Beyond scattering and absorption: Perceptual unmixing of translucent liquids

Alice C. Chadwick Department of Psychology, University of Durham, Durham University Science Site, Durham, UK

George Cox Department of Experimental Psychology, University of Oxford, Oxford, UK

Hannah E. Smithson* Department of Experimental Psychology, University of Oxford, Oxford, UK

Robert W. Kentridge* Department of Psychology, University of Durham, Durham University Science Site, Durham, UK

Is perception of translucence based on estimations of scattering and absorption of light or on statistical pseudocues associated with familiar materials? We compared perceptual performance with real and computer-generated stimuli. Real stimuli were glasses of milky tea. Milk predominantly scatters light and tea absorbs it, but since the tea absorbs less as the milk concentration increases, the effects of milkiness and strength on scattering and absorption are not independent. Conversely, computer-generated stimuli were glasses of "milky tea" in which absorption and scattering were independently manipulated. Observers judged tea concentrations regardless of milk concentrations, or vice versa. Maximum-likelihood conjoint measurement was used to estimate the contributions of each physical component—concentrations of milk and tea, or amounts of scattering and absorption—to *perceived* milkiness or tea strength. Separability of the two physical dimensions was better for real than for computer-generated teas, suggesting that interactions between scattering and absorption were correctly accounted for in perceptual unmixing, but unmixing was always imperfect. Since the real and rendered stimuli represent different physical processes and therefore differ in their image statistics, perceptual judgments with these stimuli allowed us to identify particular pseudocues (presumably learned with real stimuli) that explain judgments with both stimulus sets.

Introduction

The perception of colors, textures, and shapes, and the identification of materials, are all things that we take for granted in everyday visual experience. However, the underlying mechanisms are not fully understood. There are open questions not only about the processing of incoming sensory information but also about the selection of perceptual estimates to support particular judgments. Considerable effort has been put into answering these questions for color and, to a lesser extent, texture and glossiness. Very little, however, is known about perceived translucence (Anderson, 2011). To make judgments about translucence, do we perceive and make estimates of the physical characteristics of light transport—scattering and absorption—or do we rely on heuristics[1] or pseudocues?[2] The results we present here provide the first systematic assessment of how humans perceive scattering and absorption of light within translucent materials.

Color, gloss, and translucence are perceptual properties of objects in the world. They can be thought of as estimates of the different interactions of light with the materials of which the object is made. An object's color depends upon the efficiency with which its surface reflects light of different wavelengths. The glossiness of an object depends on how randomly light is scattered when it is reflected from the object's surface. The quality of translucence depends on the way in which light is scattered and absorbed as it passes through the bulk of a material before reaching our eyes, either being scattered back to us or passing entirely through an object. However, despite the dependence of each of these perceptual properties on the object's physical properties, the information available to us when looking at an object is insufficient to recover full veridical measurements of any of the corresponding

Citation: Chadwick, A. C., Cox, G., Smithson, H. E., & Kentridge, R. W. (2018). Beyond scattering and absorption: Perceptual unmixing of translucent liquids. *Journal of Vision*, *18*(11):18, 1–15, https://doi.org/10.1167/18.11.18.

https://doi.org/10.1167/18.11.18 Received November 8, 2017; published October 29, 2018 ISSN 1534-7362

physical properties of the material. In material color perception (Smithson, 2005) and glossiness perception (Chadwick & Kentridge, 2015), it is clear that we employ heuristics or pseudocues that, while not veridical, usually produce good approximations to the qualities of a material's surface color or gloss. Much less is known about the heuristics or cues that drive translucence perception.

We are specifically interested in understanding the contributions of physical properties of light transport in a material to translucence perception and the visual mechanisms by which this occurs. We use as test stimuli photographs of translucent liquids (glasses of milky tea) whose physical constituents have been systematically varied. We obtain perceptual judgments of the physical properties of the liquid (strength and milkiness), testing the extent to which observers can perceptually unmix the liquids to judge differences in their physical makeup. With these familiar stimuli, the physical components we vary (concentrations of milk and tea) do not map independently onto changes in scatter and absorption. We therefore generate a comparison set of computer-generated stimuli that are to be subjected to the same perceptual judgments (of strength and milkiness), but in which scatter and absorption are perfectly independent. Comparison of the different perceptual responses obtained with the two stimulus sets provides strong tests of the underlying mechanisms of material perception with translucent liquids. For any image statistic that captures the pattern of observers' judgments with the real stimuli, we can establish whether the same statistic predicts behavior more generally by testing whether it also predicts the different patterns of responses when observers judge artificial stimuli.

Previous research

The perception of translucence has been relatively little studied, in contrast to work on other visual properties such as color. Metelli (1970), while not the first to study the phenomenon, made one of the first significant attempts to explain perceived transparency (the absorption component of the more general property of translucence). He outlined potential algebraic conditions under which we perceive transparency, in terms of relative figural relationships and ordinal relationships between colors in a scene. However, this was only relevant to specific limited contexts and materials. Subsequent work until the early 2000s went on to define other possible comparative relationships between materials in terms of conditions that might have to be satisfied for the visual system to define a surface as transparent. This work produced descriptive laws but did not attempt to explain how we perceive the translucence itself (Adelson & Anandan, 1990; Beck, Prazdny, & Ivry, 1984; Singh & Hoffman, 1998). Theories of translucence developed further as technology improved, with research showing that it had many other important aspects (Jensen, Marschner, Levoy, & Hanrahan, 2001; Koenderink & van Doorn, 2001; Singh & Anderson, 2002a, 2002b). It also became clear that most published work did not attempt to address the fundamental questions of how translucence was perceived, although it was evident that observers were very good at discriminating and identifying characteristics of translucence, like opacity, at short presentation times (Sharan, Rosenholtz, & Adelson, 2009). Fleming and colleagues concluded that many of the simple cues that had been proposed for the perception of translucence were unable to predict how the translucence of a material was perceived (Fleming & Bülthoff, 2005; Fleming, Torralba, & Adelson, 2004). They argued that the physics of translucence—much like the physics of surface reflectance, in color vision—were too complex for the visual system to be able to estimate using inverse optics (i.e., working backward from the stimulus received to the material most likely to have produced it based on mental models of light transport). In addition, those authors argued that simple image statistics were inadequate when used alone. Nevertheless, they proposed that perceived translucence was achieved by means of parsing scenes into key regions and gathering image statistics from those regions, though the precise statistics remained unknown.

More recently, Fleming, Jäkel, and Maloney (2011) noted that much of the work conducted up to that point had focused mainly on thin filters rather than solid materials, and had predominantly drawn conclusions based on inferences from the physics of light transport without empirical tests. They concluded that distortions in the perceived shape of objects, caused by thick transparent objects, were an important part of the evidence indicating the translucence of material volumes, and that this information could comprise an additional class of cues for the visual system. Gkioulekas et al. (2013) also identified that multiple scattering (forward and backward scattering of light within a volume) contributes to the translucent appearance of materials and showed that the phase function of scattering, which describes the distribution of scattering directions, can contribute to a translucent appearance, changing the blurring and brightness of objects seen through the translucent material. This is not, however, necessarily a measure that the visual system attempts to approximate when making judgments of translucence. Similarly, Motoyoshi (2010) considered image properties that might be diagnostic of perceived translucence, identifying a role for the spatial contrast in specular highlight and body components of the image. Beyond pure physical characteristics, it has also been found that contextual variation—such as lighting direction—has a significant effect on perceived translucence; front lighting made translucent objects appear more translucent, and back lighting made them appear more opaque (Xiao et al., 2014).

Our approach

The first stage in our strategy is to estimate the way in which perceived translucence depends on two familiar physical parameters—the strength and milkiness of tea—that have systematic effects on light transport. To obtain such estimates we employ a maximum-likelihood conjoint-measurement (MLCM) task (Knoblauch & Maloney, 2012). We used images of real tea, varying in the concentrations of milk and tea, and we asked observers to make judgments of either milkiness regardless of tea strength or tea strength regardless of milkiness. Using glasses of milky tea as stimuli provided a convenient and familiar way of asking observers about the nature of the translucent material, as questions could be framed in terms of tea strength or milkiness rather than more abstract terms such as absorption or scattering. Importantly, tea concentration primarily affects absorption, and milkiness primarily affects scatter (Aernouts et al., 2015; Narasimhan et al., 2006), although when tea and milk are combined there are interactions between them (Hasni et al., 2011). The two physical factors, strength and milkiness, could affect perceptual judgments in three different ways, each best described by a different type of model. One possibility is that only one physical property contributes to judgments of the corresponding perceived property—for example, milk concentration might be the sole determinant of perceived milkiness. If this were the case, an *independent* model—which uses only a single physical variable to model performance in each task—would provide the most parsimonious fit to the data. Alternatively, both physical factors might affect perceptual judgments but do so separately, so that the effect of a change in one physical factor will remain the same no matter what the strength of the other physical factor. An *additive* model is the most parsimonious way of describing this scenario. Finally, a *saturated* model is required if the two physical variables interact with each other in the effect they have on perceptual judgments. We fit these three models to our observers' judgments using MLCM analysis, and the model of best fit is determined using a nested-hypothesis test.

The second stage in our strategy is to use computer-generated images of tea in which scattering and absorption of light are manipulated independently, crucially without the natural interaction that occurs in mixtures of real tea and milk. Observers were again asked to make judgments of tea strength regardless of milkiness or milkiness regardless of tea strength, with these new stimuli. Running experiments with rendered stimuli in addition to real stimuli allows us to perform two tests of the underlying perceptual processes driving observers' perceptual judgments of milkiness and strength. First, we can ask whether these judgments depend separably on scatter and absorption, the two fundamental processes of light transport that are manipulated to produce the rendered images. Second, because the images of real tea and the images of rendered tea vary in different ways, we can test whether candidate pseudocues have the explanatory power to predict differences in performance with the two stimulus sets. We link the physical parameters that describe the two stimulus sets to the respective perceptual judgments by simulating the behavior of an ideal observer in an MLCM task who has access to pseudocues derived from a range of image statistics. It is important to note that we are not looking to identify cues that are immune to changes in context or the environment, or even to test whether cues are specifically necessary for identifying translucence. We are trying to test whether a single heuristic model can explain the judgments of a single perceptual property with two quite different sets of images. The stimuli are designed so that if this is the case, it suggests that an approach to understanding translucence perception based on pseudocues is more appropriate than one based on inverse optics.

In the real world, there are of course additional sources of information as one moves. Using photographs of real stimuli makes these more comparable to rendered images, and we acknowledge that the images do not incorporate the full range of information in real stimuli.

Method

To investigate the basis of perceptual judgments of translucent materials, we asked our observers to make decisions about the properties of a translucent liquid: tea. Each stimulus combined different concentrations of a black-tea solution and differing amounts of milk. We asked our observers to make perceptual judgments about either the perceived strength of the tea (ignoring the milkiness) or the milkiness (ignoring the strength of the tea).

MLCM requires stimuli that vary in at least two physical dimensions—in this case, the concentration of black tea and the amount of milk. We generated stimuli that covered approximately the same perceptual range in both dimensions, with a physical spacing that was chosen to linearize the estimated response function. Such conditions are useful for subsequent data analysis and are likely to optimize the conditions for rejecting the independent model and detecting an interaction in the saturated model. To find the desired stimulus levels for MLCM, we first ran maximum-likelihood difference scaling (MLDS; Knoblauch & Maloney, 2012) with a wide range of candidate stimuli. Using the MLDS results, we then selected stimuli to use in MLCM that were matched in discriminability (expressed by d') for the two physical parameters (generating a perceptually linear "tea-space"). We thus ensured that any interactions found in the main MLCM experiment were not due to scale differences in discriminability in the range of physical parameters selected (this approach has been used elsewhere; Hansmann-Roth & Mamassian, 2017; Rogers, Knoblauch, & Franklin, 2016).

In MLDS, two pairs of stimuli differing along one physical dimension were presented simultaneously, and observers were asked to decide which of the two pairs showed a greater perceptual difference. The two pairs (quadruples) were chosen in accordance with the methods of Knoblauch and Maloney (2012). The maximum-likelihood perceptual scale in units of d' was computed using the Knoblauch and Maloney (2008) MLDS package for R (R Core Team, 2017). Having obtained the MLDS scales, we used them to choose four values of the physical variables that were perceptually equally spaced (in discriminability). The set of stimuli for the MLCM experiment comprised all possible combinations of these values for the physical variables. Since the tea-space included four levels of milkiness and four levels of tea strength, we had 16 MLCM stimuli in total.

In MLCM, pairs of stimuli that differed from one another in terms of both physical variables were presented to observers. On each trial, different pairs of stimuli were presented. Observers were instructed to make a judgment about just one perceptual dimension per task, reporting in which stimulus the tea appeared either milkier or stronger. All possible combinations of stimuli were presented.

Data were analyzed with the Knoblauch and Maloney (2014) MLCM package for R to test which of the three classes of model—independent, additive, or saturated—provided the most parsimonious fit to the perceptual estimates (expressed in units of d'). The models fit latent parameters that describe the contribution of specific levels, or combinations of levels, of the physical parameters to the intensity of the perceptual quality about which the judgments were made.

In an additive model we assume that the contributions of the physical properties scattering s and absorption a to the percept milkiness M add together:

$$\Delta(i,j,k,l) = \left(\psi_i^{M:s} + \psi_j^{M:a}\right) - \left(\psi_k^{M:s} + \psi_l^{M:a}\right), \quad (1)$$

where Δ is the perceived difference in milkiness, $\psi_i^{M:s}$ and $\psi_j^{M:a}$ are the latent parameters referring to the milkiness percepts from a stimulus with the ith level of scatter and the jth level of absorption, and $\psi_k^{M:s}$ and $\psi_l^{M:a}$ are the latent parameters referring to the milkiness percepts from a stimulus with the kth level of scatter and the lth level of absorption. The probability of reporting that the milkiness M of an item with physical scattering and absorption (i, j) is greater than that of an item with scattering and absorption (k, l) is

$$P[\Delta(i,j,k,l)] = P\left[\left(\left(\psi_i^{M:s} + \psi_j^{M:a}\right) - \left(\psi_k^{M:s} + \psi_l^{M:a}\right)\right) > \varepsilon\right] \quad (2)$$

$$P[\Delta(i,j,k,l)] = 1 - \Phi\left(\left(\psi_i^{M:s} + \psi_j^{M:a}\right) - \left(\psi_k^{M:s} + \psi_l^{M:a}\right)\right), \quad (3)$$

where ε is a Gaussian random variable, $\varepsilon \sim \mathcal{N}\left(0, \sigma^2\right)$, representing observer decision noise.

The likelihood function for a single pair of stimuli with physical values (i, j, k, l) having possible response values 1 and 0 is

$$\mathcal{L}(r|i,j,k,l) = P[r=1]^r P[r=0]^{1-r}, \quad (4)$$

where the expected value of r is the probability of making the decision.

We can take the same approach to fitting the other models. An independent model, where perceived milkiness is dependent only on physical scattering, has no absorption parameters. A saturated model, in which physical scattering and physical absorption interact, has separate terms for every combination of physical scattering and absorption, $\psi_{(i,j)}^{M:a,s}$. An independent model fits three latent parameters: one for each value of the physical variable used except the first level, which is fixed. An additive model fits six hypothetical parameters: one for each level of each physical variable, except the first levels of each variable. A saturated model needs separate parameters for every combination of the levels of the physical variables except the combined first levels of each variable, in this case fitting 15 parameters.

When testing the goodness of fit of the models, we used the log likelihood values of each model in nested-hypothesis tests to establish whether there were significant differences between the amounts of deviance (minus twice the log likelihoods) explained by each model. We interpret the results by first comparing the saturated model with the additive model, to see if the additional parameters in the saturated model yield a significantly better model fit. If this is not the case, then we go on to test the additive model against the less complex independent model, again determining whether additional parameters produce a significantly better fit. Where these tests are inconclusive in identifying a model of best fit, we follow this up with a comparison of the saturated and independent models. These nested-hypothesis tests are based on a chi-squared approximation of the distribution of log likelihood ratios (Mood, Graybill, & Boes, 1974).

Methods for real-stimuli experiment

Stimuli

The tea solution we used for each stimulus combined different concentrations of black-tea solution and differing amounts of milk, to produce a fixed volume of liquid in each glass. Further details are given in Supplementary Material, Table S1. Photographs were taken with a Nikon D80 in manual mode, without exposure compensation. The real glasses of tea were placed against a white infinity-curve backdrop in a windowless lab lit from above by a halogen ceiling light and from the front, top, and side by a single D65 daylight desk lamp. An initial set of 14 stimuli was produced for MLDS testing: seven varying in milkiness at the lowest strength of tea and seven varying in tea strength at the lowest level of milkiness. Using the results of the MLDS (see Supplementary Figure S1), a new set of stimuli was generated for MLCM testing, yielding a final set of 16 stimuli (see Supplementary Table S1 for the amount of each liquid used to create the final set of volumes, and Figure 1 for the set of photographic stimuli).

Observers

To determine appropriate levels of milkiness and strength for the physical stimuli, we ran repeated MLDS experiments, revising the stimuli each time, with a pool of five observers. There was good interobserver consistency, and the final perceptually uniform scale was derived from the MLDS data of a single observer. A total of sixteen observers took part in the MLCM tasks, with eight completing each condition (judging either milkiness or strength). All observers were aged 18–25 years and had normal or corrected-to-normal vision.

Apparatus

Stimuli were presented using a MATLAB program (MathWorks, Natick, MA) on a calibrated NEC CRT display, controlled by a CRS Visage (Cambridge Research Systems, Rochester, UK). Responses were collected using a multibutton input device (Cedrus, San Pedro, CA).

MLDS procedure

In a single trial, a fixation point was presented for 0.6 s, followed by a blank screen for 0.4 s, after which a quadruple (two separate pairs of images) was presented on the display screen for 3 s. In each of three blocks, 126 trials were presented, to give a total of 378 trials. The observer judged which of the two pairs of images had a greater perceptual difference (of either milkiness or tea strength) and indicated their decision by pressing one of two keys. The next trial was then triggered. Separate measurements were obtained for stimuli that differed in milkiness or in strength.

MLCM procedure

In a single trial, a fixation point was presented for 0.4 s, and a pair of stimuli were then presented sequentially on the screen for 1.5 s each, separated by a blank screen for 0.5 s. In each of three blocks, 136 trials were presented, to give a total of 408 trials. Observers were asked to respond indicating which of the two appeared to be either milkier or stronger. The next trial was then triggered. Observers were randomly allocated to one of the two MLCM conditions, making judgments of either milkiness or of strength.

Methods for rendered-stimuli experiment

Stimuli

Stimuli were created by simulating a glass tumbler, a volume of liquid, and a scene in Blender (v2.68; 2013), an open-source 3-D computer graphics program that can model 3-D scenes and objects. Images were then rendered using the LuxRender ray-tracing renderer (http://www.luxrender.net), which simulates physical properties of materials, including their light-transmitting and light-scattering properties. It is based on physically-based ray tracing (Pharr & Humphreys, 2004) and uses physically-based equations to simulate the propagation of light and its interaction with materials in a scene in a physically realistic way.

The properties of the liquid were varied by manipulating the degree of physical light absorption and physical light scattering (the absorption spectrum used produced a brown liquid at high values, and the scattering spectrum was uniform and so produced a white milky liquid at high levels of scatter; see Supplementary Material, section 3). The models of the

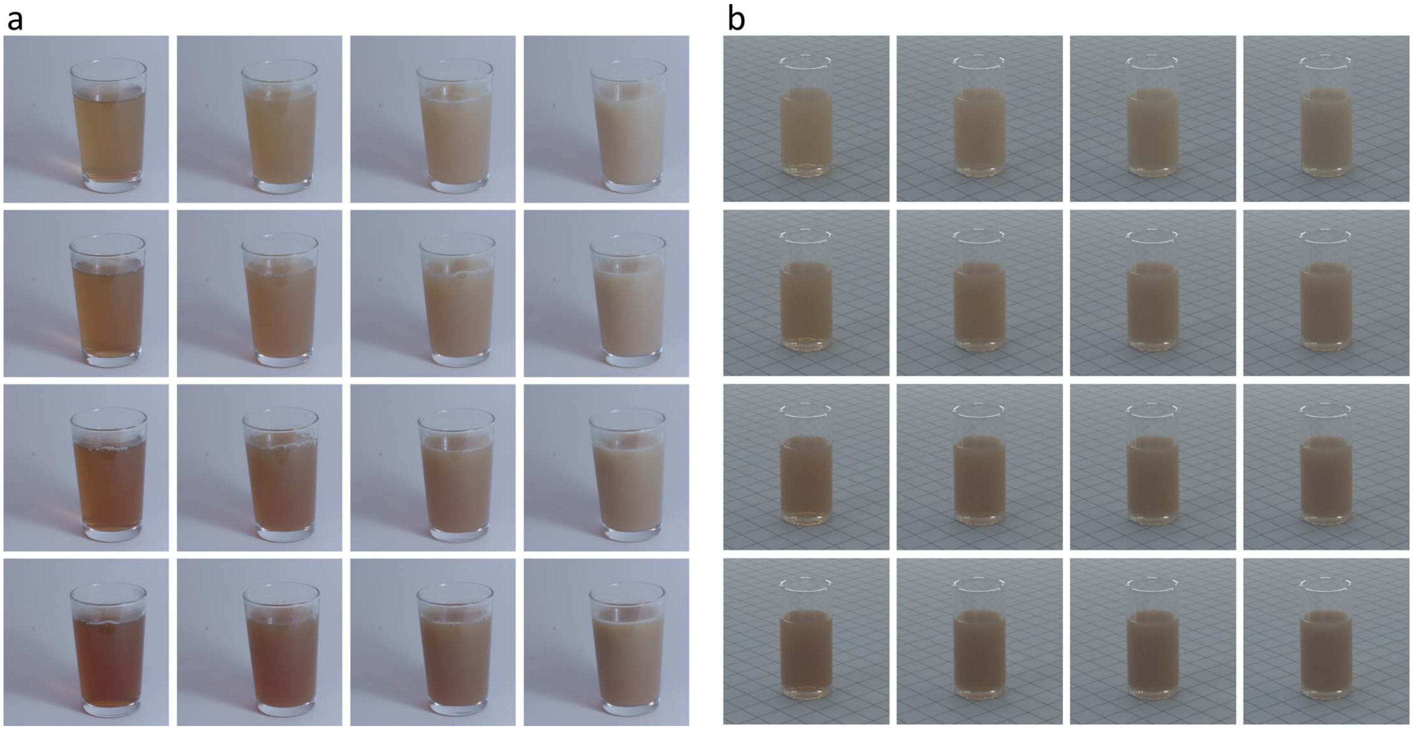

Figure 1. (a) The set of 16 real stimuli used in the MLCM task, with milkiness increasing from left to right and tea strength increasing from top to bottom. (b) The set of 16 rendered stimuli used in the MLCM task, with simulated milkiness (scattering) increasing from left to right and simulated tea strength (absorption) increasing from top to bottom. (These images have been adapted for publication purposes.)

liquid in the tumbler were illuminated by a real-world lighting probe using natural light distributions (Debevec & Malik, 1997), with a single large area lamp behind the glass providing some additional weak diffuse illumination. The original stimulus set comprised 625 rendered images—that is, all combinations of 24 levels of physical absorption and 19 levels of physical scattering. For MLDS testing we selected 11 levels of each variable (see Supplementary Figure S2). MLDS testing allowed us to choose sets of four values for scattering and absorption that were perceptually equally spaced (in discrimination space) to produce a final set of 16 stimuli for the MLCM experiment.

Apparatus

The apparatus was the same as that used with the real stimuli, with the exception of the monitor, which was a calibrated 17-in. ViewSonic CRT color display.

Observers

Three observers participated in both the MLDS tasks and the MLCM tasks, with an additional two observers participating in the MLCM tasks. All were aged 18–25 years and had normal or corrected-to-normal vision.

MLDS procedure

The three observers each completed both conditions of the MLDS tasks, on consecutive days. The stimulus sequence was the same as for the experiment with real stimuli, with minor differences in timing. Instructions for observers were also the same. Each of three blocks presented 330 trials in a randomized order, giving 990 trials in total.

MLCM procedure

All observers completed both conditions of the MLCM tasks on consecutive days. Each condition consisted of four blocks of 180 trials, giving 720 trials in total. The stimulus sequence was the same as for the experiment with real stimuli, with minor differences in timing, and pairs were shown side by side rather than sequentially.

Image statistics for both real and rendered images

Each image was converted from RBG to HSV, and the mean, standard deviation, skew, and kurtosis were calculated for both the saturation and the value (where value is akin to brightness or intensity in this context).

Since light propagation through an attenuating material depends on path length (as summarized by the Beer–Lambert law), we also extracted information about the saturation gradient extending downward from the surface of the liquid, summarized by the space constant on a single exponential fit to the image data. Our model simulations assumed an ideal observer whose decisions in the MLCM task were determined by the relative values of image statistics. For example, the decision as to whether one image was perceived as milkier than another would be based on which of the pair had a lower mean saturation, with an error term that allows performance to be scaled in d' units. We fitted the simulation data to the saturated MLCM parameter estimates of perceived milkiness and strength, and used the adjusted r^2 values from the fits to evaluate how well the candidate image statistics accounted for the observed data.

Results

Results from photographs of real tea

In the comparison of saturated, additive, and independent model fits produced by MLCM analysis of perceived tea strength, the independent model was rejected for all eight observers. The additive model was found to account for a significantly greater amount of the variation (discounting the additional parameter estimates), with $p < 0.01$ in all cases (Supplementary Table S2). This demonstrates that *both* physical variables always contributed to perceptual judgments in this task. In the perceived-milkiness task, the independent model was rejected for seven of the eight observers (Supplementary Table S3). For most observers in both tasks, the physical variables contributed in additive combination to perceptual judgments. So, when judging milkiness, every level of the distractor variable (tea strength) contributed a fixed offset to the perceptual judgment. The same was true for judgments of perceived tea strength, in that milkiness produced a fixed offset to the perceptual judgment (Figure 2a and 2b). Only two of the eight observers needed a model more complex than the additive one to fit judgments of tea strength, and only one needed more than an additive model to fit judgments of milkiness ($p < 0.01$ in all cases). For perceived tea strength, two of the additive observers (CM and DT) showed a clear tendency for increasing milkiness to increase perceived tea strength, whereas for three others (AG, CG, and GT) there was a clear tendency for it to decrease perceived tea strength. In the milkiness task, increased tea strength decreased perceived milkiness for all observers apart from observer BC. Observers' judgments were largely driven by the physical parameter that we asked them to judge, as regressions of the additive models showed greater contribution of "relevant" variables toward parameter estimates (as seen most clearly in the additive plots in Figure 2a and 2b).

To support reliable and useful estimation of the physical constituents of tea mixtures, whenever concentrations of either tea or milk increase, so should perceptual estimates of their concentration. This is exactly what we found with our observers. However, unmixing was not perfect: Additive rather than independent models dominated the results, indicating consistent additive contributions of milkiness on judgments of strength and vice versa. The magnitude of these additive contributions was small but significant across the discriminable scale. Observers are therefore far from perfect at unmixing real tea and milk but are nevertheless responding to tea and milk concentrations in a systematic manner. What exactly, then, are they doing?

The information available for visual perception of the material properties of objects derives from the physical properties of a material, which in turn determine the nature of its interaction with light. Real materials exhibit complex combinations of a variety of types of physical light transport, such as absorption and scattering, operating over multiple spatial scales. In tea specifically, the particles of predominantly light-absorbing tea agglomerate onto fat globules in predominantly scattering milk, so as milk concentration increases, the absorbing power of a given concentration of tea decreases (Hasni et al., 2011).

Perhaps the way real milk and tea interact in their effects on scattering and absorption interfered with unmixing. If our perceptual judgments of mixtures were based purely on estimates of scattering and absorption of the separate constituents, without taking proper account of the interactions between milk and tea that affect scattering and absorption in real tea (Hasni et al., 2011), then images of artificial computer-rendered tea in which milkiness affected only scattering and tea strength affected only absorption, without complex interaction, should be easier to unmix. Furthermore, perceptual judgments with the two classes of stimuli (real and rendered) that arise from different physical processes and that therefore differ in their image statistics would allow us to test the extent to which particular pseudocues (presumably learned with real stimuli) can explain judgments.

Results from rendered stimuli

The rendered stimuli appeared highly realistic, and observers thought that they were images of real glasses of tea. We asked observers exactly the same questions as

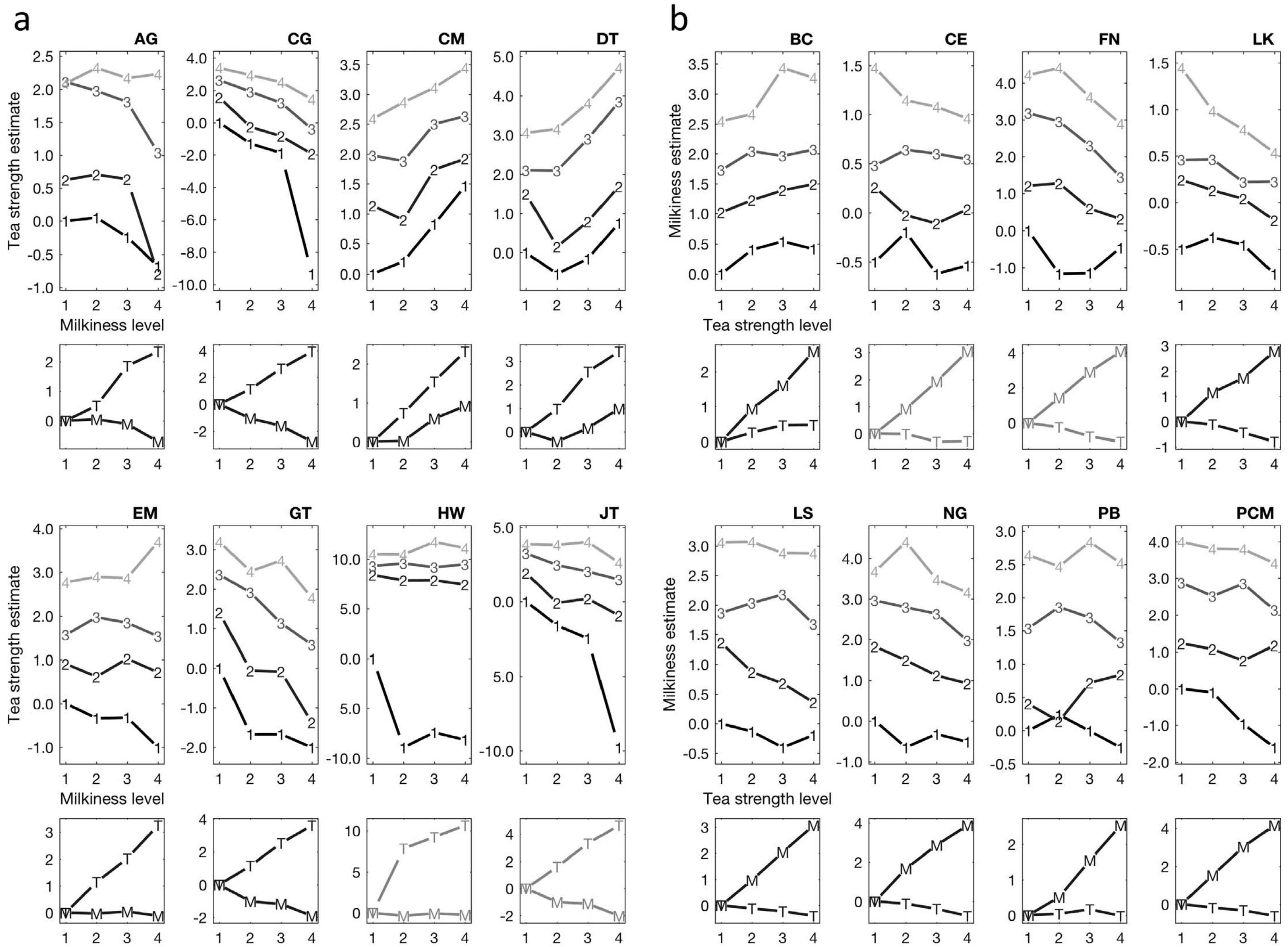

Figure 2. The results of MLCM analyses from the real-tea experiment. A pair of plots is presented for each observer. The upper plots with four lines show parameter estimates from saturated models. The *y*-axis (in units of d') always represents the estimated value of the perceptual property: (a) perceived tea strength and (b) perceived milkiness. The *x*-axis always represents levels of one physical parameter: (a) milk concentration and (b) tea concentration. The four lines represent levels of the second physical parameter: (a) tea concentration and (b) milk concentration. For the physical parameters, the scales from 1 to 4 on both the *x*-axis and between the lines indicate increases in concentration. These plots have characteristic properties for the three models we test. If the data are best fit by an independent model, we should expect the second physical parameter to have no effect on judgments, and so the four lines representing different values of this parameter should therefore be identical. If the data are best fit by an additive model, then the effect of the second physical parameter should be the same at all levels of the first parameter, and so the four lines should be parallel. If the lines are not parallel, this indicates that a saturated model is required. The lower plots with two lines show parameter estimates from additive models. The axes are the same as for the saturated plots in both panels. In these plots the line labeled M denotes the contribution from physical milkiness and the line labeled T denotes the contribution from physical tea strength. For results where the additive model provided the best fit, the additive graph is in bold. The slopes of the two lines indicate the relative additive contributions of the two physical parameters to perceptual judgments.

in the task with real images, that is, to judge tea strength or milkiness. Although observers felt they could estimate each perceptual property separately and thought the two to be conceptually distinct, the perceptual estimates we extracted with MLCM were described by complex saturated combinations of the two physical properties for the majority of observers (see Figure 3a and 3b). For tea strength, the perceptual estimates of four of five observers were best fitted by a saturated model (AC, PM, YB, and EG), and for milkiness that was true of three of five observers (AC, PM, and YB), with an additive model fitting the estimates of the others ($p < 0.01$ in all cases). The independent model never provided an optimal fit (Supplementary Tables S4 and S5). In perceptual terms, variation in scattering and absorption did not map separably onto perceived milkiness and

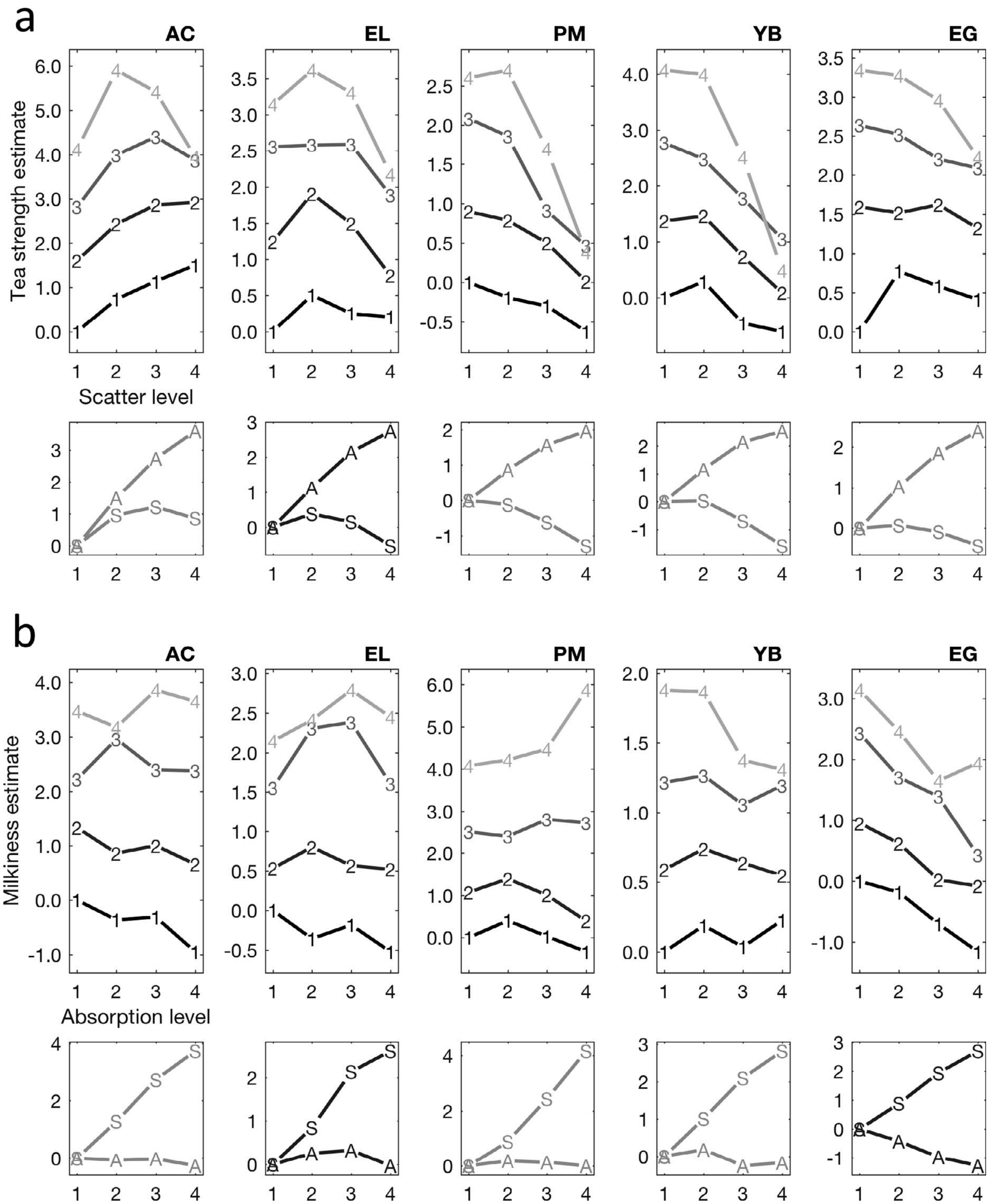

Figure 3. The results of MLCM analyses from the rendered-tea experiment. The format of the plots and the way in which they should be interpreted are analogous to those of Figure 2. The only difference between experiments is that in this case the physical parameters, represented on the *x*-axis and between lines, were scatter and absorption rather than milk concentration and tea concentration. Correspondingly, in the lower plots the line labeled S denotes the contribution from physical scatter and the line labeled A denotes the contribution from physical absorption.

strength. Observers cannot therefore be basing their judgments on independent estimates of the light-transport properties of absorption and scattering in the constituent liquids.

Perhaps rather than estimating light-transport properties and inferring material properties from them, observers use pseudocues in the image that are direct heuristic estimates of concentrations of specific mate-

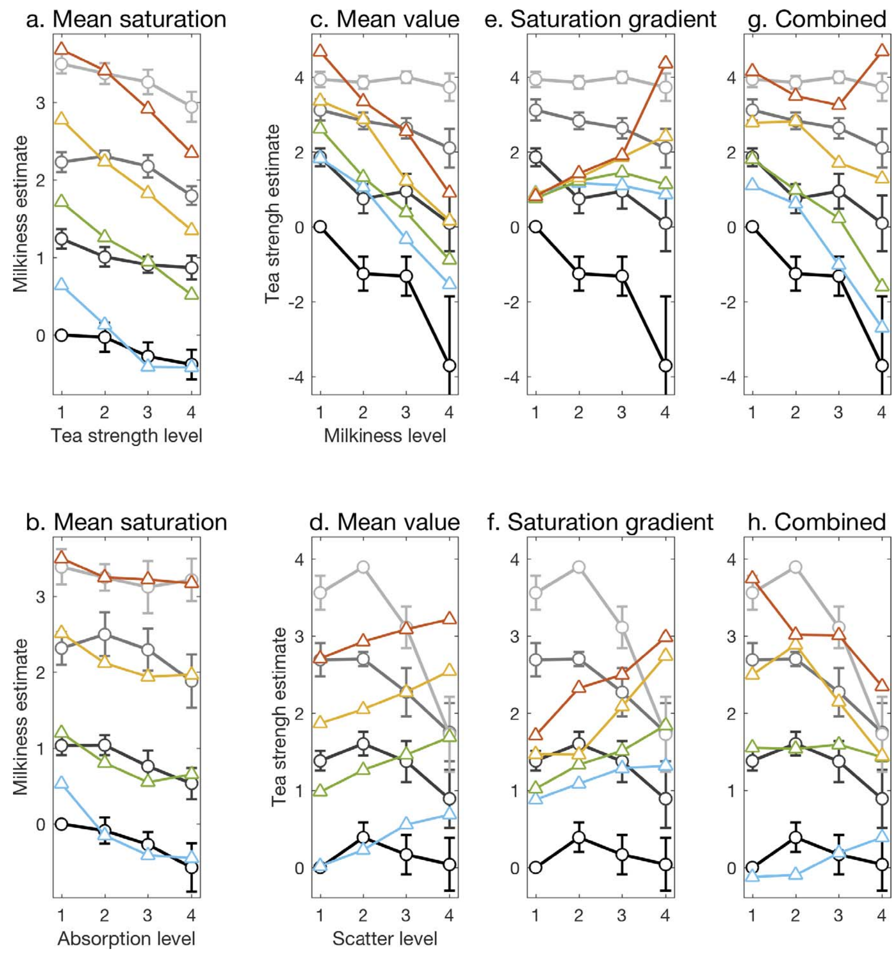

Figure 4. The results from the ideal-observer MLCM simulations in comparison to the MLCM parameter estimates from the saturated models of the two experiments averaged across observers. The top row and bottom rows refer to real and rendered stimuli, respectively. The pair of panels at the extreme left (a and b) refer to milkiness judgments. The three pairs of panels on the right (c and d, e and f, g and h) refer to tea-strength judgments. In all cases, the grayscale lines and symbols represent parameter estimates from human observers (with 95% confidence intervals across observers), and the colored lines and symbols represent the parameter estimates from ideal observers whose responses are governed by candidate image statistics. (a–b) Milkiness estimates with ideal-observer responses based on the mean of saturation, adjusted $r^2 = 0.920$ and 0.970 for real and rendered. (c–d) Strength estimates with ideal-observer responses based on the mean of value, adjusted $r^2 = 0.535$ and 0.670 for real and rendered. (e–f) Strength estimates with ideal-observer responses based on the space constant of gradients of saturation, adjusted $r^2 = 0.096$ and 0.207 for real and rendered. (g–h) Strength estimates with ideal-observer responses based on a weighted sum of the mean of value and gradients of saturation, adjusted $r^2 = 0.812$ and 0.894 for real and rendered.

rials (Chadwick & Kentridge, 2015). By this reasoning, they should be better at perceptual unmixing when physical dimensions correspond to previously encountered materials, and applying learned pseudocues to our rendered images should produce less effective performance. The data suggest that this is the case. With rendered stimuli, complex interacting combinations of the physical dimensions contribute to perceptual judgments of strength. These can produce nonmonotonic effects—sometimes increasing scatter (adding more milk) makes the liquid look stronger, and sometimes less.

Image statistics as pseudocues

If observers are not making estimates of light-transport properties, how are they unmixing the real materials? We conducted a number of simulations of the MLCM task with an ideal observer who has access to pseudocues based on image statistics of the tea region of the images. The averaged data and simulations are shown in Figure 4. Before averaging, the data shown in Figures 2 and 3 were normalized to each observer's maximum d' values, and the normalized average was rescaled to the average d' range for that task. When calculating best fits of the ideal-observer estimates with the averaged estimates from our observers, we used a simple linear model fit with an offset term and scaled contribution(s) from the image statistic(s) under test. Since there is no variability in the extracted image statistics, scaling is necessary to account for the noisy decisions of the real observers. We used the adjusted r^2 values from the fits to evaluate how well the candidate image statistics accounted for the observed data. Mean color saturation (the S of HSV) provided a good explanation of performance in the milkiness task with real tea (adjusted $r^2 = 0.920$). The *same* pseudocue was also best at reproducing performance on the rendered task (adjusted $r^2 = 0.970$), even though the patterns of perceptual estimates of the two tasks were quite different. In the tea-strength tasks, no simple statistics provided a good account of behavior. For this task, the inclusion of spatial information was crucial for getting a good fit to the data. A linear mixture of mean value (the V of HSV) and color-saturation gradient (from the top surface of the liquid into the tea volume, summarized by a fitted exponent describing the space constant of variation in saturation as light penetrates the volume) provided a good account for the real images (adjusted $r^2 = 0.812$). Again, the statistic that successfully accounted for performance in the task with real stimuli also produced the best fit to the perceptual estimates of the rendered stimuli (adjusted $r^2 = 0.894$).

We additionally tested whether the data from individual observers could be accounted for using the same image statistics. For milkiness judgments, mean color saturation provided a very good explanation (adjusted $r^2 > 0.870$) for all five observers viewing rendered tea and for six out of eight observers viewing real tea. The remaining two observers, BC and PB, gave r^2 values of 0.642 and 0.778, respectively. For strength judgments, linear mixtures of mean value and color-saturation gradient provided acceptable fits to the data (adjusted $r^2 > 0.7$) for all five observers viewing rendered tea and for five out of eight observers viewing real tea. The remaining three observers—CM, DT, and HW—gave r^2 values of 0.417, 0.571, and 0.413, respectively. For all but two observers, both parameters contributed significantly to the fit ($p < 0.005$), with individual differences in performance consistent with differential weighting of the two cues. For CM the fit was poor and the mean-value parameter did not reach significance, and for AC the fit was very good with only mean value and no significant contribution from saturation gradient. Full details of the fits are provided in Supplementary Tables S6–S9.

Discussion

The real-tea task draws on observers' everyday experience and asks them to make judgments about concrete properties that they clearly understand rather than about abstractions—for example, physical parameters such as scattering or absorption coefficients, or perceptual qualities such as translucence and transparency. Nevertheless, when we ask someone to judge milkiness from an image of milky tea, we are presenting them with a complex unmixing task. The observers have information only about the distributions of brightness and color within an image, which are insufficient to recover the light-transport properties affected by milkiness and tea strength, and yet we show empirically that they are able to make judgments that are reliable and effective. Their judgments are dominated by the property, be it milkiness or tea strength, that they are asked to estimate. The additive models we derived from MLCM show that although there are contributions to their judgments from the factor to be ignored, those contributions are relatively small (on average about 15% for the milkiness task and about 30% for the strength task).

In the second stage of our study we used rendered stimuli to investigate the mechanisms underlying these judgments. Perceptual unmixing of rendered tea was less successful than the unmixing of the real tea images. For real tea, judgments departed from the true concentrations only by a small additive contribution from the distracting variable, whereas judgments with rendered tea typically required modeling full interactions between the two physical variables. Observers were unable to effectively unmix the rendered images of tea despite the fact that one might regard differences within the set of rendered images as being less complex than those within the set of real images—in the rendered tea, scattering and absorption parameters are manipulated independently as one moves between steps in the tea-space, in contrast to the interactions in the effects of milk and tea on scattering and absorption. Since the contributions of the physical parameters to perceptual judgments are less complex for real tea than for the simplified renderings, this suggests that observers were making judgments based on characteristics that encapsulated the real interactions of the material rather than making simplified estimates of the light scatter or absorption.

There are individual differences evident in the MLCM analyses of both experiments. It is not surprising that we find these individual differences. If a trial is ambiguous (as indeed many were, as the tasks were designed to challenge observers) and observers cannot find consistent cues, then differences will be inevitable as they try out different strategies to make judgments about difficult comparisons. However, over and above these individual differences, there are global differences between conditions, particularly in the models required to account for the data. With rendered stimuli, saturated models were more likely to be required, whereas additive models were sufficient to account for judgments with real stimuli.

There were clear differences in the appearance of the sets of real and rendered images. The real stimuli are necessarily a less homogeneous set than the computer-generated set. Nevertheless, the two stimulus sets were spanned by similar ranges of perceptual discriminability, as measured by the d' measure in the preliminary MLDS task and shown in the range of extracted perceptual estimates from MLCM (Figures 2 and 3). The number of trials differed between experiments; however, when we reanalyzed the data, subsampling the same number of trials for each, there were no meaningful changes to our findings.

In the third stage of our study, we made use of the differences between the images of the real and rendered stimuli. Specifically, we tested whether it is possible for the same statistical pseudocue(s) to predict the performance of observers in the two experiments, which have quite different outcomes. We are not attempting to find a pseudocue that is invariant under changes in viewing conditions—we are simply aiming to establish whether the use of pseudocues could provide a plausible mechanism underlying performance in this study.

We found that a weighted sum of the same two statistical pseudocues was the best predictor of perceptual estimates of tea strength for judgments of both real and rendered stimuli. Similarly, a different pseudocue was the best predictor of performance in the milkiness task with both types of stimulus. The similarities of statistical-model predictors between the real and rendered images suggest that there is a level of consistency in the mechanisms underlying observers' judgments of them. The common mechanisms these results imply are used across experiments may provide insights into observers' strategies for estimating a specific material property.

The majority of previous studies of translucence perception have been based on achromatic stimuli, so the role of color statistics has not been thoroughly tested. One notable exception is that the perception of wetness of a surface has been shown to rely on saturation: Image regions with a higher saturation appeared wetter, particularly for images with a wide color distribution (Sawayama, Adelson, & Nishida, 2017). The present data show that, in judgments of milkiness, mean saturation was a good predictor of performance. This is consistent with the underlying physics: Highly scattering liquids direct more of the incident, whitish illumination to the camera, desaturating the image. In the tea-strength judgment, mean value was a significant predictor of performance, with darker images consistent with higher absorption. However, mean value is also affected by milkiness, and strength affects mean saturation, so these image statistics alone are not separable indicators of strength and milkiness. Good fits to observer data for strength judgments were obtained only by including an additional pseudocue that included a spatial component. The spatial distribution of information within an image is well known to be vital for making judgments of material properties (Landy, 2007; Motoyoshi, 2010; Sawayama & Nishida, 2018). Here we find that a chromatic gradient of saturation, summarized by the space constant on a single exponential fit, is the common pseudocue whose contribution predicted behavior with real and rendered stimuli. In a purely absorbing material, with no scattering component, transmittance depends on the concentration of the absorbing species and path length of the light through the material. Scatter increases the effective path length (Ben-David, 1995, 1997). For the tea images, the saturation gradient from the top surface of the liquid provides a measure of light penetration into the volume and selective absorption of some wavelengths, and it shortens with absorption and with scatter for rendered tea and with strength and with milkiness for real tea. Importantly, although the saturation gradient varies with both physical parameters, it does so differently from mean value. So a weighted sum can be used to obtain separable estimates of strength and milkiness, as indicated in the behavioral data. Further exploration of spatially specific statistics might yield even more effective pseudocues; one would want to design stimuli that powerfully discriminate between different spatially specific statistics.

Our conjecture that observers' judgments are based on statistical pseudocues (and not on direct estimates of the physical components of light transport) makes specific predictions about differences in performance between experiments with the real and rendered stimuli. When observers use a pseudocue that is optimized for discriminating properties of natural materials with stimuli whose properties are unlike those natural materials, we would predict that their judgments will err. In our study we found that when observers made judgments about rendered stimuli, there were clear deviations from monotonicity, meaning that as the concentration of milk increases, the judgment of milkiness may decrease. This maladaptation to rendered stimuli is presumably inherited from the observers' experiences with natural materials.

Our finding that it is entirely possible for simple perceptual heuristics to produce very different responses for what appear to be very similar stimuli opens the way to empirical studies of perception of material properties with complex underlying physical determinants. Asking observers questions about those physical factors would be all but impossible. However, when we constructed artificial stimuli broadly resembling tea, yet in which we could control aspects of light transport, we could still ask concrete questions about natural material properties. Our results suggest that a single strategy describes performance in both experiments. Taking a similar approach in which comparisons are made between real stimuli and rendered stimuli resembling them, but in which physical properties can be rigorously controlled, might reveal whether other perceptual properties are truly unitary and are also estimated using an individual heuristic.

The discovery of image statistics that can approximate real observers' responses in both of the experiments illustrates that there may well be a shortcut to achieve modeling of a material substance. This shortcut takes the more complex interactions of the volume into account without necessarily estimating the physical variables of light transport, since it is possible to find pseudocues that implicitly capture the effects that the physical dimensions have on the perceptual dimensions. Color signals may be important for judging more than surface spectral reflectance, since gradients of color can provide summary information about the complex optics of light penetration into a volume.

Conclusion

These results have implications for the fundamentals of translucence perception. We can of course perceive translucence; it is clear, however, that we do not independently perceive the physical determinants of the composition of light reaching the eye, in that we do not perceive translucent substances as simple linear mixtures of the scatter and absorption properties of the constituents. Our results support the suggestion that observers are employing pseudocues without requiring knowledge of physical light-transport factors and without using those physical properties as an intermediate step in material perception. Using real and rendered stimuli allowed us to generate conditions that make explicit, separable predictions for direct estimation of light transport versus reliance on pseudocues, and so test this hypothesis. Pseudocues that are derived from properties of specific real materials are capable of accounting for complex physical interactions within those materials, and are therefore perhaps more useful than estimations of the physical properties of light transport when determining the best proxies to use in making perceptual decisions about objects in the real world.

More generally, these results provide evidence in favor of a general theory of vision which denies that we see the physical determinants of volumetric or surface properties. The visual system does not "know" the physical laws of light scattering and absorption, and does not perform calculations of inverse optics to estimate these properties.

Keywords: translucence, material, perception

Acknowledgments

Informed consent was obtained for each observer in accordance with regulations by the Ethics Committees of the Department of Experimental Psychology, University of Oxford, and the Department of Psychology, Durham University. HES, RWK, and ACC designed the research; ACC and GC performed it; and all authors contributed to data analysis and drafts of the article. We declare we have no competing interests. ACC was supported by a studentship funded by Procter and Gamble; HES was supported by the Wellcome Trust (094595/Z/10/Z).

*HES and RWK are co–senior authors.

Commercial relationships: none.
Corresponding author: Alice C. Chadwick.
Email: alice.c.chadwick@gmail.com.
Address: Department of Psychology, University of Durham, Durham University Science Site, Durham, UK.

Footnotes

[1] A nonanalytic method of approximation.

[2] *Cues* will refer to image properties that covary with object properties, whereas *pseudocues* or *perceptual cues* will refer to their perceptual correlates.

References

Adelson E. H., & Anandan, P. (1990). Ordinal characteristics of transparency. In P. Kahn (Ed.), *AAAI Workshop on Qualitative Vision, Boston MA* (pp. 77–81). San Mateo, CA: Morgan Kaufmann.

Aernouts, B., Van Beers, R., Watté, R., Huybrechts, T., Lammertyn, J., & Saeys, W. (2015) Visible and near-infrared bulk optical properties of raw milk. *Journal of Dairy Science*, *98*(10), 6727–6738.

Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology*, *21*(24), R978–R983.

Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics*, *35*(5), 407–422.

Ben-David, A. (1995). Multiple-scattering transmission and an effective average photon path length of a plane-parallel beam in a homogeneous medium. *Applied Optics*, *34*(15), 2802–2810.

Ben-David, A. (1997). Multiple-scattering effects on differential absorption for the transmission of a plane-parallel beam in a homogeneous medium. *Applied Optics*, *36*(6), 1386–1398.

Blender: A 3D modelling and rendering package [Computer software]. (2013). Retrieved from http://www.blender.org

Chadwick, A. C., & Kentridge, R. W. (2015). The perception of gloss: A review. *Vision Research*, *109*, 221–235.

Debevec, P. E., & Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. *Proceedings of the 24th annual Conference on Computer Graphics and Interactive Techniques* (pp. 369–378). New York, NY: ACM.

Fleming, R. W., & Bülthoff, H. H. (2005). Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception*, *2*(3), 346–382.

Fleming, R. W., Jäkel, F., & Maloney, L. T. (2011). Visual perception of thick transparent materials. *Psychological Science*, *22*(6), 812–820.

Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, *4*(9):10, 798–820, https://doi.org/10.1167/4.9.10. [PubMed] [Article]

Gkioulekas, I., Xiao, B., Zhao, S., Adelson, E. H., Zickler, T., & Bala, K. (2013). Understanding the role of phase function in translucent appearance. *ACM Transactions on Graphics*, *32*(5), 1–19.

Hansmann-Roth, S., & Mamassian, P. (2017). A glossy simultaneous contrast: Conjoint measurements of gloss and lightness. *i-Perception*, *8*(1), 1–16.

Hasni, I., Bourassa, P., Hamdani, S., Samson, G., Carpentier, R., & Tajmir-Riahi, H. (2011). Interaction of milk α- and β-caseins with tea polyphenols. *Food Chemistry*, *126*(2), 630–639.

Jensen, H. W., Marschner, S. R., Levoy, M., & Hanrahan, P. (2001). A practical model for subsurface light transport. *Proceedings of SIGGRAPH 2001, Annual Conference Series* (pp. 511–518). New York, NY: ACM.

Knoblauch, K., & Maloney, L. T. (2008). MLDS: Maximum likelihood difference scaling in R. *Journal of Statistical Software*, *25*, 1–26, http://www.jstatsoft.org/v25/i02/.

Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. New York: Springer.

Knoblauch, K., & Maloney, L. T. (2014). MLCM: Maximum likelihood conjoint measurement, R package version 0.4.1 [Computer software]. Retrieved from https://CRAN.R-project.org/package=MLCM

Koenderink, J. J., & van Doorn, A. J. (2001). Shading in the case of translucent objects. *Proceedings of SPIE*, *4299*, 312–320.

Landy, M. (2007, May 10). Visual perception: A gloss on surface properties. *Nature*, *447*, 158–159.

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, *3*(1), 59–66.

Mood, A., Graybill, F., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

Motoyoshi, I. (2010). Highlight–shading relationship as a cue for the perception of translucent and transparent materials. *Journal of Vision*, *10*(9):6, 1–11, https://doi.org/10.1167/10.9.6. [PubMed] [Article]

Narasimhan, S. G., Gupta, M., Donner, C., Ramamoorthi, R., Nayar, S. K., & Jensen, H. W. (2006). Acquiring scattering properties of participating media by dilution. *ACM Transactions on Graphics*, *25*(3), 1003–1012.

Pharr, M., & Humphreys, G. (2004). *Physically based rendering: From theory to implementation*. Burlington, MA: Morgan Kaufmann.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software]. Retrieved from https://www.R-project.org/

Rogers, M., Knoblauch, K., & Franklin, A. (2016). Maximum likelihood conjoint measurement of lightness and chroma. *Journal of the Optical Society of America A*, *33*(3), A184–A193.

Sawayama, M., Adelson, E. H., & Nishida, S. Y. (2017). Visual wetness perception based on image color statistics. *Journal of Vision*, *17*(5):7, 1–24, https://doi.org/10.1167/17.5.7. [PubMed] [Article]

Sawayama, M., & Nishida, S. (2018). Material and shape perception based on two types of intensity gradient information. *PLoS Computational Biology*, *14*(4), e1006061, https://doi.org/10.1371/journal.pcbi.1006061.

Sharan, L., Rosenholtz, R., & Adelson, E. (2009). Material perception: What can you see in a brief glance? *Journal of Vision*, *9*(8): 784, https://doi.org/10.1167/9.8.784. [Abstract]

Singh, M., & Anderson, B. L. (2002a). Perceptual assignment of opacity to translucent surfaces: The role of image blur. *Perception*, *31*(5), 531–552.

Singh, M., & Anderson, B. L. (2002b). Toward a perceptual theory of transparency. *Psychological Review*, *109*(3), 492–519.

Singh, M., & Hoffman, D. D. (1998). Part boundaries alter the perception of transparency. *Psychological Science*, *9*(5), 370–378.

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1458), 1329–1346.

Xiao, B., Walter, B., Gkioulekas, I., Zickler, T., Adelson, E., & Bala, K. (2014). Looking against the light: How perception of translucency depends on lighting direction. *Journal of Vision*, *14*(3):17, 1–22, https://doi.org/10.1167/14.3.17. [PubMed] [Article]