The perceptual representation of transparency, lightness, and gloss
Barton L. Anderson
School of Psychology, The University of Sydney, Australia

To appear in: Oxford Handbook of Perceptual Organization
Oxford University Press
Edited by Johan Wagemans

1. Theoretical preliminaries

The adaptive role of vision is to provide information about the behaviorally relevant properties of our visual environment. Our evolutionary success relies on recovering sufficient information about the world to fulfill our biological and reproductive needs while avoiding environmental dangers. The attempt to understand vision as a collection of adaptations to specific computational problems has shaped a growing body of research that treats vision as a decomposable collection of “recovery” problems. In this view, perceptual outputs are understood as approximately ideal solutions to specific recovery problems, which have been dubbed the ‘natural tasks’ of vision (Geisler & Ringach, 2009). From this perspective, the science of understanding visual processing proceeds by identifying an organism’s natural tasks, evaluating the information available to perform each task, developing models of how to perform a task optimally, and discovering the mechanisms that implement these solutions.

The first aspect of this method of approach – the identification of ‘natural tasks’ – is arguably the most important because it defines the problem that needs to be solved. It is also the least constrained. Any environmental property can be hypothesized to be something that could have adaptive value and therefore something that might provide a selective advantage to anyone equipped to recover it. Presumably, however, only some aspects of our environment were involved in directly shaping the evolution of our senses. The scientific challenge is to differentiate properties that actually exerted selective pressure in shaping the design of our senses from those that merely came along for the ‘evolutionary ride’ (perceptual ‘spandrels’). But there is currently no principled means of making such distinctions. For example, a general argument could be (and has been) made that the computation of surface lightness would be useful because it provides information about an intrinsic property of the external world, but it is much harder to fashion a clear argument about how the recovery of surface albedo provides a specific adaptive benefit, or that any such benefit played a role in natural selection.

The second aspect of the adaptationist approach – identifying the information available for a computation – is in principle more constrained. Natural scenes are replete with information that could be used to sense a particular world property. Once a recovery problem has been identified, it is possible to inventory the sources of information that exist in the natural world that can be used to sense it. However, most recovery problems in vision (such as shape, depth, color,
lightness, etc.) are considered in isolation, often in informationally impoverished laboratory settings. This approach has led to the nearly universal acceptance of a belief in the poverty of the stimulus: the presumption that the images do not contain sufficient information to recover the aspects of the world that we experience. This view is typically defended by demonstrating that it is impossible to derive a unique solution for a specific recovery problem based on the information available in the images. Perception is construed as the outputs of a collection of under-constrained problems of probabilistic inference, which are solved with the aid of additional information, assumptions, or constraints. So construed, it is natural to turn to probability theory for guidance into how to solve such inference problems ideally, which typically entails the application of Bayes’ theorem (see Feldman’s chapter, this volume).

The third aspect of the adaptationist program is ostensibly the easiest, and is where theory meets data. Percepts or perceptual performance of observers is compared to that of the Bayesian ideal, constructed on a set of priors and likelihoods. When data and the Bayesian ideal are deemed sufficiently similar, the explanatory circle is considered closed: the fit between model and data is upheld as evidential support for the specification of the natural tasks, the selection of priors and likelihoods needed to perform the inference, and the claim that perception instantiates a form of Bayesian inference. All that remains is the discovery of the mechanisms that instantiate such computations.

The preceding describes what may currently be considered one (if not the) dominant view on how to approach the study and modeling of visual processes. My own view departs in a number of significant ways from this approach, which shapes both my selection of problems and the theoretical approach taken to account for data. One of the main goals of this chapter is to provide an overview of how my approach has shaped work in three areas of surface and material perception: transparency, lightness, and gloss. The gist of my approach may be articulated as follows. First, I assume that the attempt to identify the ‘natural tasks’ of vision – i.e., the computational ‘problems’ that visual systems putatively evolved to solve – is at best a guessing game, and at worst a theoretical fiction. Some of the ‘problems’ our visual systems seem to solve may be epiphenomenal outputs, not explicit adaptations. Second, the claim that vision is an ill-posed inference problem is a logical consequence of treating vision as a collection of recovery problems, for which it can be shown that there is no closed form solution that can be derived from the information that is currently available. But if the putative ‘recovery problem’ is misidentified, or the ‘information available for solving it’ is artificially restricted (such as typically occurs in laboratory environments), then it may not be vision that is ill-posed, but our particular understanding of visual processing that is misconstrued.

An alternative approach is to begin with what we visually experience about the world, and attempt to determine what image properties modulate these experiences. The question is not whether there is sufficient information in the images to specify the true states of the world, but rather, whether there is sufficient information to explain what we experience about the world. This approach is neutral as to the “computational goals” of the visual system, or if even whether the idea of a computational goal has any real meaning for biological systems. Whereas the recovery of a world property can be shown to be under-constrained by argument, the question
whether there is sufficient information available to explain what we experience about the world is an empirical question.

2. Disentangling images into causal sources

We experience the world as a collection of 3D objects, surfaces, and materials that possess a variety of different phenomenological qualities. The reflectance and transmittance properties of a material, together with its 3D geometry, structure light in ways that modulates our experience of shape, lightness, color, gloss, texture, and translucency. Some image structure also arises from the idiosyncratic distribution of light sources in a scene – the illumination field. To a first approximation, this list of surface and material properties tend to be experienced as separate sources of image structure, despite the fact that they are conflated in the image. Much research into perceptual organization has focused on how the visual system fills-in missing information or groups image fragments into a global structure or pattern. While such phenomena are an extremely important aspect of our visual experience, one of the other fundamental organizational problems involves understanding how the visual system disentangles different sources of image structure into the distinct surface and material qualities that we experience. In what follows, I consider a variety of segmentation problems in the perception of surface and material attributes, and the insights that such problems shed on the broader theoretical issues raised above.

2.1. Transparency

One of the perceptually most explicit and theoretically challenging forms of image segmentation occurs in the perception of transparency. Historically, the study of transparency focused on achromatic surfaces, which was largely due the seminal influence of Metelli’s model of transparency (Metelli, 1970, 1974a, 1974b, 1985)(see Gerbino’s chapter, this volume). The perception of (achromatic) transparent surfaces generates two distinct impressions: its perceived lightness and its perceived opacity or ‘hiding power.’ Metelli’s model was based on a simple physical device known as an episcotister: a rapidly rotating disc with a missing sector. The proportion of the disk that is ‘missing’ determines the amount of light transmitted from the underlying surfaces through the episcotister blades, which is the physical correlate of a transparent surface’s transmittance. The lightness (or albedo) of the transparent surface corresponded to the color of the paint used on the front surface of the episcotister, which determines the color of the transparent layer (or for achromatic paints, its lightness). Metelli’s model was restricted to “balanced” transparency, which referred to conditions where the episcotister had a uniform reflectance and transmittance, reducing each to a single scalar (number). For the simple bipartite fields Metelli used as backgrounds, this allowed equations for the total reflected light in the regions of overlay to be written as a sum of two components: a multiplicative transmittance term, which determined the weight for the contribution of the underlying surface; and an additive term, which corresponds the light reflected by the episcotister surface. By construction, Metelli considered displays containing two uniformly colored background regions, which gave him a system of two equations and two unknowns that could be solved in closed form. A significant body of work showed that the perception of
transparency is often well predicted by Metelli’s episcotister model: balanced transparency is perceived when displays were consistent with the episcotister equations, but generally not otherwise. Note that Metelli’s model served double duty as both a physical model of transparency and a psychological model of the conditions that elicit percepts of transparency.

Despite these successes, Metelli himself noted a curious discrepancy between the predictions of the episcotister model and perception: A light episcotister looks less transmissive than dark episcotister (Metelli, 1974a). From a “recovery” point of view, this constitutes a perceptual error, and hence non-ideal performance, but almost no experimental work was conducted to understand this deviation from the predictions of Metelli’s model. We therefore performed a series of experiments to test whether the physical independence of opacity and lightness is observed psychophysically (Singh & Anderson, 2002). Observers matched the transmittance of simulated surfaces that varied in lightness, and the lightness of transparent filters that varied in transmittance. We found that lightness judgments were modulated by simulated transmittance, and transmittance judgments were modulated by simulated variations in lightness. Thus, although the transmittance and reflectance of transparent layers are physically independent parameters in Metelli’s model, they are not experienced as being independent perceptually.

What theoretical conclusions can be drawn from these results? Metelli’s model treated a physical model of transparency as a perceptual model of transparency. Our findings of mutual ‘contamination’ of the transmittance and lightness of the transparent filter implies one of two possibilities: (1) there is no simple correspondence between the dimensions of a physical model and a perceptual model, or (2) that Metelli’s model is the wrong physical model on which to base theories of perceived transparency. With respect to (1), Metelli’s model equates the perceived opacity of an episcotister with its physical transmittance, and hence cannot explain why light episcotisters look more opaque than dark episcotisters. The dependence of perceived opacity on lightness can be readily understood, however, if the visual system relied on image contrast to assess the hiding power of transparent surfaces. A light episcotister reduces the contrast of underlying surface structure more than an otherwise identical dark episcotister, and hence, should appear more opaque if the visual system uses image contrast to assess perceived opacity. Indeed, it seems almost inevitable that the visual system utilizes contrast to judge the perceived opacity of transparent filters, since contrast determines the visibility of image structure in general. But this implies that the visual system is using the ‘wrong’ image properties to generate our experience of a world property, and hence will almost always result in the ‘wrong’ answer. From the perspective of explaining our experience, such issues are largely irrelevant; the only issue is whether there is sufficient information in the image to explain what it is we experience about the world, not whether such percepts are veridical.

---

1 This reduction in contrast occurs for almost any definition of contrast, which includes a divisive normalization term that is a function of integrated or mean luminance in the region over which contrast is defined. Unfortunately, there is currently no general definition of contrast that adequately captures perceived contrast in arbitrary images, so the precise way in which contrast is reduced depends on the definition of contrast used in a particular context.
Alternatively, it could be (and has been) argued that the discrepancy between perception and Metelli’s model merely provides evidence that there is something wrong with Metelli’s model, and does not impact on the more general claim that perception can be identified with the recovery of some physical model. Faul and Ekroll (2011) have made precisely this argument. They contend that a subtractive filter model better captures the perception of chromatic transparency, and hence may be a more appropriate model of achromatic transparency as well. Although there is currently insufficient data to determine which of these alternatives is ultimately correct for achromatic stimuli, Faul and Ekroll reported substantial discrepancies between their filter model and perceived transparency when the chromatic content of the illuminant was varied, despite demonstrating that there was theoretically sufficient information for a much better level of performance (Faul & Ekroll, 2012). At this juncture, there is currently no physical model that maps directly onto our experience of transparent surfaces, and it is largely a matter of scientific faith that such a model may ultimately be discovered.

2.2. Lightness
The perception of lightness also has been treated as a kind of segmentation problem. For achromatic surfaces, the term lightness (or albedo) refers to a surface’s diffuse reflectance. The light returned to the eye is a conflated mixture of the illuminant, surface reflectance, and 3D pose. There is currently extensive debate over the computations, mechanisms, and/or assumptions that are responsible for generating our experience of lightness (see Gilchrist’s chapter, this volume). There are four general theoretical approaches to the problem of lightness: scission (or layers models), equivalent illuminant models, anchoring models, and filter or filling-in models. I consider each model class in turn.

2.2.1. Models and theories of lightness

Scission models
Scission models assert that the visual system derives lightness by explicitly segmenting the illuminant from surface reflectance in a manner analogous to the decomposition that occurs in conditions of transparency. Such models have been dubbed layers, scission, or intrinsic image models (Adelson, 1999; B. L. Anderson, 1997; B. L. Anderson & Winawer, 2005, 2008; Barrow, Tenenbaum, Hanson, & Riseman, 1978; A. L. Gilchrist, 1979). In models of lightness, scission models assert that the visual system teases apart the contributions of reflectance, the illuminant, and 3D pose. Although some authors associate scission (or intrinsic image) models with veridical perception (Gilchrist et al., 1999), there is nothing inherent in scission models that mandates this association. The concept of scission entails a claim about a particular representational format or process of image decomposition that is presumed to underlie our experience of lightness. The hypothesized segmentation processes responsible for generating the putative layered representation may or may not result in veridical lightness percepts depending on how (and how well) the visual system performs the hypothesized decomposition.

Equivalent illumination
One model that is conceptually related to layers models is the equivalent illumination model (EIM) developed by Brainard and Maloney (2011). As with layers models, the EIM assumes that the visual system recovers surface reflectance by factoring the image into two components: an
estimate of the illuminant (which they term an “equivalent illuminant”) and surface reflectance. Whereas layers models have assumed that there is an explicit representation of both the illuminant and surface reflectance, the same is not necessarily true for the EIM. The EIM is a two-stage model which asserts that the visual system begins by generating an estimate of the illuminant, and uses this information in a second stage to derive surface reflectance properties from the image data. This model remains mute as to how the visual system estimates the parameters of the estimated illuminant from images and also remains uncommitted as to the any representational format the EI may take. The main experimentally assessable claim is that it predicts that the parametric structure of color or lightness matches can be described by some EIM. The approach of the EIM can be understood as follows: Given a set of reflectance matches, is it possible to find a model of the illuminant that is consistent with the matches? Note that there is no presumption that the particular EIM that putatively shapes observer’s matches is veridical; the only claim is that observers’ lightness matches are shaped by some EIM. Indeed, the benefit of this class of model is that it can in principle account for both veridical matches and/or the specific pattern of failures in veridicality.

**Anchoring theory**
A third theoretical approach to lightness is captured by anchoring theory, which was developed in an attempt to account for a variety of systematic errors in the perception of lightness (A. Gilchrist et al., 1999). Unlike layers or EIM models, there is no explicit factorization of the illuminant and reflectance in anchoring theory. Rather, anchoring theory asserts that perceived lightness is derived through a set of heuristic rules that the visual system uses to map luminance onto perceived lightness. There are two main components to anchoring theory (see Gilchrist’s chapter, this volume). First, following Wallach (1948), luminance ratios are used to derive information about relative lightness. When the full 30:1 range of physically realizable reflectances are present in a common illuminant, the true reflectance of surfaces can be derived on the basis of these ratios alone. However, in scenes containing less than this full 30:1 range, some additional information or rule is needed to transform ambiguous information about relative lightness into an estimate of absolute surface reflectance. For example, an image containing a 2:1 range of luminances could be generated by surfaces with reflectances of 3% and 6%, or 5% and 10%, 40% and 80, ad infinitum. Anchoring theory asserts that this ambiguity must be resolved with an anchoring rule, such that a specific relative image luminance (such as the highest) is mapped onto a fixed lightness value (such as white). All other lightness values in a scene are putatively derived by computing ratios relative to this anchor value. A number of fixed points are possible (e.g., the average luminance could be grey, the highest luminance could be white, or the lowest luminance could be black), but a variety of experiments, especially those from Gilchrist’s lab, have suggested that in many contexts, the highest luminance is perceived as white.

**Filtering and filling-in models**
A third approach to lightness treat lightness percepts as the outputs of local image filters applied directly to the images (Blakeslee & McCourt, 2004; Dakin & Bex, 2003; F. Kingdom & Moulden, 1988, 1992; Shapiro & Lu, 2011). Such approaches typically do not distinguish between perceived lightness (perceived surface reflectance) and brightness (perceived
luminance), at least not explicitly in the construction of the model. Rather, a new image is generated from a set of transformations applied to the input image. In a strict sense, filter models are not truly lightness models, since they simply transform one image into another image. Such models are more appropriately construed as models of brightness than lightness, since there is no explicit attempt to represent surface reflectance, or distinguish reflectance from luminance. Their relevance to understanding lightness depends on the extent to which the distinction between brightness and lightness makes biological or psychological sense for a given image or experimental procedure. Like anchoring models, filter approaches to lightness do not explicitly segment image luminance into separate components of reflectance and illumination.

In a related manner, a variety of filling-in models have been proposed that do not explicitly distinguish lightness and brightness (Grossberg & Mingolla, 1985; Paradiso & Nakayama, 1991; Rudd & Arrington, 2001). Such models invoke a two stage process: one that responds to the magnitude and orientation of ‘edges’ (oriented contrast) and/or gradients, and a second process that propagates information between such localized ‘edge’ responses to generate a fully “filled-in” or interpolated percept of brightness or color.

2.2.2. Evaluating theories of lightness
As noted in a recent article, the topic of lightness and brightness has historically been quite divisive (F. A. Kingdom, 2011). One source of disagreement involves the very distinction between brightness and lightness. Although such constructs are easily distinguished from each other with regard to their intended physical referents, it is not clear that (or when) such distinctions have psychological meaning. The distinction between lightness and brightness is particularly problematic for the kinds of displays that are typically studied in either lightness or brightness studies. In almost all cases, the targets of interest have a single, uniform luminance (or approximately so), and are embedded in highly simplified geometric and illumination contexts. For scenes depicting real or simulated surfaces, the surfaces of interest are typically flat, matte, and arranged in a single depth and/or illuminant. They typically lack information about the light field, such as that provided by specular reflections, 3D structure, shading, and inter-reflections. It is perhaps not surprising, then, that the field remains divided as to the proper way to understand how such impoverished displays are experienced, since it is unclear whether the distinction between lightness and brightness is psychologically meaningful in many of these displays. In what follows, I will consider some recent evidence relevant for each of the theories of lightness described above.

The core claim of scission models is that our experience of lightness involves the decomposition of the input into separable causes. One of the difficulties in assessing scission models is that it is not always clear whether (or when) such separation occurs, or what criteria that should be applied to determine whether such decomposition occurs. One can begin by posing a question of sufficiency: Can scission induce transformations in perceived lightness when it is phenomenally apparent? The most phenomenologically compelling sense of scission occurs in conditions of transparency, which requires the satisfaction of both geometric and photometric conditions. One technique for inducing scission involves manipulating the relative depth and photometric relationships of stereoscopic Kanizsa figures such as those depicted in Figure 1.
When the grey, wedge-shaped segments of the Kanizsa figure’s inducing elements in Figure 1 are decomposed into a transparent layer overlying a white disk (second and fourth rows of Figure 1), they appear substantially darker than when the same grey segment appears to overlie a dark disk (first and third rows of Figure 1). Note that the color of the underlying circular inducing element appears to be “removed” from the grey wedge-shaped segments and attributed to the more distant layer, which putatively transforms the perceived lightness of the transparent layer. Note also that the direction of the lightness transformation depends on which layer observers are asked to report. If observers are asked to report the color of the far layer underneath the grey sectors of the top image, they report it as appearing quite dark (nearly black), since this is the color of the interpolated disc. But if they are asked to report the near layer of the transparent region, they report it as appearing quite light.

Figure 1. Stereoscopic Kanizsa figure demonstrating the role of scission on perceived lightness for two different grey values. The small pie shaped inducing sectors are the same shade of dark grey in the top two rows, and the same shade of light grey in the bottom two rows. When the left two images are cross fused, or the right two images divergently fused, an illusory diamond is experience. Note that the diamonds in the 1st and 3rd row appear much lighter than their corresponding figures in the 2nd and 4th rows. Adapted from Anderson (1998, TICS).

In order to provide more conclusive evidence for the effects of scission on perceived lightness, I constructed stereoscopic variants of Figure 1 using random noise textures. The goal was to induce transparency in a texture such that the light and dark “components” of the texture
would perceptually segregate into different depth planes. An example is presented in Figure 2. When the left two columns are cross-fused, vivid percepts of inhomogeneous transparency can be observed: The top image appears as dark clouds overlying light disks, and the bottom appears as light clouds overlying dark disks. Note that the lightest components of the texture in the top image appear as portions of the underlying disc in plain view, whereas the same regions in the bottom image appear as the most opaque regions of the light clouds in the bottom image (and vice versa for the dark regions). We subsequently showed that similar phenomena could be observed in non-stereoscopic displays. In these images, scission was induced by embedding targets in surrounds that contain textures that selectively group with either the light or dark ‘components’ of the textures within the targets (Figure 3). As with their stereoscopic analogues, the white and black chess pieces are actually physically identical (i.e., contain identical patterns of texture). Note that the luminance variations within the texture of the chess piece figures are experienced as variations in the opacity of a transparent layer that overlie a uniformly colored surface. The opacity of the transparent surface is greatest for luminance values that most closely match the surround along the borders of the chess pieces (dark on top, light on the bottom), and the least opaque when for luminance values that are most different from the surround (light on top, dark on the bottom). Note that the lightest regions within the targets on the dark surround appear in plain view, and the darkest regions within the targets appear in plain view on the light surround. This bias is evident for essentially all ranges of target luminance tested, although this perceptual fact is in no way mandated by the physics of transparency, particularly for underlying surfaces that do not appear black or white.

Figure 2. Stereoscopic noise patterns can also be decomposed into layers in ways that induce large transformations in perceived lightness. If the left two images are cross fused or the right two images divergently fused, the top image appears to split into a pattern of dark clouds overlying light discs (top), or light clouds overlying dark disks (bottom). The textures in the top and bottom are physically identical. Adapted from Anderson (1999, Neuron).
Figure 3. Scission can also be induced by a selective grouping the light and dark components of texture of the targets (chess pieces) with the surround. The textures within the chess pieces in the top and bottom images are identical, but appear as dark cloud overlying light chess pieces on the top, and light clouds overlying dark chess pieces on the bottom. Adapted from Anderson & Winawer (2005, Nature).

These phenomena demonstrate that scission can induce striking transformations in perceived lightness in conditions of transparency, but it does not address the broader question of whether scission plays a role in generating our experience of lightness in conditions that do not generate explicit percepts of multiple layers or transparency.

EIMs also assert that the perception of surface color and lightness is derived by decomposing the image into estimates of the illuminant and surface reflectance. The evidence in support of this model is, however, phenomenologically indirect. Work from Brainard’s and Maloney’s labs have demonstrated that the parametric structure of a variety of matching data can be explained with a two-stage model in which the first stage involves an estimation of the illuminant (an “equivalent illuminant”), which is then used to derive observers’ reflectance matches from the input images (Brainard & Maloney, 2011).

Unlike scission models or EIMs, anchoring theory asserts that lightness is derived without explicitly decomposing the images into an explicit representation of illumination and reflectance. The central premise of anchoring theory is that the visual system solves the ambiguity of
lightness by treating a particular relative luminance as a fixed (anchor) point on the lightness scale (namely, that the highest luminance as white), independent of the level of illumination or absolute luminance values in a scene. To test this claim, we constructed both paper Mondrians displayed in a otherwise uniformly black laboratory, and simulated Mondrians displayed on a CRT in a dark black lab room (B.L. Anderson, de Silva, & Whitbread, 2008). In all cases, the highest luminance in the room was the central target patch of the Mondrian display. We varied both the reflectance range and illumination level of the former (i.e., paper Mondrians), and the simulated reflectance range and simulated illuminant levels of the latter simulated Mondrians. For restricted reflectance ranges (3:1 or less), we found that the highest luminance could vary in perceived lightness as a function of illumination. For our simulated illuminants and Mondrian displays, observers’ lightness matches (expressed as % reflectance) were a logarithmic function of (simulated) illuminant, rather than an invariant “white” as predicted by anchoring theory. These results suggest that the apparent “anchoring” of luminance to “white” is a consequence of the particular experimental conditions that have been used to assess this model, rather than reflecting an invariant “anchor point” used to scale other lightness values.

Some recent data has provided some strong evidence against an explicit illumination estimation model, and more generally, any most that relies on luminance ratios to compute perceived lightness (such as anchoring theory). Radonjić, Allred, Gilchrist, and Brainard (2011) conducted experiments depicting checkerboard displays in a display capable of displaying an extremely large dynamic range, and found that observers mapped a very high dynamic range (~10,000:1) onto an extended lightness range of 100:1, which spanned from “white” to “dark black” (the darkest values were obtained using glossy papers). Such behavior would not be expected for any model that attempts to infer a physically realizable illuminant, or any realizable reflectance ratios of real surfaces, as embraced by anchoring theory or the EIM.

One common assumption of anchoring theory and the EIM is that the visual system explicitly attempts to extract an estimate of lightness that corresponds to the physical dimension of surface albedo. The results of Radonjić et al. (2011) provide compelling evidence against this view. Just as our experience of transparency may not have any direct correspondence to the physical dimensions that modulate perceived transparency (such as transmittance), the perception of lightness may not represent an approximation of the physical dimension of surface albedo. The results of Radonjić et al. provide evidence that directly challenge any attempt to interpret the visual response as a “best guess” as to the environmental sources that produced their stimuli, since there is no combination of surface reflectance and illuminant that can produce such stimuli (at least in a common illuminant). I will return to this general point in the general discussion below.
3. Gloss

The experience of gloss is another aspect of our experience of surface reflectance that has received a growing amount of experimental attention. Whereas the concept of surface lightness has been cast as the problem of understanding how we experience the diffuse reflectance of a surface, the perception of gloss is typically cast as the problem of understanding how we experience the specular ‘component’ of reflectance. From a generative point of view, the diffuse and specular “components” of reflectance are treated as computationally separable. So construed, the problem of gloss perception involves understanding how the visual system segments the image structure generated by specular reflectance from diffuse reflectance (and all other sources of image structure).

The apparent intractability of this problem has inspired attempts to find computational shortcuts to avoid the complexity of this decomposition problem. One approach asserts that the visual system uses simple image statistics that do not require any explicit decomposition of the images into distinct components of reflectance to derive our experience of gloss. Motoyoshi, Nishida, Sharan, and Adelson (2007) argued that perceived gloss was well predicted by an image’s histogram or sub-band skew, a measure of the asymmetry of the pixel histogram (or response of center-surround filters) respectively. This claim was evaluated for a class of stucco surfaces with a statistically fixed level of surface relief that were viewed in fixed illumination field. In these conditions, glossy surfaces generated images with a strong positive skew, whereas matte surfaces generated surfaces with negative skew. The attractive feature of this kind of model is that it potentially reduces a complex mid-level vision problem into a comparatively simple problem of detecting low-level image properties.

However, subsequent work has shown that our experience of gloss cannot be understood so easily (B. L. Anderson & Kim, 2009; Kim & Anderson, 2010; Kim, Marlow, & Anderson, 2011; P. Marlow, Kim, & Anderson, 2011; Olkkonen & Brainard, 2010, 2011). One of the main problems with the proposed image statistics is that they fail to take into account the kind of image structure that predicts when gloss will or won’t be perceived. Specular highlights, and specular reflections more generally, must appear in the ‘right places’ on surfaces to elicit a percept of gloss (see Figure 4). From a physical perspective, specular highlights cling to regions of high surface curvature. The perception of gloss also requires highlights to appear in specific places and have orientations consistent with surface shading for a surface to appear glossy, a geometric constraint that is not captured by histogram or sub-band skew.
The perception of gloss depends critically on highlights appearing in the “right places” of a surface’s diffuse shading profile. In A, the highlight appears near the luminance maxima of the diffused shading profile and have similar orientations, and the surface appears relatively glossy. In B, the highlights have been rotated so that they appear with random positions and orientations relative to the diffuse shading profile, and do not appear glossy. Adapted from Anderson & Kim (2009, JOV).

Although these results suggest that the visual system in some sense “understands” the physics of specular reflection, there are other findings that reveal that the extent of any such understanding is limited. The perception of gloss has been shown to interact with a surface’s 3D shape and its lighting conditions, which are physically independent sources of image variability (Ho, Landy, & Maloney, 2008; P. J. Marlow, Kim, & Anderson, 2012; Olkkonen & Brainard, 2011). These interactions have been observed by a variety of authors and have resisted explanation. Indeed, these interactions are difficult to understand from a physical perspective, since gloss and 3D shape are independent sources of image structure. However, we recently presented evidence that these interactions can be understood as a consequence of a simple set of image cues that the visual system uses to generate our experience of gloss, which are only roughly correlated with a surface’s physical gloss level (P. J. Marlow et al., 2012). Some of the intuition shaping this theoretical proposal can be gained by considering the surfaces depicted in Figure 5. All of the surfaces in these images have the same physical gloss level, yet appear to vary appreciably in perceived gloss. Each column contains surfaces with a common degree of relief, and each row contains images that were placed in an illumination field with the same direction of the primary light sources. We varied the structure of the light field, the direction of the primary light sources, and 3D surface relief. Observers perform paired comparison judgments of the perceived gloss of all surfaces, where they chose which of a pair of surfaces was perceived as glossier. The data revealed complex interactions between the light field and surface shape on gloss judgments.
Figure 5. Interactions between 3D shape and perceived gloss as a function of the illumination field. All of the images in this image have the same physical gloss level, but do not appear equally glossy. The images in the top row were rendered in an illumination field where the primary light sources were oriented obliquely to the surface, and the images in the second row were illuminated in the same illumination field with the primary light sources oriented towards the surface. Adapted from Marlow, Kim, & Anderson (2012, Current Biology).

As can be seen in Figure 6, the variation of the illumination field and shape had a significant impact on the sharpness, size, and contrast of specular highlights in these images. We reasoned that if observers were basing their gloss judgments on these cues, then it should be possible to model observers’ gloss judgments with a weighted combination of these image cues. However, there is currently no known method for computing these cues directly from image. We therefore had independent sets of observers judge each of these cues, and tested whether it was possible to predict gloss judgments with a weighted sum of these cues. We found that a simple weighted sum model was capable of predicting over 94% of the variance of the other observers’ gloss judgments. Thus, although the perception of surfaces with the same physical gloss level can appear to vary significantly in perceived gloss, these effects can be understood with a set of relatively simple, albeit imperfect, “cues” that the visual system uses to generate our experience of gloss.
Figure 6. Data and model fits for the experiments we performed on the interactions between perceived gloss, 3D shape (as captured by a measure of surface relief), and the illumination field. The stimuli were viewed either with or without stereoscopic depth (the ‘disparity’ and ‘no disparity’ conditions respectively). The different colored curves in each graph correspond to a different illumination direction of a particular illumination field (called ‘Grace’). The gloss judgments are in the two top right panels. The panels on the left represent the judgments of a separate group of observers of four different cues to gloss: the depth, coverage, contrast, and sharpness of specular reflections. The panel labeled “skew” was computed directly from images. The dotted lines in the two graphs on the top right correspond to the best fitting linear combination of the cues on the left, which account for 94% of the variance of gloss judgments. The weights are denoted in the boxes adjacent to the small arrows in the center of the graphs. Adapted from Marlow, Kim, & Anderson (2012, Current Biology).
4. The perceptual organization of surfaces and materials

The last few decades have witnessed an explosive increase in models that have treated visual processes as a collection of approximately ideal “solutions” to particular computational problems. Such models are explicitly teleological: they treat a desired outcome, goal, or task as the organizing force that shapes the perceptual abilities they are attempting to model. Evolutionary theory serves as the engineering force that putatively drives biological systems toward optimal solutions. This modeling process hinges critically on the ability to specify the “natural tasks” that were putatively shaped by evolution. The justification for the adaptive importance of a particular “natural task” typically takes a generic form: an environmental property is treated as having evolutionary significance because it is an intrinsic property of the world. Thus, any animal capable of accurately recovering that property would gain an adaptive advantage. The properties to be recovered – the “tasks” of vision – are defined in with respect to particular physical sources of variability. Our experience of lightness is treated as the visual system’s solution to the problem of recovering the albedo of a surface. Our experience of transparency is treated as the perceptual solution to a particular generative model of transparency (such as Metelli’s episotister model or Faul and Ekroll’s filter model). And our experience of gloss is understood as the visual system’s attempt to estimate the specular component of surface reflectance.

One of the assumptions of this approach is that the dimensions of psychological variation are assumed to mirror the sources of physical variation. This assumption is explicit in both Metelli’s model, which treated the episotister as both a physical and psychological model of transparency, and the EIM of Brainard and Maloney, which asserts that the visual system generates a “virtual” model of the illuminant to recover color and lightness. The perception of gloss has also been studied as a kind of “constancy” problem, which involves recovering the specular “component” of reflectance.

A main theme of this chapter is to question the adequacy of this conceptualization of vision. Rather than attempting to guess the “natural tasks” and an animal, I view the goal of perceptual theory to discover the “natural decompositions” of representational space, i.e., to discover the psychological dimensions that capture the space of our experiences. The preceding focused on our experience of transparency, lightness, and gloss. Each of these attributes can be identified with a particular physical property of surfaces and material, which can be described in physical terms independently of any perceptual system. Such descriptions assume that the visual system plays no part in defining the attributes that it putatively represents; the dimensions are given by identifiable sources of variation in the world, which the visual system is attempting to recover, not by intrinsic properties of the visual system. We are left discussing how well the visual system encodes or recovers a particular world property, rather than how the visual system contributes to shaping the dimensions of our vision experience.

The preceding suggests that this general approach fails to explain a number of different phenomena in surface and material perception. The perception of surface opacity does not follow Metelli’s model of transmittance. We argued that one of the main reasons for this failure
was that Metelli’s model is based on a ratio of luminance differences, where are not available to a visual system that transforms retinal luminance into local contrast signals. We showed that our matching data were well predicted by a model in which observers matched contrast ratios, rather than luminance difference ratios. One of the key points of our model was to define transmittance in a way that was consistent with intrinsic coding properties of the visual system, even if this results in the failure to compute physically accurate measure of surface opacity. This general approach of a physiologically motivated model has also been pursued by a recent model of these results by Vladusich, who proposed an alternative model of our transmittance matching data (Vladusich, 2013). He shows that our transmittance matching data can be captured with a modified version of Metelli’s model in which log luminance values are used instead of luminance values (Vladusich, submitted). Like our model, the choice to use Log luminance values cannot be derived from the physics of transparent surfaces; they are derived from intrinsic response properties of the visual system.

The different theories of lightness perception are even more contentious and diverse than those found in the transparency literature. One of the basic issues involves the distinction between lightness and brightness. The perception of lightness is then defined as the perception of diffuse (achromatic) surface reflectance, whereas brightness is defined as the perception of image luminance. The presumption is that these physical distinctions have psychological meaning. But this is far from self-evident. The majority of work on lightness has used 2D (flat) matte displays of surfaces with uniform albedos, for which the distinction between lightness and brightness is arguably least valid (or meaningful) perceptually. For some experimental conditions, observers’ matching data will differ substantially if instructed to match either brightness or lightness. But in others, a difference in instructions may make little or no difference. Consider, e.g., the problem of matching the “brightness” versus the “lightness” of the checker-shadow illusion. A given patch appears a particular shade of grey, and there is no evidence that observers could distinguish its brightness and lightness. In support of this view, we found that the perception of lightness increased as a function of its luminance in both simulated and “real” Mondrian displays. Moreover, the data of Radonjić et al. (2011) demonstrate that observers will readily map a physically unrealized set of luminances, spanning 4 orders of magnitude, onto a lightness scale two orders smaller. These results are impossible to reconcile with models that treat the problem of lightness as a recovery problem, since the range of reflectances in a natural scene can only span a range of ~30:1.

In the perception of gloss, we found that observer’s experience of gloss can be well predicted by a set of simple cues that are only imperfectly correlated with the physical gloss of a surface. Gloss is not defined with respect to some physically specified dimension of surface optics, but with respect to a set of cues the visual system uses as a proxy for an objectively defined surface property.

What general understanding can be gleaned from these patterns of results? All of these results reveal the insufficiency of attempting to identify psychological dimensions of our experience with physical sources of image variability. The fact that we have a particular experience of lightness, gloss, or transparency does not imply that the dimensions of our experience map onto
a particular physical dimension and/or its parameterization. The general argument used to justify “natural tasks” takes the generic form that “getting an environmental property right increases adaptive fitness.” The presumed identification of fitness with veridical perception is actually fallacious (see Hoffman, 2009; cf. Lewontin, 1996), but even if such views were accepted, they are incapable of distinguishing perceptual abilities that were actually shaped by natural selection from the “spandrels” that came along for the evolutionary ride. The fact that human observers will readily map an ecologically unobtainable range of luminance values (in a single illuminant) onto lightness estimates suggests that lightness may be one example of a perceptual spandrel. Although human observers can usually distinguish reflectance differences from other sources of image variation, the perception of absolute lightness may simply be the result of low-level processes of adaptation that allow the visual system to encode a particular range of luminance values. Indeed, I am aware of no compelling evidence or argument about why lightness constancy per se provided an adaptive advantage, or is something that the visual system is explicitly “designed” to compute. A similar argument holds for the perception of transparency and gloss. We can readily distinguish between surfaces or media that transmit light from those that do not, or distinguish between surfaces that reflect light specularly from those that do not. But the data also suggests that we do not scale these dimensions in a way that is physically correct for any of these properties.

Although it is difficult to craft a compelling argument for the specific adaptive utility of developing a physically accurate model of lightness, gloss, and transparency, the fact that we experience these different sources of variable as different underlying causes implies that the visual system is capable of at least qualitatively distinguishing different sources of image structure. This “source segmentation” is arguably one of the most important general properties of our visual system. The visual system may, in fact, be quite poor in estimating lightness in arbitrary contexts, but it is nonetheless typically quite good at distinguishing image structure generated by lightness differences from illumination changes, or variations in the opacity of a transparent surface, or from specular reflections. The identification of specular reflections as specular reflections depends on their compatibility with diffuse surface shading and 3D surface geometry, and is modulated by the structure, intensity, and distribution of image structure so identified, even if it does not accurately capture the “true” gloss level of a surface. And although the physical transmittance (or opacity) of a surface does not vary as a function of its albedo or color, the psychological analog of opacity – its “hiding power” – will for a visual system that uses contrast to determine the visibility of image structure. The visual system may not determine the “true” opacity of a surface, but nonetheless is effective at performing a segmentation that captures the presence or absence of transmissive surfaces and media.
5. Summary and conclusions

In this chapter, I have considered a number of topics in the area of surface and material perception: transparency, lightness, and gloss. The organization of these topics was largely shaped by my historical progression in conducting research into each of these domains; many alternative organizations are possible. In all of these areas of inquiry, there has been a striking tendency to treat physical models of image formation as some kind of approximation to a perceptual model of their apprehension. The precise way that a physical model “counts” as a psychological model is typically left unspecified. It appears to be based on some intuition that the visual system “knows” or “understands” the physics that of a particular surface or material attribute. I contend that one of the main goals of vision science should be to discover the dimensions of perceptual experience, and the image variables that modulate our response to them. Whereas the dimensions of physical variables can be specified independently of any perceptual system, the dimensions of perceptual experience are inherently relational, and must consider the intrinsic properties of the visual system as well as the environments in which they operate.
6. References


