalize to surfaces beyond solvent-excluded molecular surfaces but recognize that verifying this is important future work.

## 9 LIMITATIONS AND FUTURE WORK

Our work represents initial steps in understanding how visualization design can support viewers in accurately interpreting color encodings for effective surface visualization. We focus on measuring performance across a small set of common design decisions for molecular surface visualization. Exploring other aspects of design could provide a deeper understanding of how to better support color encodings and other percepts in surface visualization, such as exploring effects of interaction, ramp design, or other shadow approximations like depth darkening. Comparison to more rigorous global illumination models could help illuminate how approximations made by ambient occlusion influence surface perception. Generalizing these explorations across additional surfaces (e.g. space-filling models) or to surfaces in other domains (e.g. aerodynamics) would create a more general understanding of constancy in order to inform effective surface visualization design.

One notable trade-off of this work is that we measured perceptions for lay participants with no known shape priors for the tested molecular surfaces. While this lack of experience may improve the generalizability of these findings to other domains, experienced biochemists may perform differently—their prior knowledge of the surfaces may aid them in disambiguating shape and color. In practice, we observe that experienced biochemists prefer using

Study	Conclusion
S1	Viewers can read shadowed colors on ambient occlusion surfaces
S2	Structure helps viewers interpret shadowed colors
S3	Precise shadow information supports accurate color interpretation
S4	Luminance cues improve performance for common ramps
S5	Visualization techniques that improve depth perception enable viewers to more accurately identify shadowed colors
S6	Visualization techniques that improve shape perception may not improve performance

TABLE 1: Summary of Results

high-quality visualizations where they rely on perception rather than their own prior experiences to understand their data in context. Crowdsourcing allows us to remove potential confounds from shape priors. It also allows us to measure effects across a sampling of different viewing conditions to consider constancy in the noise of everyday viewing. We can not guarantee that our results are not biased by this sampling, but hope that our results are instead sufficiently robust—our results are consistent across both crowdsourced and in-person studies.

The task used in these studies was somewhat artificial by necessity. While identifying color values on a surface is a standard visualization task, each image in our experiments had only one colored patch. This removes potential complications due to contrast between patches or judgments from comparing multiple surface patches, both of which are possible in standard scenarios but would interfere with our ability to measure color identification performance as a function of visualization design. Future explorations might consider more complex visualizations and tasks.

## 10 CONCLUSION

Color is an effective and commonly used cue for visualizing data on surfaces. However, visualization techniques that communicate surface shape often do so using surface shading. This shading can confound data encoded using color, as colors are darkened by shadows. Lightness constancy provides a perceptual mechanism for bridging this complication, allowing viewers to interpret shadowed colors in the real world. Its effectiveness in complex synthetic environments such as surface visualizations is not well understood. In this paper, we confirmed the existence of lightness constancy for molecular surfaces rendered using ambient occlusion and present an initial exploration of how visualization design can impact the effectiveness of color encodings on these surfaces. These studies offer initial insight into how a consideration of constancy mechanisms can help guide effective visualization design.

## ACKNOWLEDGMENTS

The authors wish to thank Charles D. Hansen for his input on this project and Adrian Mayorga for his help with the protein surface visualization implementations. This work was supported in part by NSF awards IIS-1162037, CMMI-0941013 and NIH award 5R01AI077376-07.

## REFERENCES

[1] F. A. Kingdom, “Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy,” *Vision Res.*, vol. 51, no. 7, pp. 652–673, 2011.

[2] M. Olkkonen, C. Witzel, T. Hansen, and K. Gegenfurtner, “Categorical color constancy for rendered and real surfaces,” *J. Vis.*, vol. 9, no. 8, pp. 331–331, 2009.

[3] H. Landis, “Production-ready global illumination,” *Siggraph Course Notes*, vol. 16, no. 2002, p. 11, 2002.

[4] A. I. Ruppertsberg, M. Bloj, and A. Hurlbert, “Sensitivity to luminance and chromaticity gradients in a complex scene,” *J. Vis.*, vol. 8, no. 9, 2008.

[5] D. H. Foster, “Color constancy,” *Vision Res.*, vol. 51, no. 7, pp. 674–700, 2011.

[6] S. Newhall, R. Burnham, and R. Evans, “Color constancy in shadows,” *J. Opt. Soc. Am.*, vol. 48, no. 12, pp. 976–984, 1958.

[7] C. Ware, *Information Visualization*, 3rd ed. Morgan Kaufmann, 2000.

[8] E. H. Land *et al.*, *The retinex theory of color vision*. Scientific America, 1977.

[9] D. H. Brainard and W. T. Freeman, “Bayesian color constancy,” *JOSA A*, vol. 14, no. 7, pp. 1393–1411, 1997.

[10] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou, “An anchoring theory of lightness perception,” *Psych. Rev.*, vol. 106, no. 4, p. 795, 1999.

[11] P. Bressan, “The place of white in a world of grays: a double-anchoring theory of lightness perception,” *Psych. Rev.*, vol. 113, no. 3, p. 526, 2006.

[12] M. E. Rudd, “How attention and contrast gain control interact to regulate lightness contrast and assimilation: a computational neural model,” *J. Vis.*, vol. 10, no. 14, p. 40, 2010.

[13] F. Kingdom and B. Moulden, “A multi-channel approach to brightness coding,” *Vision Res.*, vol. 32, no. 8, pp. 1565–1582, 1992.

[14] M. Rutherford and D. Brainard, “Lightness constancy: A direct test of the illumination-estimation hypothesis,” *Psych. Sci.*, vol. 13, no. 2, pp. 142–149, 2002.

[15] J. Cataliotti and A. Gilchrist, “Local and global processes in surface lightness perception,” *Perception & Psychophysics*, vol. 57, no. 2, pp. 125–135, 1995.

[16] S. Grossberg and S. Hong, “A neural model of surface perception: Lightness, anchoring, and filling-in,” *Spatial Vision*, vol. 19, no. 2, pp. 263–321, 2006.

[17] J. J. Granzier, E. Brenner, and J. B. Smeets, “Can illumination estimates provide the basis for color constancy?” *J. Vis.*, vol. 9, no. 3, 2009.

[18] V. M. de Almeida, P. T. Fiadeiro, and S. M. Nascimento, “Effect of scene dimensionality on colour constancy with real three-dimensional scenes and objects,” *Perception*, vol. 39, no. 6, p. 770, 2010.

[19] M. Hedrich, M. Bloj, and A. I. Ruppertsberg, “Color constancy improves for real 3d objects,” *J. Vis.*, vol. 9, no. 4, 2009.

[20] M. Olkkonen, T. Hansen, and K. R. Gegenfurtner, “Color appearance of familiar objects: Effects of object shape, texture, and illumination changes,” *J. Vis.*, vol. 8, no. 5, 2008.

[21] S. R. Allred and D. H. Brainard, “Contrast, constancy, and measurements of perceived lightness under parametric manipulation of surface slant and surface reflectance,” *JOSA A*, vol. 26, no. 4, pp. 949–961, 2009.

[22] T. Agostini and A. Galmonte, “Perceptual organization overcomes the effects of local surround in determining simultaneous lightness contrast,” *Psych. Sci.*, vol. 13, no. 1, pp. 89–93, 2002.

[23] E. H. Adelson, “Perceptual organization and the judgment of brightness,” *Science-AAAS*, vol. 262, no. 5142, pp. 2042–2044, 1993.

[24] A. D. Logvinenko, “Lightness induction revisited,” *Perception-London*, vol. 28, pp. 803–816, 1999.

[25] E. H. Adelson and A. P. Pentland, “The perception of shading and reflectance,” *Perception as Bayesian inference*, pp. 409–423, 1996.

[26] F. A. Kingdom, “Perceiving light versus material,” *Vision Res.*, vol. 48, no. 20, pp. 2090–2105, 2008.

[27] G. Cipriano and M. Gleicher, “Molecular surface abstraction,” *IEEE TVCG*, vol. 13, no. 6, pp. 1608–1615, 2007.

[28] —, “Text scaffolds for effective surface labeling,” *IEEE TVCG*, vol. 14, no. 6, pp. 1675–1682, 2008.

[29] A. Sarikaya, D. Albers, J. Mitchell, and M. Gleicher, “Visualizing validation of protein surface classifiers,” in *Computer Graphics Forum*, vol. 33, no. 3. Wiley Online Library, 2014, pp. 171–180.

[30] R. B. Corey and L. Pauling, “Molecular models of amino acids, peptides, and proteins,” *Rev. Sci. Instrum.*, vol. 24, no. 8, pp. 621–627, 1953.

[31] L. Pauling, R. B. Corey, and H. R. Branson, “The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 37, no. 4, pp. 205–211, 1951.

[32] B. Lee and F. M. Richards, “The interpretation of protein structures: estimation of static accessibility,” *J. Mol. Biol.*, vol. 55, no. 3, pp. 379–IN4, 1971.

[33] F. M. Richards, “Areas, volumes, packing, and protein structure,” *Annu. Rev. Biophys.*, vol. 6, no. 1, pp. 151–176, 1977.

[34] W. L. DeLano, “The PyMOL molecular graphics system,” 2002.

[35] R. M. Hanson, “Jmol—a paradigm shift in crystallographic visualization,” *Journal of Applied Crystallography*, vol. 43, no. 5, pp. 1250–1260, 2010.

- [36] W. Humphrey, A. Dalke, K. Schulten *et al.*, “Vmd: visual molecular dynamics,” *J. Mol. Graphics Modell.*, vol. 14, no. 1, pp. 33–38, 1996.
- [37] R. M. Andrei, M. Callieri, M. F. Zini, T. Loni, G. Maraziti, M. C. Pan, and M. Zoppè, “Intuitive representation of surface properties of biomolecules using BioBlender,” *BMC Bioinformatics*, vol. 13, no. 4, 2012.
- [38] D. Borland, “Ambient occlusion opacity mapping for visualization of internal molecular structure,” *Journal of WSCG*, pp. 17–24, 2011.
- [39] M. Krone, M. Falk, S. Rehm, J. Pleiss, and T. Ertl, “Interactive exploration of protein cavities,” *Comput. Graph. Forum*, vol. 30, no. 3, pp. 673–682, 2011.
- [40] F. Lindemann and T. Ropinski, “About the influence of illumination models on image comprehension in direct volume rendering,” *IEEE TVCG*, vol. 17, no. 12, pp. 1922–1931, 2011.
- [41] P. Grosset, M. Schott, G.-P. Bonneau, and C. D. Hansen, “Evaluation of Depth of Field for Depth Perception in DVR,” in *Proc. IEEE Pacific Visualization*, 2013.
- [42] M. Tarini, P. Cignoni, and C. Montani, “Ambient occlusion and edge cueing for enhancing real time molecular visualization,” *IEEE TVCG*, vol. 12, no. 5, pp. 1237–1244, 2006.
- [43] M. S. Langer, H. H. Bühlhoff *et al.*, “Depth discrimination from shading under diffuse lighting,” *Perception*, vol. 29, no. 6, p. 649, 2000.
- [44] B. L. Anderson and J. Winawer, “Layered image representations and the computation of surface lightness,” *J. Vis.*, vol. 8, no. 7, p. 18, 2008.
- [45] J. M. Kennedy and J. Bai, “Cavanagh and leclerc shape-from-shadow pictures: Do line versions fail because of the polarity of the regions or the contour?” *Perception*, 2000.
- [46] L. H. Hardy, G. Rand, and M. C. Rittler, “Tests for the detection and analysis of color-blindness,” *J. Opt. Soc. Am.*, vol. 35, no. 4, pp. 268–271, 1945.
- [47] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, “A standard default color space for the internet-sRGB,” *Microsoft and Hewlett-Packard Joint Report*, 1996.
- [48] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, “The protein data bank,” *Nucleic Acids Res*, vol. 28, pp. 235–242, 2000.
- [49] M. F. Sanner, A. J. Olson, and J.-C. Spehner, “Reduced surface: an efficient way to compute molecular surfaces,” *Biopolymers*, vol. 38, no. 3, pp. 305–320, 1996.
- [50] D. Albers Szafir, M. Stone, and M. Gleicher, “Adapting color difference for design,” in *Proc. 22nd Color and Imaging Conf.*, November 2014.
- [51] M. Buhrmester, T. Kwang, and S. Gosling, “Amazon’s Mechanical Turk a new source of inexpensive, yet high-quality, data?” *Perspect. Psychol. Sci.*, vol. 6, no. 1, pp. 3–5, 2011.
- [52] J. Heer and M. Bostock, “Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design,” in *Proc. CHI*. ACM, 2010, pp. 203–212.
- [53] A. Kittur, E. Chi, and B. Suh, “Crowdsourcing user studies with Mechanical Turk,” in *Proc. CHI*. ACM, 2008, pp. 453–456.
- [54] A. L. Gilchrist and V. Annan, “Articulation effects in lightness: Historical background and theoretical implications,” *Perception*, vol. 31, no. 2, pp. 141–150, 2002.
- [55] E. H. Adelson, “Lightness perception and lightness illusions,” *The New Cognitive Neurosciences*, p. 339, 1999.
- [56] A. Gilchrist and A. Jacobsen, “Perception of lightness and illumination in a world of one reflectance,” *Perception*, vol. 13, no. 1, pp. 5–19, 1984.
- [57] K. Moreland, “Diverging color maps for scientific visualization (expanded),” in *Proceedings of the 5th International Symposium on Visual Computing*. Citeseer, 2009.
- [58] C. A. Brewer, G. W. Hatchard, and M. A. Harrower, “Colorbrewer in print: a catalog of color schemes for maps,” *Cartogr. Geogr. Inf. Sci.*, vol. 30, no. 1, pp. 5–32, 2003.
- [59] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman, and H. Pfister, “Evaluation of artery visualizations for heart disease diagnosis,” *IEEE TVCG*, vol. 17, no. 12, pp. 2479–2488, 2011.
- [60] D. Borland and R. M. Taylor, “Rainbow color map (still) considered harmful,” *IEEE Comput. Graph.*, vol. 27, no. 2, pp. 14–17, 2007.
- [61] G. Mather and D. R. Smith, “Combining depth cues: effects upon accuracy and speed of performance in a depth-ordering task,” *Vision Res.*, vol. 44, no. 6, pp. 557–562, 2004.
- [62] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, “UCSF Chimera visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [63] M. S. Langer and H. H. Bulthoff, “A prior for global convexity in local shape-from-shading,” *Perception*, vol. 30, no. 4, pp. 403–410, 2001.
- [64] J. N. Yang and S. K. Shevell, “Stereo disparity improves color constancy,” *Vision Res.*, vol. 42, no. 16, pp. 1979–1989, 2002.

- [65] D. Buckley, J. P. Frisby, and J. Freeman, “Lightness perception can be affected by surface curvature from stereopsis,” *Perception-London*, vol. 23, pp. 869–869, 1994.
- [66] D. DeCarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella, “Suggestive contours for conveying shape,” in *ACM T. Graphic.*, vol. 22, no. 3. ACM, 2003, pp. 848–855.
- [67] M. Olkkonen, T. Hansen, and K. R. Gegenfurtner, “Color appearance of familiar objects: Effects of object shape, texture, and illumination changes,” *J. Vis.*, vol. 8, no. 5, p. 13, 2008.



**Danielle Albers Szafir** is an Assistant Professor and member of the founding faculty of the Department of Information Science at the University of Colorado Boulder. Her research focuses on increasing the scalability and comprehensibility of information visualization by quantifying perception and cognition for visualization design. Her work develops interactive systems and techniques for exploring large and complex datasets in domains ranging from biology to the humanities. Prior to joining CU, she received her Ph.D.

in Computer Sciences at the University of Wisconsin-Madison and holds a B.S. in Computer Science from the University of Washington.



**Alper Sarikaya** is a Ph.D. candidate in the Department of Computer Sciences at the University of Wisconsin-Madison working under Professor Michael Gleicher in the area of data visualization. His research has focused on building visual interfaces to help viewers understand machine learning models and realize connections within large amounts of data. Prior to joining the University of Wisconsin, Alper worked in the Windows Telemetry group at Microsoft Corporation and received his B.S. in Computer Science and

Chemistry from the University of Washington.



**Michael Gleicher** is a Professor in the Department of Computer Sciences at the University of Wisconsin, Madison. Prof. Gleicher is founder of the Department’s Visual Computing Group. His research interests span the range of visual computing, including data visualization, image and video processing tools, virtual reality, and character animation techniques for films, games and robotics. Prior to joining the university, Prof. Gleicher was a researcher at The Autodesk Vision Technology Center and in Apple Computer’s Advanced Technology Group. He earned his Ph. D. in Computer Science from Carnegie Mellon University, and holds a B.S.E. in Electrical Engineering from Duke University. For the 2013-2014 academic year, he was a visiting researcher at INRIA Rhone-Alpes. Prof. Gleicher is an ACM Distinguished Scientist.