

# Perception-memory interactions reveal a computational strategy for perceptual constancy

**Maria Olkkonen**

Department of Psychology,  
Durham University, Durham, UK  
Institute of Behavioural Sciences,  
University of Helsinki, Helsinki, Finland



**Toni P. Saarela**

Department of Psychology, University of Pennsylvania,  
Philadelphia, PA, USA



**Sarah R. Allred**

Department of Psychology,  
Rutgers–The State University of New Jersey,  
Camden, NJ, USA



**A key challenge for the visual system is to extract constant object properties from incoming sensory information. This information is ambiguous because the same sensory signal can arise from many combinations of object properties and viewing conditions and noisy because of the variability in sensory encoding. The competing accounts for perceptual constancy of surface lightness fall into two classes of model: One derives lightness estimates from border contrasts, and another explicitly infers surface reflectance. To test these accounts, we combined a novel psychophysical task with probabilistic implementations of both models. Observers compared the lightness of two stimuli under a memory demand (a delay between the stimuli), a context change (different surround luminance), or both. Memory biased perceived lightness toward the mean of the whole stimulus ensemble. Context change caused the classical simultaneous lightness contrast effect, in which a target appears lighter against a dark surround and darker against a light surround. These effects were not independent: Combined memory load and context change elicited a bias smaller than predicted assuming an independent combination of biases. Both models explain the memory bias as an effect of prior expectations on perception. Both models also produce a context effect, but only the reflectance model correctly describes the magnitude. The reflectance model, finally, captures the memory-context interaction better than the contrast model, both qualitatively and quantitatively. We conclude that (a) lightness perception is more consistent with reflectance inference than contrast coding and (b) adding a memory demand to a perceptual task both renders it**

**more ecologically valid and helps adjudicate between competing models.**

## Introduction

In everyday life, we make use of sensory information constantly and effortlessly to guide our behavior. The seeming ease of perception belies its computational complexity, which arises because sensory signals are low-dimensional projections of an infinitely complex external world. A retinal representation of a visual scene, for instance, confounds the size and shape of objects with their distance and pose and the properties of surfaces with those of the ambient illumination. This uncertainty about the source of a sensory signal is compounded by the inherent variability in sensory encoding. A central goal of neural information processing is to transform these noisy and ambiguous sensory signals into behaviorally relevant representations of world properties, in other words, to achieve perceptual constancy. But even after decades of research, we are far from understanding the computational and neural underpinnings of perceptual constancy (e.g., DiCarlo, Zoccolan, & Rust, 2012; Kingdom, 2010). Our purpose here is to elucidate the computational strategies used by humans when estimating surface properties from ambiguous sensory input. We focus on the achromatic aspect of surface

Citation: Olkkonen, M., Saarela, T. P., & Allred, S. R. (2016). Perception-memory interactions reveal a computational strategy for perceptual constancy. *Journal of Vision*, 16(3):38, 1–21, doi:10.1167/16.3.38.

doi: 10.1167/16.3.38

Received April 27, 2015; published February 24, 2016.

ISSN 1534-7362



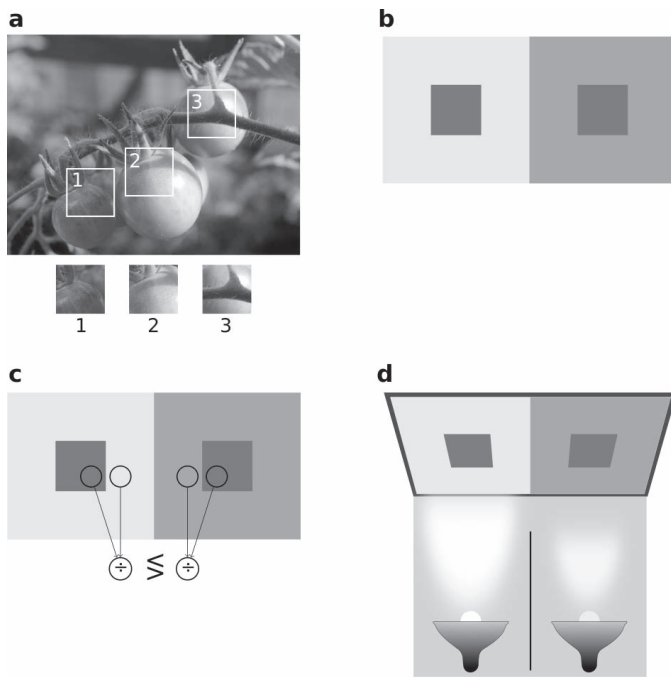


Figure 1. (a) The light reflected from surfaces varies across a scene both due to variations in surface reflectance (1, 3) and illumination gradients (2). (b) In the simultaneous lightness contrast illusion, the two central targets appear to differ in lightness even though the reflected light matches. The effect has been explained based on contrast matching (c) or on reflectance estimation (d). (c) The contrast hypothesis. The visual system matches the two targets when their edge contrasts are equal. Here, the edge contrasts are unequal even though the target luminances match. Thus, the squares appear different. (d) The reflectance estimation hypothesis. The visual system attributes the luminance change between the surrounds to an illumination change. Because the targets are equal in luminance, the reflectance of the targets must differ. Thus, the targets appear different in lightness (perceived reflectance).

reflectance, lightness, although the same principles are likely to hold for chromatic color perception as well.

Figure 1a illustrates the constancy problem for the perception of surface lightness. In natural scenes, luminance edges can be due to reflectance or shading differences. Separating the retinal light signal into these different underlying causes (sometimes termed *layers*) is intractable without additional constraints, derived from heuristics or prior information about the world (Adelson & Pentland, 1996; Kingdom, 2008; Zaidi, 1998). Evidently, extracting invariant, stable information from the light signal is necessary for the sensory information to be behaviorally useful. Although the early stages in the processing of light are relatively well understood, much less is known about the transformation of these early signals into invariant lightness percepts (e.g., Kingdom, 2010).

Several lightness illusions demonstrate the ambiguity experienced by the visual system about the causes of retinal light signals. In these illusions, physically identical stimuli are perceived differently depending on viewing context. A classical example is the simultaneous lightness contrast illusion, in which the same gray patch appears either dark or light depending on its surround (Figure 1b). Insofar as they reflect the attempts of the visual system to parse an ambiguous scene into invariant representations, illusions can shed light on the underlying computations. Indeed, several theories of lightness perception have been proposed based on human responses to various ambiguous displays (e.g., Allred & Brainard, 2013; Anderson & Winawer, 2008; Blakeslee & McCourt, 2012; Bloj et al., 2004; Brainard & Maloney, 2011; Gilchrist et al., 1999; Land & McCann, 1971; Maertens, Wichmann, & Shapley, 2015; Murray, 2013; Rudd & Zemach, 2005; Vladusich, 2012). The majority of these theories fall under one of two frameworks: In one, lightness is derived from operations on border contrasts at one or more spatial scales (e.g., Blakeslee & McCourt, 2012; Land & McCann, 1971; Maertens et al., 2015; Rudd & Zemach, 2005; Wallach, 1948); in another, lightness is directly estimated by segmenting the sensory signals into surfaces and illuminants based on prior constraints (e.g., Allred & Brainard, 2013; Murray, 2013; Purves, Williams, Nundy, & Lotto, 2004). Each theoretical framework can account for a subset of the experimental results, but there is no consensus as to which ultimately describes human performance better, and we are not aware of any direct comparisons between the two frameworks. To advance our understanding of the computations that underlie human lightness constancy, we use a novel psychophysical paradigm together with computational implementations of both frameworks, which we term *contrast* and *reflectance* models.

Contrast models emphasize the role of low-level processes such as ratio coding at luminance borders for perceived lightness; quantitative models in this framework have successfully accounted for many perceptual phenomena, such as the classical simultaneous lightness contrast illusion (Blakeslee & McCourt, 2012; Dakin & Bex, 2003; Kingdom & Moulden, 1992; Rudd & Zemach, 2005; Spehar, Debonet, & Zaidi, 1996). Although the models in this framework differ considerably in terms of implementation, all model outputs depend on local luminance ratios. The classical lightness contrast effect arises in these models from different luminance ratios at the target-surround border (Figure 1c); for the targets to match in perceived lightness, the physical edge contrasts need to match. Although this framework is attractive in its simplicity, it fails to explain some well-known lightness phenomena, such as the effect of spatial configuration on perceived lightness (e.g., Adelson, 1993; Anderson &

Winawer, 2008; Bloj & Hurlbert, 2002; Gilchrist, 1977; Hillis & Brainard, 2007b; Knill & Kersten, 1991; Purves, Shimpf, & Lotto, 1999; Schirillo, Reeves, & Arend, 1990).

The reflectance estimation approach, on the other hand, emphasizes inferential processes for estimating lightness from the incoming light signal (Adelson & Pentland, 1996; Allred & Brainard, 2013; Bloj et al., 2004; Purves et al., 1999, 2004; Schirillo & Shevell, 1997; von Helmholtz, 1867). This framework assumes that observers have knowledge about likely states of the world that they then use to constrain perceptual estimates. In the case of the classical simultaneous contrast illusion, observers assume that the luminance difference between the two sides of the display is due to a difference in illumination rather than to a difference in background material (Figure 1d). This assumed illumination difference causes the two targets to appear different in lightness, because the reflected light (luminance) matches. Although the reflectance estimation framework is powerful in explaining several lightness phenomena not amenable to contrast coding accounts (such as the effects of spatial configuration), few model implementations exist (but see Allred & Brainard, 2013; Bloj et al., 2004; Murray, 2013).

Finally, the existing model implementations in either framework can account for only purely *perceptual* effects. But real-world sensory estimation tasks commonly place demands on both *perceptual* and *memory* processing: To identify previously seen surfaces in a new scene (e.g., when looking for a lost item of clothing), one has to *both* discount illumination variation (such as shadows in Figure 1a) to estimate reflectance *and* employ working memory to compare percepts with the memorized surfaces. In the lightness perception literature, *perceptual* errors have been used to drive theory; the dependence of perceived lightness on depth relationships, for example, demonstrates that local contrast is not the only determinant of perceived lightness. A separate visual memory literature has revealed that working memory processing increases the sensory noise of a representation (Pasternak & Greenlee, 2005), leading to errors in memory-dependent estimation (Ashourian & Loewenstein, 2011; Jazayeri & Shadlen, 2010; Olkkonen, McCarthy, & Allred, 2014). Accumulating evidence from behavioral and neurophysiological studies shows further that the underlying neural processes for perception and working memory are closely related (Ester, Serences, & Awh, 2009; Harrison & Tong, 2009; Kang, Hong, Blake, & Woodman, 2011; Magnussen & Greenlee, 1999; Pearson & Brascamp, 2008; Serences, Ester, Vogel, & Awh, 2009; Supér, Spekreijse, & Lamme, 2001). Although a small number of studies have considered some aspects of memory in a color constancy task (Allen, Beilock, & Shevell, 2011; Jin & Shevell, 1996; Ling & Hurlbert,

2008; Uchikawa, Kuriki, & Tone, 1998), the independence of perceptual and memory demands in perceptual constancy has not been characterized.

Here we take advantage of the fact that contrast and reflectance models generate different predictions about the interaction between perceptual and memory effects. This allows us to adjudicate between competing frameworks by comparing human and model performance in a novel psychophysical task that factorially combines perceptual and memory demands. To anticipate, the reflectance model is consistent with the pattern of biases in human lightness matching when perceptual and memory demands are combined, whereas the contrast model is not. The results prove our experimental approach fruitful for differentiating between competing models of human performance in perceptual tasks across domains.

## Experiment: Lightness matching

Before implementing computational models of lightness perception, we sought to establish human lightness matching performance in a novel task that combines perceptual and memory demands in a  $2 \times 2$  factorial design. After reporting the results from psychophysics, we turn to model implementations and their comparison to human results.

## Methods

### Participants

Five Rutgers undergraduates participated after signing informed consent. Each participant ran all experimental conditions. Participants were compensated \$10/hr. Participants had normal or corrected-to-normal visual acuity and normal color vision as assessed with the Ishihara color plates. The experimental procedures were approved by the Rutgers University Institutional Review Board and adhered to the Declaration of Helsinki.

### Apparatus

Stimuli were displayed on a calibrated CRT monitor with a spatial resolution of  $1,024 \times 786$  pixels, 85-Hz refresh frequency, and a 16-bit luminance resolution via the DATAPixx box (VPixx Technologies, Inc., Saint-Bruno, QC, Canada). We calibrated the monitor once a month by measuring the output of the three monitor primaries with a Photo Research PR655 spectroradiometer and gamma-correcting the linear intensity values with standard methods (Brainard, Pelli, &



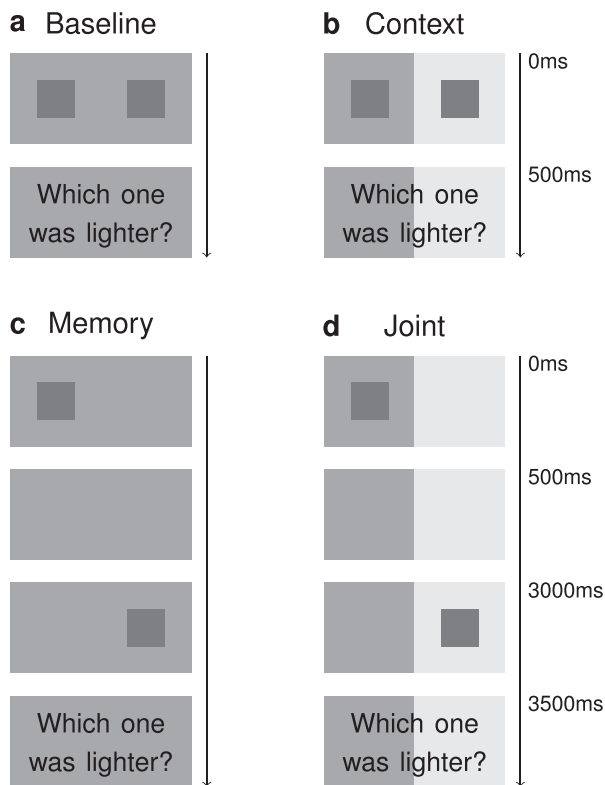


Figure 2. The experimental conditions. (a) In the baseline condition, both stimuli were displayed simultaneously on a surround that was either uniformly dark (shown) or light gray. (b) In the context condition, both stimuli were displayed simultaneously on a light and dark surround. (c) In the memory condition, the reference stimulus was displayed in the first interval either on the left (shown) or on the right side. The test stimulus was displayed after a 2.5-s delay on the opposite side. The surround was uniformly dark (shown) or light gray. (d) In the joint condition, the reference was displayed in the first interval either on the left (light surround; shown) or on the right (dark surround). Test was displayed on the other side after a 2.5-s delay. In each condition, observers were instructed to indicate which stimulus appeared lighter (see text for details).

Robson, 2002) before sending them to the DATAPixx box.

MGL functions (<http://justingardner.net/doku.php/mgl/download>) and custom software were used for stimulus display and data collection in Matlab (Mathworks, Inc., Natick, MA).

### Stimuli

We characterized lightness perception with a lightness comparison task based on the classical simultaneous contrast illusion (see Figure 1b). We modified the basic simultaneous contrast display to separately measure the effect of perceptual context (Figure 2b), memory (Figure 2c), and both together (Figure 2d) on

perceived lightness. To complete the factorial design, we included a baseline condition without context or memory demands (Figure 2a).

The two stimuli that observers compared on each trial, reference and test, were  $1.8^\circ$  square patches displayed on both sides of a central fixation cross at  $3^\circ$  eccentricity. We used three different reference luminances (4.4, 6.9, 10.9  $\text{cd/m}^2$ ), all of which were darker than the surrounds. We used only decrements because the factorial design constrained the number of references, and decrements have been found to be more prone to simultaneous contrast effects than increments (e.g., Arend & Spehar, 1993). The luminance of the test patch for a given reference was determined on each trial according to a staircase procedure, described below. The reference and test stimuli were displayed on different sides of the display on surrounds that were either the same (symmetric) or different (asymmetric) in luminance, depending on condition. The luminances for the dark and light surrounds were 11.7 and 18.5  $\text{cd/m}^2$ , respectively.

### Procedure

Observers viewed the display from a 94-cm distance controlled with a chin rest. Stimulus timing varied according to condition, but the general procedure was as follows. On each trial, observers saw the reference and test for 0.5 s (displayed either simultaneously or with a 2.5-s delay), after which they indicated with a key press which stimulus appeared lighter. “Lighter” was defined as the stimulus that appeared whiter on a continuum from black to white. The instructions did not contain any reference to surfaces or illuminants nor cues that might direct the observer to a given scene interpretation. The following trial started after a response and a 0.5-s intertrial interval.

Test luminance for a given reference on each trial was determined by an adaptive staircase procedure. Four interleaved staircases with different decision rules converged on different points of the psychometric function (roughly 12.5, 25, 75, and 87.5 percentiles). Using four decision rules allowed us to measure the proportion of “lighter” responses for a large range of test luminances and to fit a psychometric function to the data. We defined perceived lightness for each reference as the 50th percentile of the psychometric function (the point of subjective equality [PSE]). We also derived the precision of each estimate, or the discrimination threshold, as the difference between the 50th and the 75th percentile of the psychometric function (see below for more details on the fitting).

The four experimental conditions shown in Figure 2 were presented in blocks with counterbalanced order across observers. In conditions in which the two surrounds differed (context and joint), the darker

surround was displayed on the left. The left-right location of the reference was always randomized. Separate staircases controlled tests for each reference on either surround. With three references, two surrounds, four staircases, and 20 trials per staircase, this resulted in 240 trials for the symmetric surround conditions (baseline, memory) and 480 trials for the asymmetric surround conditions (constancy and joint). The blocks with symmetric surrounds had fewer trials because they were run separately for the dark and light surrounds to keep adaptation constant within a block.

Each observer ran two versions of the delay conditions (memory and joint). In the *blank* version, the delays between the reference and test were blank, only showing the fixation cross (these are shown in Figure 2b, d). In the *distractor* version, two distractor stimuli were displayed for 0.5 s during each delay period. The distractors had the same spatial dimensions and locations as the reference and test stimuli. Distractor luminances were selected from a Gaussian distribution approximately 1.5 just-noticeable-differences from a given reference stimulus in either direction. There was no task related to the distractors. We included the distractor conditions because we hypothesized that they might increase estimation noise and thus exacerbate any potential biases. However, analysis showed no significant differences between distractor and blank versions of the task (Supplementary Figure S1), and so we pooled the data across the two versions for the memory and joint conditions.

Each observer ran all six conditions twice with different block orders for the two repetitions. As some blocks were rather long (30–45 min), observers were encouraged to take short breaks between trials. In addition, observers were able to take longer breaks between blocks. The whole experiment took about ten 1-hr sessions per observer.

### Data analysis

*Derivation of appearance and precision:* Data from the two repetitions in each condition were pooled for analysis, and we fit psychometric functions (cumulative Gaussian) to the pooled data (Wichmann & Hill, 2001a, 2001b). Perceived lightness of the reference was defined as the PSE in each condition (i.e., the 50th percentile of the psychometric function). Precision was defined as the discrimination threshold (i.e., the difference between the 75th and 50th percentile of the psychometric function). Appearance bias (i.e., appearance shift from veridical) was defined as the difference between the PSE in a given condition and the veridical reference luminance.

*Independence analysis:* Effects of memory and of context on lightness matches can be characterized by comparing the PSE in each condition to the veridical

reference luminance: Both memory and context can bias lightness matches to a varying extent. If these effects on lightness appearance are independent, lightness appearance bias in the joint condition should be a fully additive combination of the separate memory and context biases. Let  $c = C(r)$  be a context transfer function, where  $r$  is the reference value and  $c$  the reference lightness as quantified by a PSE. This function describes the effect of context on perceived reference lightness. Similarly, the effect of memory on reference lightness can be described by a memory transfer function  $m = M(r)$ . If context and memory exert independent influences on the perceived lightness of a reference, the combined match should be given by the concatenation of the two transfer functions,  $cm = C(M(r))$ . In practice, we computed the predicted joint matches by taking the measured memory matches ( $m$ ) as new references and deriving context matches ( $cm$ ) to those by interpolating/extrapolating from the existing context matches ( $c$ ).

After deriving the predicted joint matches, we quantified the independence of memory and context effects in the measured joint matches with an additivity index ( $AI$ ):

$$AI = k(\text{bias}_{\text{obs}} - \text{bias}_{\text{pred}}), \quad (1)$$

where  $k = 1$  for a dark surround and  $k = -1$  for a light surround so that the context effect has the same sign for both surrounds, and  $\text{bias}_{\text{obs}}$  and  $\text{bias}_{\text{pred}}$  are the observed and predicted biases, respectively. An additivity index of zero indicates full additivity; negative values indicate subadditivity, and positive values superadditivity.

### Results

First, we present data from one representative observer, after which we show data in the aggregate. Example psychometric functions for one observer are shown in Figure 3a through c. These functions describe the proportion of “test lighter” responses for the whole range of test luminances, for one reference luminance and one experimental condition per panel. The effects of context and memory separately and jointly are shown in the three panels through a comparison with a baseline psychometric function. A horizontal shift of the experimental function relative to baseline indicates an effect on appearance (bias), whereas a change in slope indicates an effect on precision (threshold).

As expected, surround luminance affected the perceived lightness of the targets: In Figure 3a, the psychometric function for the context condition (green) is shifted rightward from the baseline function (black), indicating that the reference stimulus on a *dark* surround appeared *lighter* than the test stimulus on the light

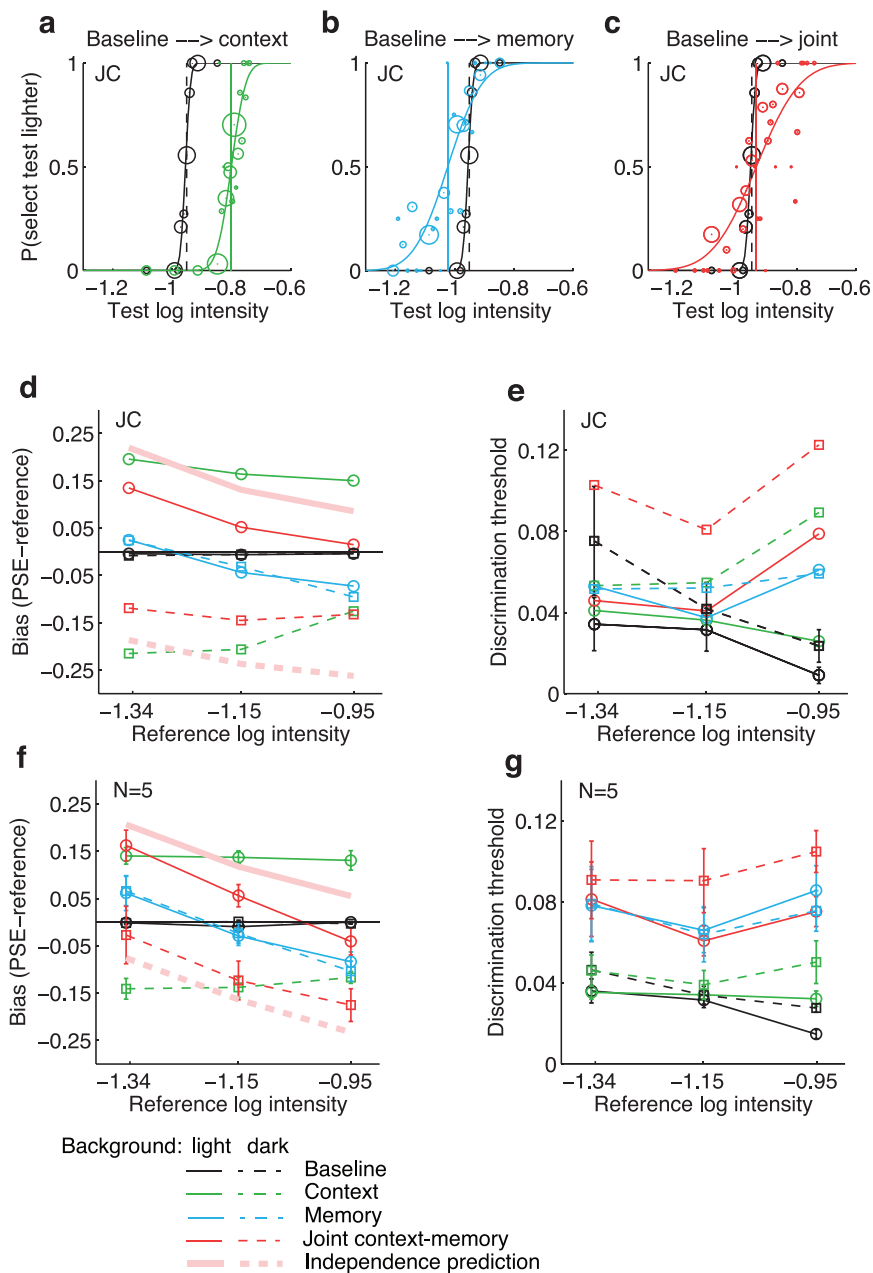


Figure 3. Psychophysics: Bias and thresholds. (a–c) Example psychometric functions with cumulative Gaussian fits are shown for observer JC. The black curve in each panel shows data for the baseline condition for one reference on the dark surround. The psychometric functions for the same reference/surround pair are shown in the context (a), memory (b), and joint (c) condition. Data points show the probability of selecting the test as the lighter stimulus; marker size indicates the number of trials for each data point. (d) Bias, as defined as the difference between each point of subjective equality (PSE) and the reference value for all reference/surround pairs and conditions for observer JC. Solid lines show data for the dark reference surround; dashed lines show data for the light reference surround. Colors are as in (a–c). Thick pink lines illustrate the independence prediction for the joint bias (see text). (e) Thresholds for the same observer are shown for each of the three references. Colors and line styles are as in (d). (f, g) Bias and thresholds averaged across five observers. Error bars are  $\pm 1$  SEM. Thick pink lines in (f) show the average independence predictions.

surround (as a higher test intensity was needed for a perceptual match). This effect is consistent with previous reports on the simultaneous lightness contrast illusion.

More surprisingly, memory retention affected both lightness appearance and precision: The memory

function (blue) is both shifted and scaled in comparison with baseline (Figure 3b). The leftward shift indicates that this particular reference (the lightest reference in the set) appeared *darker* than veridical with a 2.5-s delay between reference and test. We analyze this bias

more closely below. The shallower slope of the memory function indicates that the added delay also made responses less precise.

In Figure 3c, we plot data from the joint condition that combines context change with memory demands. Lightness judgments were less precise in the joint condition, just as they were in the memory-only condition. Surprisingly, however, the context effect on appearance was considerably *lessened* by the addition of memory load. This can be seen by comparing the small shift of the psychometric function in Figure 3c (red vs. black) to the large shift in Figure 3a (green vs. black). In other words, the effect of context on lightness judgments was smaller when the comparison was made with memory load.

Figure 3a through c shows the data for just one reference stimulus; next, we examine data across all references and conditions. We quantify appearance bias as the difference between the perceived (PSE) and the veridical reference value in each condition. Figure 3d shows the bias in each condition for the example observer from Figure 3a through c, and Figure 3f shows the bias averaged over all observers.

First, the effect of luminance context on lightness judgments was similar for all reference values, shown by the flat green lines. The context effect was also similar in absolute magnitude regardless of whether the reference was on the dark or light surround (solid and dashed green lines are approximately mirror images). These results replicate the classic finding in simultaneous lightness contrast displays. Second, the memory bias was always toward the mean of all the reference values, large for extreme references, and small for the middle reference. This pattern suggests a central tendency bias caused by memory retention. Third, the joint effect of context and memory was somewhere in between the pure context and memory effects (red lines). The joint bias had the negative slope of a central tendency bias, but it was shifted in the direction of the context bias.

To evaluate whether the memory and context biases acted independently in the joint condition, we compared the joint data to an independence prediction obtained by essentially adding the pure memory and context biases (see the details in the Methods section). The independence prediction is marked with thick pink lines in Figure 3d and f. The joint biases were consistently subadditive; the effects of memory and context on lightness estimates are thus not independent. We characterized the independence of context and memory in the whole data set with an index that measures the distance of each joint PSE from the PSE expected from full additivity. The indices were significantly negative, indicating subadditivity [ $t(59) = -7.6$ ,  $p < 10^{-5}$ ; Figure 4].

So far, we have focused on the effects of context and memory on lightness *appearance*. Next, we will turn to the effects of context, memory, and reference lumi-

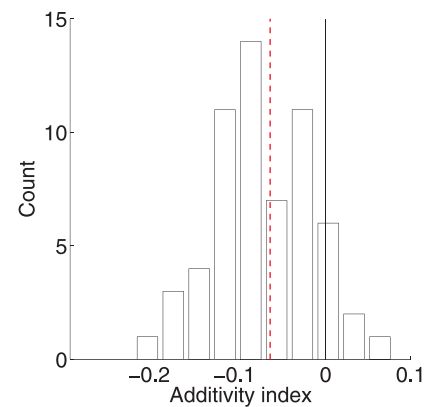


Figure 4. Nonadditivity of context and memory. A histogram of additivity indices calculated for each joint match. Vertical black line indicates the independence of memory and context biases. Negative values indicate subadditivity; positive values superadditivity. Vertical dashed red line shows the median index, which was significantly negative (see text).

nance on the *precision* of lightness matches, which we quantify as discrimination thresholds. As expected, memory demands had a negative overall impact on precision: Discrimination thresholds were generally higher in the memory conditions than in the no-memory conditions (blue and red lines in Figure 3g). This can be seen more clearly in Figure 5a, which shows thresholds for memory conditions to be consistently larger than for the no-memory conditions. In contrast, a context difference between the reference and test did not affect thresholds systematically (green lines in Figure 3g and Figure 5b).

The effect of reference luminance on thresholds varied across conditions. In the baseline condition, thresholds decreased with increasing reference intensity, which for our stimuli corresponds to decreasing contrast (black lines in Figure 3e and g). This pattern did not hold in the other conditions, in which thresholds were often smallest for the middle reference, as shown by the v-shape of the thresholds as a function of reference value in Figure 3e and g. Was this pattern systematically related to appearance bias? According to a probabilistic estimation strategy, a decrease in precision should increase the weight on prior information in the final estimate, strengthening estimation biases. The data shown in Figure 6b confirm this prediction: Less precise judgments tended to be more biased both in the memory (blue symbols,  $\rho = 0.57$ ,  $r^2 = 0.33$ ,  $p < 10^{-5}$ ) and in the joint condition (red symbols,  $\rho = 0.4$ ,  $r^2 = 0.16$ ,  $p = 0.001$ ).

## Intermediate discussion

We confirmed previous findings on lightness perception *without* memory demands in the baseline and



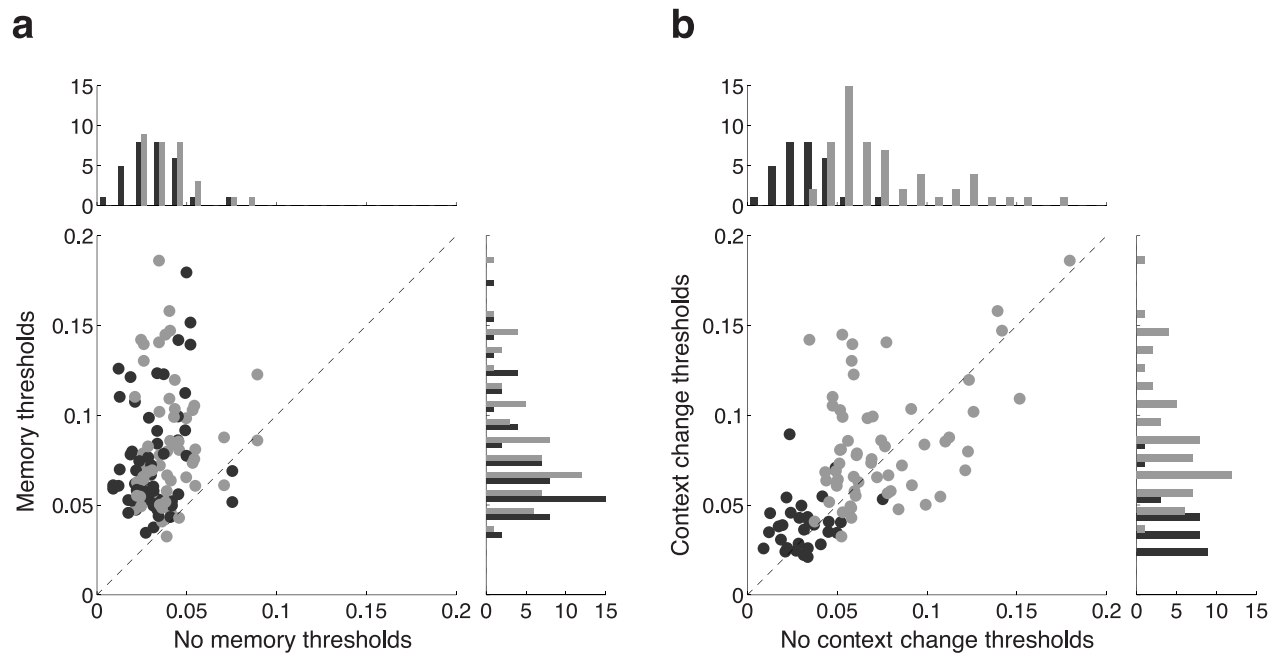


Figure 5. Effect of memory and context on thresholds. (a) Memory: Thresholds in the delay conditions (memory, joint) are plotted against thresholds in the simultaneous conditions (baseline, context). The delay conditions without and with distractors are plotted against the same simultaneous conditions, so each simultaneous threshold (x-axis) is plotted twice. Symmetric and asymmetric conditions are indicated with dark and light symbols, respectively. Marginal plots show corresponding threshold histograms. (b) Context: Thresholds in the asymmetric conditions (context, joint) are plotted against thresholds in the symmetric conditions (baseline, memory). Simultaneous and delayed conditions are indicated with dark and light symbols, respectively. Marginal plots show corresponding threshold histograms.

context conditions: Perceived lightness of a central target depended on surround luminance (Arend & Goldstein, 1987; Blakeslee, Reetz, & Mccourt, 2009; Heinemann, 1955), and discrimination thresholds in the baseline condition were lowest for smallest target contrasts, an effect sometimes referred to as *crispening* (Whittle, 1986).

Adding a 2.5-s retention interval between the reference and test stimulus revealed more surprising and novel patterns in the lightness matches. First, delayed lightness matches were biased toward the central stimulus value. Central tendency or range biases have previously been reported for stimulus estimates for line length (Ashourian & Loewenstein, 2011; Crawford, Huttenlocher, & Engebretson, 2000; Duffy, Huttenlocher, Hedges, & Crawford, 2010; Huttenlocher, Hedges, & Vevea, 2000), spatial frequency (Huang & Sekuler, 2010), and interval duration estimation (Jazayeri & Shadlen, 2010) and recently for color (Olkkonen & Allred, 2014; Olkkonen, McCarthy, & Allred, 2014). Although our experimental design does not discriminate between a bias toward the middle reference and a bias toward the mean luminance of the whole stimulus collection, the latter seems more likely if we assume that observers were paying attention to all stimuli nearly equally. Across stimulus domains, these biases have been interpreted as an optimal solution to

disambiguating uncertain sensory information: As the sensory representation of a stimulus becomes more uncertain in memory, an ideal observer should rely more on prior information about the stimulus ensemble to make their estimate (Ashourian & Loewenstein, 2011; Jazayeri & Shadlen, 2010). This interpretation is

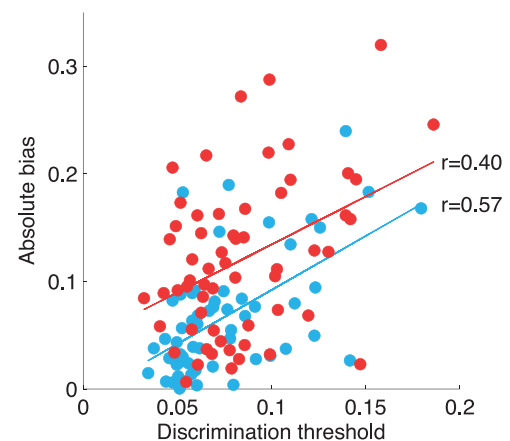


Figure 6. Relationship between bias and thresholds. Absolute bias values in the memory (blue) and joint (red) conditions are plotted against discrimination thresholds for each psychometric function. Correlation coefficients are noted next to each linear regression line.



supported by our current finding that precision and bias were correlated in the whole data set. It is worth noting that bias and precision are not routinely measured jointly in the color perception literature (but see Hillis & Brainard, 2005, 2007a, 2007b, for related work), and a relationship between bias and precision in color perception or memory has only recently been demonstrated (Bae, Olkkonen, Allred, & Flombaum, 2015; Olkkonen et al., 2014).

To our knowledge, this is the first report to reveal an interaction between perceptual context and memory processes for lightness processing (for hue, see Olkkonen & Allred, 2014). The joint effect of contextual and memory processing on perceived lightness seems puzzling: Adding a delay between the reference and test, viewed on different surrounds, *weakened* the effect of this context difference. One might expect the opposite based on traditional accounts of color adaptation: As the surrounds are visible throughout the whole trial, there should be more adaptation to the surrounds when a delay is added, resulting in a larger context bias (Fairchild & Lennie, 1992). Note that this result also means that lightness constancy became *poorer* because of memory. It is possible that this interaction is indicative of a more complex computational strategy that strives to be optimal, although in this case, it results in less constancy. In the next section, we explore this phenomenon by implementing two prominent constancy theories—*contrast coding* and *reflectance estimation*—and comparing them to human data.

## Two probabilistic models of lightness perception

We compare two classes of model in explaining the psychophysical data: the *contrast model* and the *reflectance model*. The contrast model observer compares stimuli by their local contrast, defined as the ratio of the center and surround luminance. A contrast-based model could be implemented in several ways. We chose a simple version of the contrast model that is known to produce the simultaneous lightness contrast illusion. The reflectance model observer compares stimuli in terms of their inferred reflectance.

We formulate both models probabilistically. This allows a principled way for the model observers to combine sensory evidence with prior knowledge about stimulus properties (Knill & Richards, 1996). Although our contrast model differs from previous implementations in that it is probabilistic, it has otherwise the same flavor: It uses border contrast to calculate perceptual matches. Amending existing implementations was necessary to be able to model memory effects.

In both models, the model observer makes noisy measurements of the presented stimuli and combines this sensory evidence with prior information about stimulus values to infer the values of interest (either contrast or reflectance). Probabilistic models can successfully account for several perceptual effects (without memory), including color estimates under uniform illumination (Brainard et al., 2006), lightness estimates under nonuniform illumination (Allred & Brainard, 2013), lighting direction (Stone, Kerrigan, & Porrill, 2009), speed perception (Stocker & Simoncelli, 2006), and orientation perception (Girshick, Landy, & Simoncelli, 2011). The probabilistic framework also provides a natural way to model the effect of memory on perception. Memory retention affects sensory uncertainty (Pasternak & Greenlee, 2005), and increased uncertainty, coupled with prior expectations about stimulus statistics, can account for central tendency biases in memory conditions (Ashourian & Loewenstein, 2011; Jazayeri & Shadlen, 2010). In both models, we model the effect of memory as increased noise in the sensory evidence.

The generative models and inference processes of both model observers are presented in Figure 7. The input to the contrast model (Figure 7a) is the log border contrast between the center and surround ( $k_1$  and  $k_2$  for the two contrasts). The model observer makes noisy measurements ( $m_1$  and  $m_2$ ) of these contrasts (noise is assumed to be normally distributed). Given a measurement, each contrast has a certain likelihood to have given rise to the measurement. The model observer combines the likelihoods with prior probabilities to compute the posterior probabilities for the contrast values, given the measurements  $p(k_i|m_i)$ . As in other probabilistic models of delayed estimation, priors were based on the statistics of the stimuli used in the experiments.

We implemented two variants of the contrast model. The first variant had a single prior that was centered on the mean contrast across conditions (single-prior variant). In the second variant, the prior was centered on the mean in each condition separately (multiple-prior variant; see the Appendix for details). Based on the computed posteriors, the model observer chooses the stimulus that more probably had lower contrast (that is, was lighter, as the stimuli were decrements). Figure 7c illustrates this inference process. Because this model observer matches stimuli by contrast, it naturally produces the simultaneous lightness contrast effect: If the center patches have equal luminances but their surrounds differ, they will have unequal edge contrasts and thus they will be perceived differently. A perceptual match (50% probability of choosing one stimulus over the other) occurs when the contrasts match. To achieve this match, the test luminance has to be adjusted to a value that differs from the reference luminance.

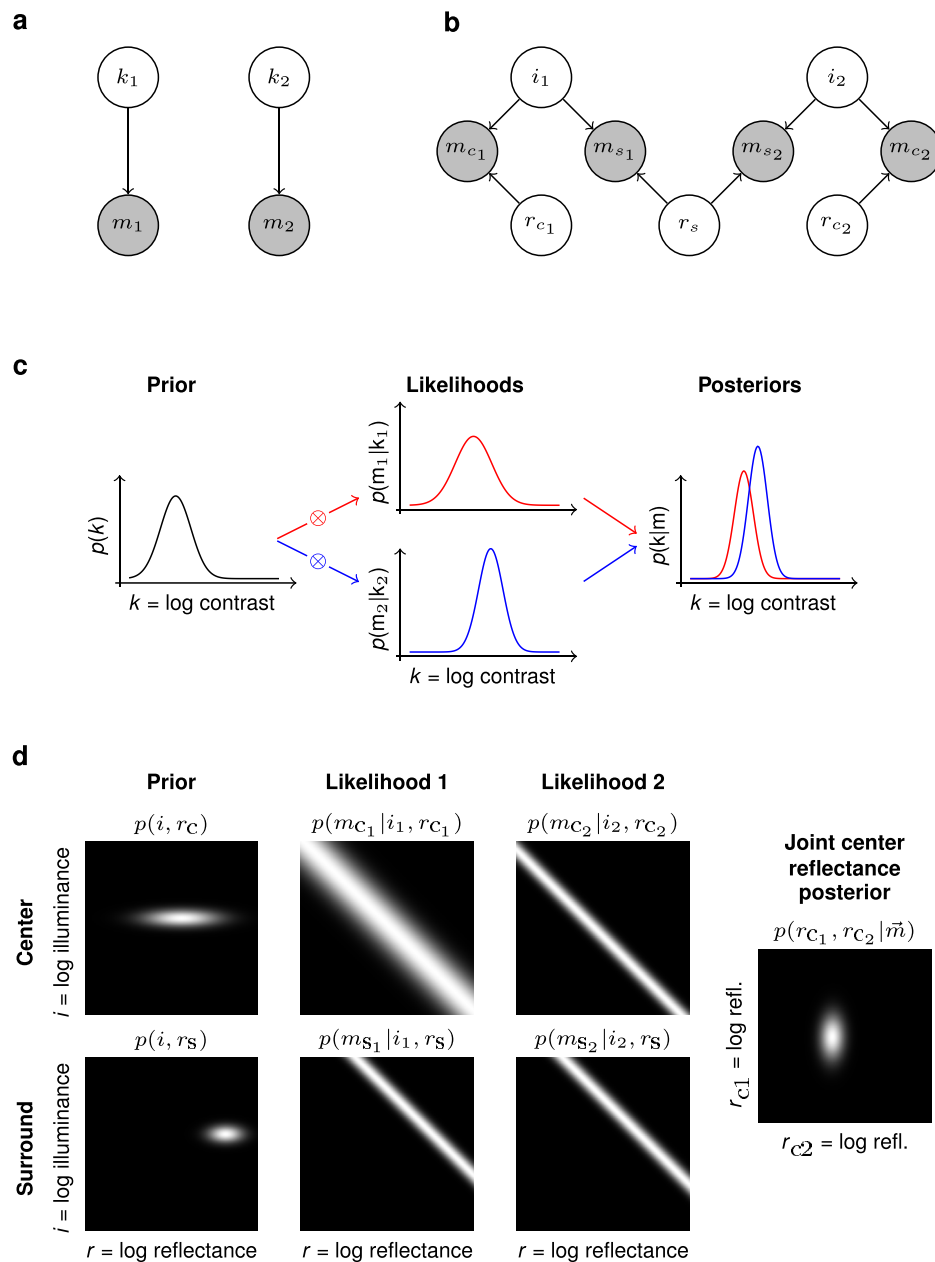


Figure 7. Two probabilistic models of lightness perception. Panels a and b show how the measurements are generated in each model and present the inference problem. Panels c and d illustrate the inference process on a single trial. (a) The contrast model. The variables  $k_i$  represent the log border contrasts of the two stimulus patches. The model observer makes a noisy measurement of each contrast ( $m_i$ ) and from these makes an inference about the contrasts  $k_i$ . Given a measurement, each contrast is associated with a likelihood. To infer which patch was lighter (had lower contrast), the observer combines the likelihoods with prior information and chooses the patch that had the higher probability of having a smaller contrast. (b) The reflectance model. The observer makes four noisy log luminance measurements ( $m_i$ ), one of each surround and center. The observer assumes each log luminance is the sum of log illuminance and log reflectance (only the illuminance and reflectance variables are shown; the intervening luminance is omitted). The observer infers center reflectance from the measurements by combining the likelihoods for log illuminance and reflectance with prior information and computing the posterior distribution for the two center reflectances. The observer chooses the center that more probably had the higher reflectance. (c) An example of the inference and decision process of the contrast model. Given the noisy measurement, each log contrast is associated with a likelihood, as shown in the middle panels for the two stimuli to be compared. The observer combines these likelihoods with a prior distribution over log contrast (left panel) to compute the posterior distribution for log contrast given the measurements (right panel). The observer chooses the stimulus that more probably had the smaller contrast. (d) Inference and decision process of the reflectance model. The model observer has a separate prior for log center

→

←

reflectance, surround reflectance, and illuminance. These are shown as two-dimensional priors in the left-hand panels. Given the measurement, each reflectance-illuminance pair is associated with a likelihood (middle panels, for the two centers and two surrounds). Combining the likelihoods with the prior distributions and integrating out the other variables, the observer computes the posterior distribution for the two center reflectances and chooses the stimulus that more probably had a higher reflectance.

The input to the reflectance model is log luminance, which in the real world depends on both illumination and reflectance. The generative model and inference problem for the reflectance model are described in Figure 7b. This model observer makes four noisy log luminance measurements, one for each surround ( $m_{s1}$  and  $m_{s2}$ ) and one for each center ( $m_{c1}$  and  $m_{c2}$ , again with normally distributed error). Each measurement results in a two-dimensional illumination-reflectance likelihood. The observer then infers the center reflectances and chooses the center that more probably had a higher reflectance. Figure 7d illustrates this computation graphically. The likelihood functions (middle columns in Figure 7d) are ambiguous about reflectance, because several different combinations of surface reflectance and illumination can produce any given log luminance value. Priors are therefore needed to constrain the reflectance values (Figure 7d, left column). The observer infers reflectance by combining the likelihood with priors for illumination and reflectance and computing the posterior probability for center reflectances given the measurements  $p(r_{c1}, r_{c2} | m_{c1}, m_{c2}, m_{s1}, m_{s2})$ . Importantly, the observer assumes that the surround reflectance is uniform across the display (there is a single surround reflectance variable,  $r_s$ ) and any luminance change between the two surrounds is due to a change in illumination (two illumination variables,  $i_1$  and  $i_2$ ). This constraint produces the simultaneous lightness contrast effect: Consider the case in which the two center patches have equal luminances but the surround luminances differ. The model observer assumes a constant surround reflectance and thus imputes the luminance change to an illumination difference across the stimulus. But if the illumination is different for the two center patches, they cannot have the same reflectance (as they have equal luminance). Because this model observer bases its lightness judgments on reflectance, the two centers will be perceived as different.

In all models, we assume that a delay between the first (reference) and the second (test) stimulus adds noise to the first measurement. This added noise produces the memory bias we observed: The noisier measurement leads to a wider likelihood, which causes the posterior distribution to be drawn more toward the prior, biasing perception of the stimulus. The prior is centered on the (average) reference value. Therefore, the sign of the bias depends on whether the reference is lighter or darker than this average: A reference that is

darker than average is matched with a lighter test than it is without the delay, and vice versa for a lighter-than-average reference. The effect of this added noise can be seen in the example likelihoods in Figure 7c, d. The implementation of all models is outlined in detail in the Appendix.

### Comparison with human data

We ran the models in all experimental conditions and extracted the PSEs and discrimination thresholds from the proportion-lighter data as with the human observers. We adjusted the parameters of the models to give as good a match as possible to the average human data (in terms of PSEs and thresholds; see Appendix for details).

Several features emerge from a side-by-side comparison of model and human data (Figure 8; human data reproduced from Figure 3d). First, both contrast and reflectance models produce a memory effect (seen as the negative slope in the blue lines). But in the case of the single-prior contrast model, there is an additional offset depending on whether the reference was on the dark (solid line) or light (dashed line) surround: The middle reference has a nonzero bias, unlike the human data and the reflectance model. The nonzero bias for the contrast model results from using one prior for all experimental conditions (see the Appendix for explanation). If the prior is allowed to shift with experimental condition, as is the case in the multiple-prior contrast model, this asymmetry disappears. The reflectance model (Figure 8d) matches human observers even with a fixed prior.

Second, all models produce the simultaneous lightness contrast effect (context condition; green lines), but the mechanism behind the effect is different for the contrast and reflectance models as explained above. The magnitude of the effect produced by the contrast model depends only on border contrast and is clearly larger than the effect in the human data. The reflectance model produces an effect that is comparable to human observers.

Finally, the reflectance model reproduces the sub-additivity of the memory and context effects observed in the human data (red lines). The predicted joint effect of independent memory and context biases is shown by the thick light-red lines as before. The reflectance model shows a joint effect that is smaller than predicted

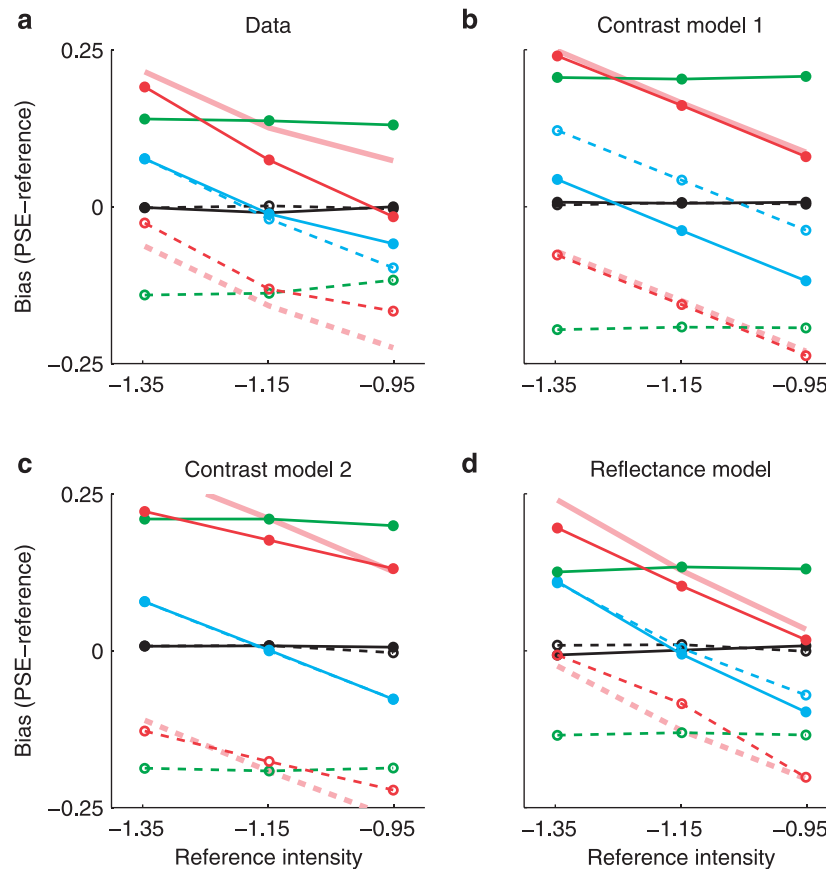


Figure 8. Comparison of the biases in the human data and the three models. The upper left-hand panel shows the observed biases as in Figure 3d. The other panels show the biases produced by the two contrast models and the reflectance model. Line colors: black, baseline; blue, memory; gray, constancy; red, joint condition. The independence predictions (thick pink lines) were computed for the models in the same way as for the data.

(Figure 8d). The effect in the single-prior contrast model matches the prediction, that is, shows no subadditivity (Figure 8b). The multiple-prior contrast model shows a degree of subadditivity, but its magnitude depends on the reference luminance in a way that is not consistent with the human data (Figure 8c). In the case of the multiple-prior contrast model, the subadditivity is due to the changing prior: The prior in the memory-only condition is different from the prior in the context and joint conditions, because the average reference contrast was different in these conditions (the surround was uniform in the memory-only condition and bipartite in the other conditions). These different priors bias perception to different extents, leading to apparent subadditivity when compared side by side. Note that the memory effect produced by this model is not identical in the joint and memory-only conditions: The slopes of the red (joint) and blue (memory) lines are different in Figure 8c. This is inconsistent with the human data, in which the slopes for the memory and joint conditions are indistinguishable (Figure 8a).

The reflectance model, which has two priors that remain the same across conditions, does produce the

same memory effect in the memory-only and joint conditions (Figure 8d). The memory effect, characterized by the slope, occurs because the added noise in both memory conditions biases perception toward the average stimulus value, which is the same in the memory and joint conditions. But at the same time, the added noise puts less constraint on illuminance. As the noise, and uncertainty, of the sensory signal increases, the observer gives more weight to the prior distribution. The model observer assumes there are two illuminants, but these two illuminants have a common prior distribution. Therefore, when uncertainty increases and the prior is given more weight, the sensory measurements become more consistent with the two illuminants being closer together, which decreases the context effect. The context effect is weakened because surround reflectance is assumed uniform—the context effect in this model comes from the assumption of two illuminants, and the closer those illuminants are to each other, the weaker the effect of context. Thus, compared with the memory-only and context-only conditions, the reflectance model produces the same memory effect but



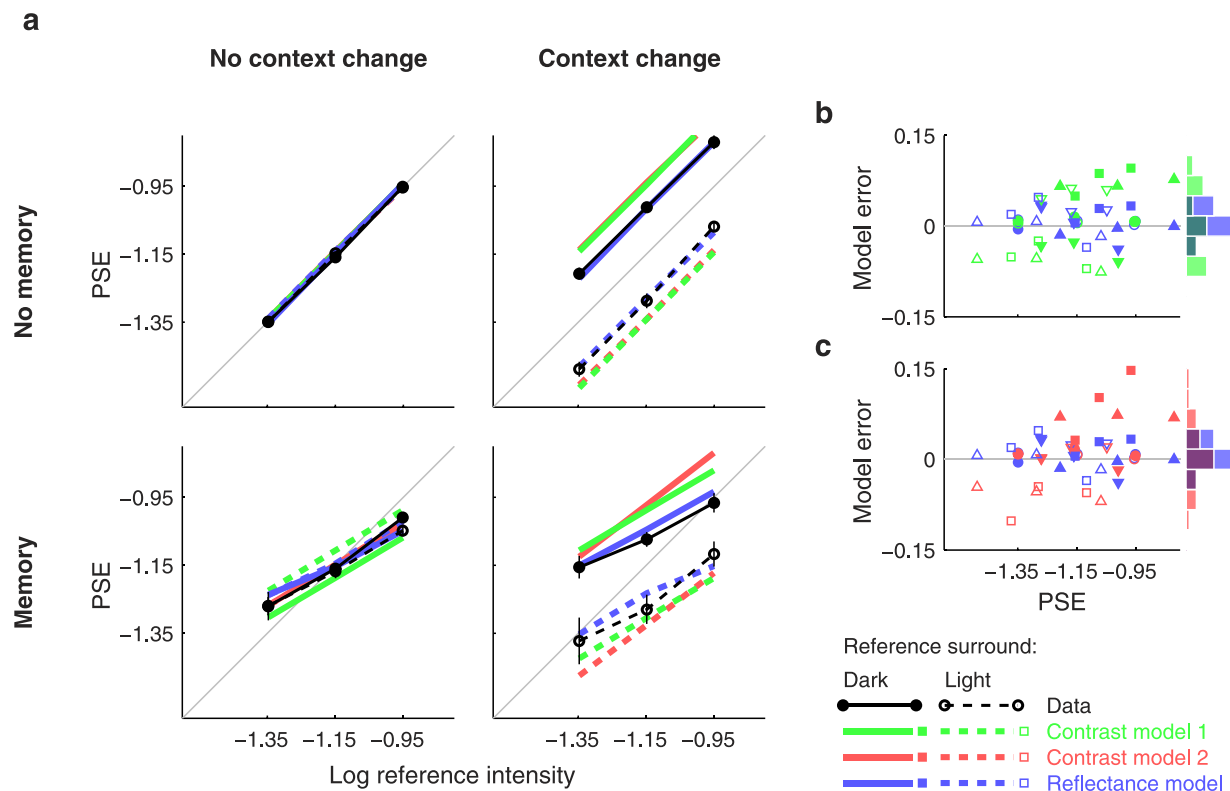


Figure 9. (a) Comparison of PSE values in the human data and the models. The PSE values are plotted against the reference luminance values. Human data are plotted in black symbols and lines; the thick colored lines show the model fits. Blue, reflectance model; green and red, contrast models. (b, c) Analysis of errors in the model PSE values. The error (difference between the model PSE and the observed PSE) is plotted against the observed PSE. The histograms to the right show the distribution of errors. Blue, reflectance model; green and red, contrast models. The reflectance model errors are plotted in both (b) and (c) to ease comparison with both contrast models.

a smaller context effect in the joint condition, consistent with human data.

Figure 8 shows the appearance biases in each condition for the human and model observers in separate panels. An alternative way to compare the models to the human data is to plot the PSEs for human and model observers in the same panel for each condition separately. The four panels in Figure 9 show data from human and model observers in each of the conditions, shown in a grid to illustrate the factorial effects of memory and context. Average human data are shown in black, the reflectance model in blue, and the contrast models in green and red. Overall, the reflectance model PSEs are clearly a better fit to the observed PSEs; this can be seen by comparing the error between model estimates of PSEs and observed PSEs in Figure 9b, c. On average, the reflectance model PSEs (blue) were closer to the observed values than the contrast model PSEs (green in b, red in c).

We also extracted model discrimination thresholds. In all models, thresholds in the memory conditions were higher than in the no-memory conditions (the memory-related noise in the model was nonzero; this is what leads to the memory bias described above). All

models, however, showed a systematic deviation from the observed thresholds (see Supplemental Figure S2). First, in the human data, thresholds tended to decrease with reference intensity, especially in the baseline condition. The models do not show this behavior. Second, the modeled thresholds in the memory conditions are lower than in the human data (increased memory noise would lead to better fits for the thresholds, but it would also worsen the fit for PSEs). There was, however, much more variation in thresholds than in PSEs across the human observers.

The contrast model has three parameters, whereas the reflectance model has four (see the supplemental material for all model parameters). To compare the goodness-of-fit of the models while taking the number of parameters into account, we computed the reduced  $\chi^2$  values for each (the raw  $\chi^2$  values divided by degrees of freedom). Comparison of the  $\chi^2$  values confirms that the reflectance model gives the best account of the PSE data ( $\chi^2 = 4.05$  for Contrast Model 1, 2.67 for Contrast Model 2, and 1.31 for the reflectance model). Note that the second contrast model is in a way more complicated than either of the other models, because there the prior depends on the condition. If we allowed the tuning of

the prior in the reflectance model also, it would probably give an even better fit. For threshold data, the contrast models fit the human data better ( $\chi^2 = 1.35$  for Contrast Model 1, 0.97 for Contrast Model 2, and 1.88 for the reflectance model), but considering both PSEs and thresholds together, the reflectance model gives the best fit ( $\chi^2 = 2.52$  for Contrast Model 1, 1.70 for Contrast Model 2, and 1.45 for the reflectance model).

## Discussion

We studied the strategies used by humans to judge lightness from ambiguous sensory data. Our novel psychophysical paradigm factorially combined perceptual context and memory demands, and the data provide support for a model of lightness perception in which observers base lightness judgments on inferred reflectance and not on border contrast. We focused here on lightness judgments, but as similar context dependencies exist in both achromatic and chromatic color perception (see, e.g., Maloney & Schirillo, 2002), we expect the same principles to hold in full-color scenes. Furthermore, both memory and context effects have been separately reported in other perceptual domains, such as orientation (e.g., Goddard, Clifford, & Solomon, 2008; Fischer & Whitney, 2014) and motion (e.g., McKeefry, Burton, & Vakrou, 2007; Nawrot & Sekuler, 1990) perception; an exciting avenue for future research is to investigate general principles of constancy computations by testing for perception-memory interactions in many perceptual domains.

The addition of a memory load into a perceptual task introduced systematic errors: a central tendency bias that did not add simply to the (perceptual) simultaneous contrast effect. This bias is similar to range effects reported in other stimulus domains (Ashourian & Loewenstein, 2011; Crawford et al., 2000; Duffy et al., 2010; Huang & Sekuler, 2010; Huttenlocher et al., 2000; Jazayeri & Shadlen, 2010) and in the hue dimension of color (Olkkonen et al., 2014). The most interesting and novel experimental finding, however, is the subadditivity of the memory and context biases. If these two biases were independent, one would expect them to add in a joint condition. We should then be able to predict the joint memory-context bias from the separately measured memory and context biases. This was not the case: The observed joint bias was considerably smaller than the prediction. In other words, observers were *less constant* when evaluating surface lightness across a memory delay compared with simultaneous comparison. Lightness theories do not tend to make predictions about memory effects, so it is not clear how this result fits into

existing frameworks. But in the chromatic color domain, we recently found a similar effect for hue (Olkkonen & Allred, 2014), and de Fez, Capilla, Luque, Pérez-Carpinell, and del Pozo (2001) reported slightly poorer constancy when observers matched Munsell chips across an illuminant change with a memory delay compared with simultaneous matching. To gain a better understanding about what might cause this subadditivity between context and memory effects, we investigated how different models of lightness perception can account for the biases. The  $2 \times 2$  design we used, especially with the joint memory-context condition, provided a rich enough data set against which to test these models.

We evaluated two classes of computational framework invoked most often to explain contextual effects in lightness perception: models that use border contrast as a proxy to reflectance (e.g., Blakeslee & McCourt, 2012; Land & McCann, 1971; Rudd & Zemach, 2005; Wallach, 1948), and models that form explicit reflectance (and sometimes illumination) estimates based on scene cues and prior information (e.g., Allred & Brainard, 2013; Bloj et al., 2004; Murray, 2013). The two approaches use fundamentally different strategies to provide stable lightness percepts: one relies on a type of heuristic (contrast) whereas the other tries to solve the (ill-defined) problem of inferring stimulus reflectance.

Contrast models successfully account for a variety of perceptual phenomena. These models rely on straightforward calculations operating on contrast signals rather than on complex inferences about attributes that are not directly observable (such as reflectance). However, contrast models fail when border contrast signals do not correlate with reflectance in a straightforward manner, such as when scenes contain complex illumination or geometric structure (Bloj, Kersten, & Hurlbert, 1999; Gilchrist, 1977; Maertens et al., 2015). On the other hand, reflectance estimation models successfully account for many perceptual phenomena in simple and complex scenes, but only a few implementations exist (Allred & Brainard, 2013; Brainard et al., 2006; Murray, 2013; also see Bloj et al., 2004, for an analogous “equivalent illuminant” model). Reflectance (or illumination) estimation models are necessarily more complex than contrast models as they embrace the ambiguity of luminance being a product of reflectance and illumination. To alleviate this ambiguity, the models use prior constraints on reflectance and illumination. In our probabilistic formulations, both contrast and reflectance model observers used prior information to constrain the percepts. Although the contrast model does not necessarily need prior information (without a prior, the model observer could still perform inference based on maximum likelihood), it does not produce a memory bias without a prior. We

thus augmented the contrast model to include memory effects to make the two models more comparable.

In our hands, both types of model produce memory and context effects, but the reflectance model is quantitatively closer to human performance. Moreover, only the reflectance model accurately captures the subadditivity of memory and context biases. The mechanism for the memory bias is similar in both models: Keeping the sensory evidence in memory increases uncertainty, which leads to more weight being given to prior information. The prior is centered on the average stimulus value, drawing the perceived value toward the mean. The mechanism for the context effect, on the other hand, is very different in the two models, as explained in the modeling section. The reflectance model produces a context bias that is comparable with the one measured with human observers. The context bias of the contrast model, however, is larger than observed with humans. There is no parameter in the model that could be adjusted to remedy this, as the match depends only on the border contrast. More complex contrast-based models could, however, correctly account for the effect of context (e.g., Gilchrist et al., 1999; Rudd, 2014). The main difference between the models, finally, is their ability to account for the subadditivity of the memory and context biases. The reflectance model shows this subadditivity, giving a fairly good fit both qualitatively and quantitatively, whereas the first contrast model does not produce any subadditivity. To evaluate whether a more complex contrast model would better match human performance, we additionally implemented a contrast model with separate priors for each condition, reflecting the possibility that the observers switched priors between every block. This model produced the correct pure memory bias but did not quite capture the pattern of additivity between context and memory. To recapitulate, the subadditivity in the reflectance model is caused mainly by the two putative illuminant estimates being drawn toward the prior. The shift toward a common prior means that the effective difference between the two estimated illuminants becomes smaller in memory. But the context effect is a result of the two illuminants being different, so the shift also reduces the context effect. Because the contrast model only has a prior for border contrast (one or many, depending on the version), there is no mechanism for the two backgrounds to become more similar in memory. This is the key difference between the contrast and reflectance models and presumably the reason why the reflectance model does better in explaining human performance. Thus, by adding a temporal dimension that affected internal estimation noise, we were able to separate the predictions from contrast coding and reflectance estimation models. Comparison of modeling results to human data suggests that human observers employ a reflectance estimation strategy when estimating lightness from an ambiguous center-surround display.

Implementing the models, we made assumptions about the priors. First, we assumed the prior distributions to be centered on the mean stimulus value. We find this reasonable, especially as it leads to the central tendency bias demonstrated by both the current study and earlier reports. The second assumption concerned the shape of the distribution. We assumed the prior to have a normal distribution (see model descriptions for details), although we naturally cannot be sure that this is the case. Strictly, if the prior reflects a learned stimulus distribution, it could be a series of very narrow peaks placed at the stimulus values already encountered. Or it could be a uniform distribution from the lowest to the highest encountered value. Both of these alternatives seem too strict given the uncertainty in sensory encoding. Some earlier studies have decoded the prior used by observers (Girshick et al., 2011; Stocker & Simoncelli, 2006). We made no such attempt, especially given the complexity of the reflectance model, which is beyond the scope of our article. The main purpose of the modeling was to illustrate the differences in the behavior of the two models, given certain types of prior distribution. Finally, alternative model structures surely exist for the contrast and reflectance-estimation models, and a different or more complex contrast model might well account for the results. The present models, however, are good exemplars of two model classes and seemed reasonable starting points in comparing probabilistic versions of contrast and reflectance-estimation models.

Some lightness models do not neatly fall under the two categories we have imposed. For instance, scission or layer models are similar to our reflectance model in that they seek to separate the retinal image into layers of surface reflectance, illumination, and transparency (Adelson & Pentland, 1996; Anderson & Winawer, 2008). The key difference to our reflectance model is that in layer models, lightness estimates are not constrained by prior information about surfaces and illuminants but rather by contrast relationships at borders and on figural information, such as border junctions. Another influential model that falls outside our classification is the anchoring model, which explains perceived lightness with a “brightest-is-white” anchoring rule specific to a given illumination framework (Gilchrist et al., 1999). The anchoring model is similar to contrast models in that it derives lightness from luminance relationships, but it has the added complexity of segmenting a scene into illumination frameworks. This allows the model to qualitatively account for the effects of spatial arrangement on perceived lightness. The gamut relativity model by Vladusich (e.g., Vladusich & McDonnell, 2014) combines assumptions from both layer and anchoring models to derive lightness estimates. Although layer and anchoring models can easily account for the



classical simultaneous lightness contrast illusion, it is unclear how they would model the effects of memory uncertainty, as they are not probabilistic. We wish to advance the general point that a computational strategy that uses constraints from prior knowledge rather than from pictorial cues alone seems more flexible in accounting for lightness perception in a dynamic, three-dimensional world. The present results show that a probabilistic estimation strategy that has access to prior information about surfaces and illuminants does indeed account for human lightness perception in a combined perceptual and memory task.

## Conclusion

Both context and memory bias the perception of lightness, but these biases are not additive: The bias in a joint memory-context task is smaller than one predicted by independent, additive biases. The addition of a memory load to a perceptual task in fact decreased the perceptual, context-induced bias. The results are consistent with a model observer that makes lightness judgments based on inferred stimulus reflectance and less well described by a model observer that uses simple border contrast for lightness judgments. More generally, these results suggest that adding realistic task demands (memory) to classical perceptual tasks may help adjudicate between competing computational frameworks across sensory domains.

*Keywords:* perceptual constancy, memory, probabilistic inference, color perception, lightness perception

## Acknowledgments

This research was supported by grants to S. R. A. (NSF CAREER BCS 0954749) and M. O. (Academy of Finland Research Fellow Scheme).

Commercial relationships: none.  
Corresponding author: Maria Olkkonen.  
Email: maria.olkkonen@durham.ac.uk.  
Address: Department of Psychology, Durham University.

## References

- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *262*, 2042–2044.
- Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference, volume 1* (pp. 409–423). New York: Cambridge University Press.
- Allen, E. C., Beilock, S. L., & Shevell, S. K. (2011). Working memory is related to perceptual processing: A case from color perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1014–1021.
- Allred, S. R., & Brainard, D. H. (2013). A Bayesian model of lightness perception that incorporates spatial variation in the illumination. *Journal of Vision*, *13*(7):18, 1–18, doi:10.1167/13.7.18. [PubMed] [Article]
- Anderson, B. L., & Winawer, J. (2008). Layered image representations and the computation of surface lightness. *Journal of Vision*, *8*(7):18, 1–22, doi:10.1167/8.7.18. [PubMed] [Article]
- Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A*, *4*, 2281–2285.
- Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 2. Reflectance variation. *Perception & Psychophysics*, *54*, 457–468.
- Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One*, *6*, e19551.
- Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*, 744–763.
- Blakeslee, B., & McCourt, M. E. (2012). When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision Research*, *60*, 40–50.
- Blakeslee, B., Reetz, D., & McCourt, M. E. (2009). Spatial filtering versus anchoring accounts of brightness/lightness perception in staircase and simultaneous brightness/lightness contrast stimuli. *Journal of Vision*, *9*(3):22, 1–17, doi:10.1167/9.3.22. [PubMed] [Article]
- Bloj, M., Ripamonti, C., Mitha, K., Hauck, R., Greenwald, S., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision*, *4*(9):6, 735–746, doi:10.1167/4.9.6. [PubMed] [Article]
- Bloj, M. G., & Hurlbert, A. C. (2002). An empirical study of the traditional Mach card effect. *Perception*, *31*, 233–246.
- Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999).



- Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, *402*, 877–879.
- Brainard, D. H., Longère, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006). Bayesian model of human color constancy. *Journal of Vision*, *6*(11):10, 1267–1281, doi:10.1167/6.11.10. [PubMed] [Article]
- Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision*, *11*(5):1, 1–18, doi:10.1167/11.5.1. [PubMed] [Article]
- Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. P. Hornak (Ed.), *Encyclopedia of imaging science and technology* (pp. 172–188). New York: Wiley.
- Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, *11*, 280–284.
- Dakin, S. C., & Bex, P. J. (2003). Natural image statistics mediate brightness ‘filling in.’ *Proceedings of the Royal Society B: Biological Sciences*, *270*, 2341–2348.
- de Fez, M. D., Capilla, P., Luque, M. J., Pérez-Carpinell, J., & del Pozo, J. C. (2001). Asymmetric colour matching: Memory matching versus simultaneous matching. *Color Research & Application*, *26*, 458–468.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*, 415–434.
- Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review*, *17*, 224–230.
- Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *Journal of Neuroscience*, *29*, 15258–15265.
- Fairchild, M. D., & Lennie, P. (1992). Chromatic adaptation to natural and incandescent illuminants. *Vision Research*, *32*, 2077–2085.
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*, 738–743.
- Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, *195*, 185–187.
- Gilchrist, A. L., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Xiaojun, L., . . . Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*, 795–834.
- Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*, 926–932.
- Goddard, E., Clifford, C. W. G., & Solomon, S. G. (2008). Centre-surround effects on perceived orientation in complex images. *Vision Research*, *48*, 1374–1382.
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635.
- Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing- and test-field luminances. *Journal of Experimental Psychology*, *50*, 89–96.
- Hillis, J. M., & Brainard, D. H. (2005). Do common mechanisms of adaptation mediate color discrimination and appearance? Uniform backgrounds. *Journal of the Optical Society of America A*, *22*, 2090–2106.
- Hillis, J. M., & Brainard, D. H. (2007a). Distinct mechanisms mediate visual detection and identification. *Current Biology*, *17*, 1714–1719.
- Hillis, J. M., & Brainard, D. H. (2007b). Do common mechanisms of adaptation mediate color discrimination and appearance? Contrast adaptation. *Journal of the Optical Society of America A*, *24*, 2122–2133.
- Huang, J., & Sekuler, R. (2010). Attention protects the fidelity of visual memory: Behavioral and electrophysiological evidence. *Journal of Neuroscience*, *30*, 13461–13471.
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220–241.
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*, 1020–1026.
- Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A*, *13*, 1981–1991.
- Kang, M.-S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, *18*, 860–869.
- Kingdom, F. A. A. (2008). Perceiving light versus material. *Vision Research*, *48*, 2090–2105.
- Kingdom, F. A. A. (2010). Lightness, brightness and transparency: A quarter century of new ideas,

- captivating demonstrations and unrelenting controversy. *Vision Research*, *51*, 652–673.
- Kingdom, F. A. A., & Moulden, B. (1992). A multi-channel approach to brightness coding. *Vision Research*, *32*, 1565–1582.
- Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, *351*, 228–230.
- Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge, MA: Cambridge University Press.
- Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, *61*, 1–11.
- Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, *25*, 1215–1226.
- Maertens, M., Wichmann, F. A., & Shapley, R. M. (2015). Context affects lightness at the level of surfaces. *Journal of Vision*, *15*(1):15, 1–15, doi:10.1167/15.1.15. [PubMed] [Article]
- Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, *62*, 81–92.
- Maloney, L. T., & Schirillo, J. A. (2002). Color constancy, lightness constancy, and the articulation hypothesis. *Perception*, *31*, 135–139.
- McKeefry, D. J., Burton, M. P., & Vakrou, C. (2007). Speed selectivity in visual short term memory for motion. *Vision Research*, *47*, 2418–2425.
- Murray, R. F. (2013). Human lightness perception is guided by simple assumptions about reflectance and lighting. In B. E. Rogowitz, T. N. Pappas, & H. de Ridder (Eds.), *Human vision and electronic imaging XVIII* (Vol. 8651, pp. 1–11). Burlingame, CA: SPIE.
- Nawrot, M., & Sekuler, R. (1990). Assimilation and contrast in motion perception: Explorations in cooperativity. *Vision Research*, *30*, 1439–1451.
- Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PLoS One*, *9*, e8648:1–11.
- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*(11):5, 1–15, doi:10.1167/14.11.5. [PubMed] [Article]
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews. Neuroscience*, *6*, 97–107.
- Pearson, J., & Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends in Cognitive Sciences*, *12*, 334–341.
- Purves, D., Shimp, A., & Lotto, R. B. (1999). An empirical explanation of the cornsweet effect. *Journal of Neuroscience*, *19*, 8542–8551.
- Purves, D., Williams, S. M., Nundy, S., & Lotto, R. B. (2004). Perceiving the intensity of light. *Psychological Review*, *111*, 142–158.
- Rudd, M. E. (2014). A cortical edge-integration model of object-based lightness computation that explains effects of spatial context and individual differences. *Frontiers in Human Neuroscience*, *8*, 1–14.
- Rudd, M. E., & Zemach, I. K. (2005). The highest luminance anchoring rule in achromatic color perception: Some counterexamples and an alternative theory. *Journal of Vision*, *5*(11):5, 983–1003, doi:10.1167/5.11.5. [PubMed] [Article]
- Schirillo, J., Reeves, A., & Arend, L. (1990). Perceived lightness, but not brightness, of achromatic surfaces depends on perceived depth information. *Perception & Psychophysics*, *48*, 82–90.
- Schirillo, J. A., & Shevell, S. K. (1997). An account of brightness in complex scenes based on inferred illumination. *Perception*, *26*, 507–518.
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*, 207–214.
- Spehar, B., Debonet, J. S., & Zaidi, Q. (1996). Brightness induction from uniform and complex surrounds: A general model. *Vision Research*, *36*, 1893–1906.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*, 578–585.
- Stone, J. V., Kerrigan, I. S., & Porrill, J. (2009). Where is the light? Bayesian perceptual priors for lighting direction. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1797–1804.
- Supèr, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science*, *293*, 120–124.
- Uchikawa, K., Kuriki, I., & Tone, Y. (1998). Measurement of color constancy by color memory matching. *Optical Review*, *5*, 59–63.
- Vladusich, T. (2012). Simultaneous contrast and gamut relativity in achromatic color perception. *Vision Research*, *69*, 49–63.
- Vladusich, T., & McDonnell, M. D. (2014). A unified account of perceptual layering and surface ap-

pearance in terms of gamut relativity. *PLoS One*, 9, e113159.

von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*. Leipzig, Germany: Leopold Voss.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, 38, 310–324.

Whittle, P. (1986). Increments and decrements: Luminance discrimination. *Vision Research*, 26, 1677–1691.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293–1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63, 1314–1329.

Zaidi, Q. (1998). Identification of illuminant and object colors: heuristic-based algorithms. *Journal of the Optical Society of America A*, 15, 1767–1776.

## Appendix

### Contrast model

The signal to the contrast model is the log-contrast  $k$  between the surround and the center of the stimulus:  $k_j = \log(L_{sj}/L_{cj}) = \log(L_{sj}) - \log(L_{cj})$ , where  $L_{sj}$  and  $L_{cj}$  are the surround and center luminances, respectively, and  $j = 1, 2$  is the index for the stimulus (log border contrast well approximates Weber contrast at low contrasts). On each trial, the model observer makes noisy measurements of the two stimulus contrasts. The measurements grow with log-contrast, and the noise is additive:

$$m_j = k_j + \epsilon_m, \quad (\text{A1})$$

where the error is normally distributed:  $\epsilon_m \sim N(0, \sigma_m)$  and thus  $m_j \sim N(k_j, \sigma_m)$ ;  $\sigma_m$  is the standard deviation of the measurement noise. The observer only has access to the measurements  $m_j$  and not to the actual contrast values. On a given trial, the observer infers the contrast values  $k_j$  from the measurements  $m_j$  as explained further below. The inference problem is illustrated in Figure 7a.

When there is a delay between the first and the second stimulus (the memory and joint conditions in the experiment), the first measurement becomes noisier with time. We assume that the delay only adds noise and does not on average bias the measurement itself. This time-dependent noise is normally distributed and adds to the measurement noise. In the delay conditions,

the measurement of the *first* stimulus is

$$m_1 = k_1 + \epsilon_m + \epsilon_t \quad (\text{A2})$$

where  $\epsilon_t \sim N(0, \sigma_t)$  is the time-dependent noise. In the delay conditions, the total noise for the first stimulus is then

$$\sigma_{\text{total}} = \sqrt{\sigma_m^2 + \sigma_t^2} \quad (\text{A3})$$

Note that the delay only adds noise to the first measurement, not to the second one.

Given a measurement  $m$  (and dropping the subindex  $j$  for the moment), each log-contrast is associated with a likelihood  $L(k) = p(m|k)$ . The observer combines the likelihood with prior information about stimulus contrast. The prior probability,  $p(k)$ , is normal with mean  $\mu_p$  and standard deviation  $\sigma_p$ . With the likelihood and the prior, the observer computes the posterior probability given the measurement

$$p(k|m) = \frac{p(m|k)p(k)}{p(m)}, \quad (\text{A4})$$

where  $p(m) = \int p(m|k)p(k) dk$ . From the posterior probabilities, the model observer computes the probability that stimulus 1 had a higher contrast than stimulus 2:

$$p(k_1 > k_2 | m_1, m_2) = \int_{-\infty}^{\infty} p(k'_1 | m_1) \int_{-\infty}^{k'_1} p(k'_2 | m_2) dk'_2 dk'_1 \quad (\text{A5})$$

Finally, the observer chooses the stimulus that had a greater probability of having a lower contrast (or higher luminance, as the stimuli were decrements). That is, counting the number of times the observer chooses Stimulus 2 over Stimulus 1, the response  $y$  is

$$y(m_1, m_2) = \begin{cases} 1 & \text{if } p(k_1 > k_2 | m_1, m_2) < 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A6})$$

The psychometric function—the probability of choosing Stimulus 2 as a function of the stimulus values—is the expected value of the response conditional on the stimulus values, given by

$$p(\text{“}k_1 > k_2\text{”} | k_1, k_2) = \int \int y(m_1, m_2) p(m_1 | k_1) p(m_2 | k_2) dm_1 dm_2 \quad (\text{A7})$$

We tested two variants of the contrast model (Contrast Model 1 and 2 in the figures). In the first variant, we fixed the mean of the prior,  $\mu_p$ , to be the mean of all the reference stimulus values against both surrounds. In the second variant,  $\mu_p$  was fixed to the mean reference value in each particular condition. We assume that the observer learns the prior during the experiment and that the width of the learned prior

depends on the range of contrasts used. There were thus three variables that we varied: the standard deviations for the prior ( $\sigma_p$ ), the measurement ( $\sigma_m$ ), and time-dependent noise ( $\sigma_t$ ).

We made no attempt to fit the model to the raw data of individual observers. We are comparing the two models (contrast and reflectance models) in their ability to explain the main characteristics of the data, most importantly the subadditivity of memory and context effects. We do this by adjusting the parameters of each model to give a good fit to the average results over observers. This adjustment of parameters is described below.

The parameters  $\sigma_m$ ,  $\sigma_t$ , and  $\sigma_p$  were fit to the average observer data by minimizing the  $\chi^2$  error between the model and observer PSEs and thresholds. The raw  $\chi^2$  is

$$\chi^2 = \sum_i \frac{(o_i - e_i)^2}{\sigma_i^2}. \quad (\text{A8})$$

Here,  $e_i$  are the modeled PSE and threshold values in condition  $i$ , and  $o_i$  are the observed values, averaged over observers.  $\sigma_i$  is the standard deviation of the observed values.

The following is a (partial) list of some key assumptions in the contrast model:

- Observers base their judgments on noisy measurements that grow with log contrast and have normally distributed error.
- The errors are uncorrelated between trials and between the two stimuli.
- Holding a contrast measurement in memory adds noise (variability) to the measurement; measurement and memory noise are independent and their variances add.
- The observer uses prior information about stimulus contrasts when making contrast estimates. This prior has the form of a normal distribution in log contrast space, centered on the mean of the stimulus ensemble. The observer learns the prior quickly during the experiment.
- The observer computes posterior probability distributions for the two contrasts and chooses the one that more probably had a higher luminance (lower contrast).

### Reflectance model

The signal to the reflectance model is the log luminance  $l$  of the two centers and surrounds. The model observer has an internal model in which the luminance is the product of illumination  $I$  and reflectance  $R$ :  $L = IR$ . Or, as the model observer does the computations in a log space, log luminance is a sum of log illuminance and reflectance:  $l = i + r$ , where  $l = \log(L)$ ,  $i = \log(I)$ , and  $r = \log(R)$ . See Figure 7b for

illustration. The observer infers the reflectance of the two center patches and chooses the one that more probably had a higher reflectance. Details of the inference are explained below.

On each trial, the observer makes noisy measurements of the four luminances. The measurements grow with log luminance and have normal error:

$$\begin{aligned} m_{cj} &= l_{cj} + \epsilon_m \\ m_{sj} &= l_{sj} + \epsilon_m \end{aligned} \quad (\text{A9})$$

where  $\epsilon_m \sim N(0, \sigma_m)$ ;  $l_{cj}$  and  $l_{sj}$  are the center and surround luminances, respectively; and  $j = 1, 2$  is again the index for the stimulus. As in the contrast model, a delay between the first and the second stimulus adds noise to the measurement of the first stimulus. We assume only the measurement for the center becomes noisier, because both surrounds were visible throughout the trial in all conditions. The measurement for the first center patch in the delay conditions is then

$$m_{c1} = l_{c1} + \epsilon_m + \epsilon_t \quad (\text{A10})$$

where  $\epsilon_t \sim N(0, \sigma_t)$  is again the time-dependent noise. The total noise for the first center patch is

$$\sigma_{\text{total}} = \sqrt{\sigma_m^2 + \sigma_t^2}. \quad (\text{A11})$$

The measurement is a function of luminance. Luminance is a function of both illumination  $i$  and reflectance  $r$ , so the likelihood function is two-dimensional:  $L(i, r) = p(m | i, r)$ . The observer's task is to infer the values of the two center reflectances. There is obviously an infinite number of illuminance-reflectance combinations that would produce a given luminance. As the measurements  $m$  are a function of luminance, estimating the reflectance from the likelihood alone would be an ill-defined problem. To reduce the ambiguity, the observer combines the likelihood with priors for illumination and surface reflectance. The inference is more complicated than in the contrast model, but the basic idea is straightforward. The observer has access only to the noisy measurements  $m$ . From the measurements, the observer infers the values of the center reflectances  $r_{c1}$  and  $r_{c2}$  (see Figure 7b). Making the optimal inference requires computing the two-dimensional posterior distribution for (log) center reflectance, given the measurements

$$\begin{aligned} p(r_{c1}, r_{c2} | \vec{m}) &= \frac{p(r_{c1}, r_{c2}, \vec{m})}{p(\vec{m})} \\ &= \frac{\int \int \int p(i_1, i_2, r_{c1}, r_{c2}, r_s, \vec{m}) dr_s di_1 di_2}{p(\vec{m})}, \end{aligned} \quad (\text{A12})$$

where the four measurement are represented as a vector:  $\vec{m} = [m_{c1}, m_{c2}, m_{s1}, m_{s2}]$ . As only the center reflectances are of interest, the observer marginalizes



over the two illuminants and the surround reflectance (the triple integral in the above equation). The (total) joint probability above can be written out (with the help of the graph in Figure 7b):

$$\begin{aligned} p(i_1, i_2, r_{c1}, r_{c2}, r_s, \vec{m}) \\ &= p(i_1, i_2, r_{c1}, r_{c2}, r_s, m_{c1}, m_{c2}, m_{s1}, m_{s2}) \\ &= p(i_1)p(i_2)p(r_{c1})p(r_{c2})p(r_s)p(m_{c1}|I_1, r_{c1}) \\ &\quad p(m_{c2}|I_2, r_{c2})p(m_{s1}|i_1, r_s)p(m_{s2}|i_2, r_s) \end{aligned} \quad (\text{A13})$$

There are three prior distributions in this joint probability: The illumination prior  $p(i)$  (same prior for the two illuminants), the center reflectance prior  $p(r_c)$  (same prior for the two centers), and the surround reflectance prior  $p(r_s)$ .

From the posterior distribution, the observer computes the probability that the log reflectance of Stimulus 2 was greater than that of Stimulus 1:

$$p(r_{c2} > r_{c1} | \vec{m}) = \int \int p(r'_{c1}, r'_{c2} | \vec{m}) I_{r'_{c2} > r'_{c1}} dr'_{c1} dr'_{c2}, \quad (\text{A14})$$

where  $I_{r_{c2} > r_{c1}}$  is the indicator function. The model observer then chooses Stimulus 2 if it has a greater probability to have a higher reflectance:

$$y(\vec{m}) = \begin{cases} 1 & \text{if } p(r_{c2} > r_{c1} | \vec{m}) > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A15})$$

This model is computationally much heavier than the contrast model. For each condition, we picked seven stimulus values from the appropriate range and ran 50 simulated trials at each point. We then fit psychometric functions to these data to extract the PSE and threshold.

We adjusted the parameters of the model to find as close correspondence as possible to the average empirical data. First, we fixed the means of the three priors with the constraints that (a) the peak of the joint log illuminance-center reflectance prior corresponded to the mean of the log reference luminances and (b) the peak of the joint log illuminance-surround reflectance prior corresponded to the average log surround luminance. In other words, if  $\mu_i$  is the mean (and the mode of the normal) illuminance prior, the mean center reflectance prior  $\mu_{rc}$  was set so that  $\mu_i + \mu_{rc}$  equaled the average reference log luminance. Similarly, the mean surround reflectance prior  $\mu_{rs}$  was set so that  $\mu_i + \mu_{rs}$  equaled the average surround log luminance.

The other parameters in the model are the prior standard deviations  $\sigma_i$  (illuminance),  $\sigma_{rc}$  (center reflectance), and  $\sigma_{rs}$  (surround reflectance); the measurement noise parameter  $\sigma_m$  and the time-dependent noise parameter  $\sigma_t$ . These parameters were adjusted by trial and error to find a close match to the empirical data. Instead of letting all three prior width parameters vary freely, we introduced a constraint: When projecting the two-dimensional prior on to the luminance axis, we required the extreme center luminance values and the two surround luminance values to have the same  $z$ -score; call the parameter that defines this score  $z$ . There were thus four free parameters in the model:  $\sigma_i$ ,  $z$ ,  $\sigma_m$ , and  $\sigma_t$ .

The following is a list of some of the key assumptions in the reflectance model:

- Observers infer the log reflectance of the center patches from noisy measurements that grow with log luminance and have normally distributed error.
- The errors are uncorrelated between trials and between different parts (center, surround) of the stimulus.
- Holding a measurement in memory adds normally distributed noise to it; measurement and time-dependent noise are independent and their variances add.
- Observers use prior information on surface reflectance and illumination when making the inference. Observers learn the range of log luminances (sum of log illuminance and reflectance) during the experiment and adjust their priors accordingly.
- Observers compute the posterior probability for the two center reflectances and choose the one that more probably had a higher reflectance

## Model comparison

The models have an unequal number of parameters and thus unequal degrees of freedom. To compare the models, we computed a  $\chi^2$  statistic for each. We first computed the “raw”  $\chi^2$  value according to Equation A8. The final, reduced  $\chi^2$  value was the raw value normalized by the degrees of freedom  $v$ :

$$\chi_N^2 = \frac{\chi^2}{v}. \quad (\text{A16})$$