

Constancy of visual working memory of glossiness under real-world illuminations

Hiroyuki Tsuda

Graduate School of Human and Environmental Studies,
Kyoto University, Kyoto, Japan



Jun Saiki

Graduate School of Human and Environmental Studies,
Kyoto University, Kyoto, Japan



Glossiness is a surface property of material that is useful for recognizing objects and spaces. For glossiness to be effective across situations, our visual system must be unaffected by viewing contexts, such as lighting conditions. Although glossiness perception has constancy across changes in illumination, whether visual working memory also realizes glossiness constancy is not known. To address this issue, participants were presented with photo-realistic computer-generated images of spherical objects and asked to match the appearance of reference and test stimuli in relation to two dimensions of glossiness (contrast and sharpness). By comparing performance in terms of the match between perception and memory, we found that both features were well recalled, even when illumination contexts differed between the study and test periods. In addition, no correlation was found between recall errors related to contrast and sharpness, suggesting that these features are independently represented, not only in perception, as previously reported, but also in working memory. Taken together, these findings demonstrate the constancy of glossiness in visual working memory under conditions of real-world illumination.

Introduction

In visual cognition research, the attribute of glossiness has been relatively little studied, even though it can provide valuable information on a wide range of cognitive activities, from perception to action. When grasping an object, evaluating the quality of food, or judging the navigability of terrain, the visual appearance of a surface, including its gloss and roughness, is essential. For example, the way we grasp objects is directly related to how slippery we expect them to be, a characteristic that can often be estimated using the appearance of glossiness (e.g., Adams, Kerrigan, & Graf, 2016). Importantly, the perception of glossiness

depends not only on the reflectance properties of a surface, but also on the conditions of illumination (the appearance of a given object can change dramatically, depending on the lighting context, as shown in Figure 1). Because visual input to the eyes confounds reflectance and illumination, and there is no certain way of distinguishing appearance due to surface from those caused by illumination (i.e., the inverse optics), visually estimating glossiness is a computationally challenging problem (Chadwick & Kentridge, 2015; Fleming, 2014).

Despite the difficulty, previous studies have shown that people adjust to changes of illumination, achieving significant, although not perfect, glossiness constancy (Dror, Willsky, & Adelson, 2004; Fleming, Dror, & Adelson, 2003); other studies have detailed the limitations of this perception (Motoyoshi & Matoba, 2012; Olkkonen & Brainard, 2010, 2011). These studies attribute their findings primarily to the characteristics of perception itself. However, to compare or match reference and test stimuli, as glossiness perception tasks generally do, one must compare current sensory input with the stored representation of an object's appearance, as seen under a previous source of illumination. For this reason, glossiness constancy may be closely interconnected with an individual's ability to maintain information in the face of interference (i.e., a shift in illumination), which is a hallmark of visual working memory (VWM). As there has been no systematic investigation of how perception and memory of glossiness may differ in terms of the precision and bias of the observer's response, it remains unclear whether the glossiness information held in working memory is robust to changes of illumination. This issue is important, not just in defining the limits of glossiness constancy, but also in assessing the capacity of working memory of ecologically realistic stimuli under natural situations, an issue that has not been adequately researched in VWM studies (Orhan & Jacobs, 2014).

Citation: Tsuda, H., & Saiki, J. (2018). Constancy of visual working memory of glossiness under real-world illuminations. *Journal of Vision*, 18(8):14, 1–16, <https://doi.org/10.1167/18.8.14>.

<https://doi.org/10.1167/18.8.14>

Received July 25, 2017; published August 28, 2018

ISSN 1534-7362 Copyright 2018 The Authors



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded From: <https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/937434/> on 08/28/2018

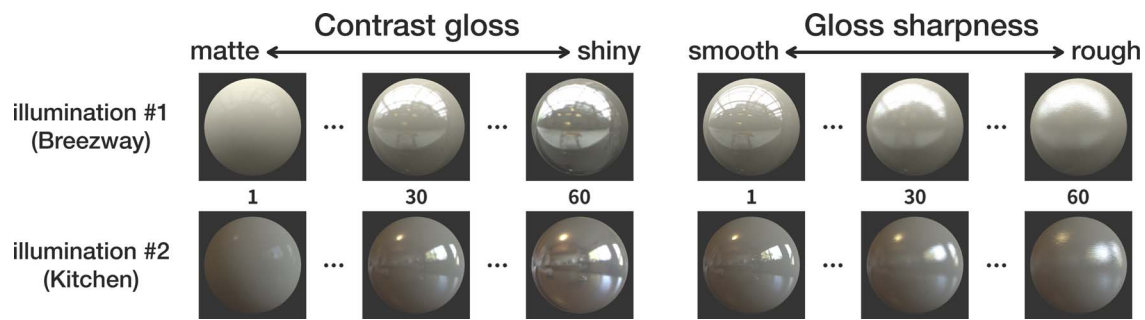


Figure 1. Stimulus images used in Experiments 1 and 2. Objects with different degrees of contrast gloss (left: matte to glossy) and gloss sharpness (right: smooth to rough) were shown. Images of each row were rendered under different illuminations (an outdoor and indoor scene).

Constancy has long been studied in the context of color and lightness perception. Relatively little is known about the perception of glossiness and its constancy, presumably because it is difficult to generate and control glossy stimuli (the ones we encounter in daily life are far more complex than simple colored patches), and also because glossiness perception has complicated nonlinearities between the physical property and sensation (Chadwick & Kentridge, 2015). Advances in computer graphics have made it possible to generate realistic renderings of objects under real-world illumination conditions. Importantly, in an influential study, Ferwerda, Pellacini, and Greenberg (2001) have identified a psychophysically based gloss space. Based on the isotropic version of Ward's physically based light reflection model (Ward, 1992), Ferwerda et al. have introduced a two-dimensional space of glossiness, which is perceptually uniform and reflects our sensitivity to glossiness variations. Each dimension of the model represents a distinct aspect of surface glossiness: (a) contrast gloss: the perceived relative brightness of specularly and diffusely reflecting areas, and (b) distinctness-of-image gloss: the perceived sharpness of images reflected in a surface (Ferwerda et al., 2001). Several glossiness perception studies, along with the current study, have employed this model to control glossy stimuli (see Figure 1 for the outcome of the renderings). We refer to distinctness-of-image gloss as "gloss sharpness" for the sake of conciseness and clarity. It is worth noting that contrast gloss and gloss sharpness are thought to be separate, independent channels for processing glossiness (Ferwerda et al., 2001). Moreover, there is no statistical dependence of matching errors between them (Fleming et al., 2003), suggesting that these are perceptually independent dimensions of glossiness.

The appearance of surface glossiness is the result of an interaction between the surface properties of an object (e.g., surface reflectance, surface geometry) and the illumination field that surrounds the object. For glossiness to be useful, the perception of glossiness must be robust to changes in the viewing context,

especially lighting conditions. Fleming et al. (2003) have shown that it is indeed robust. To prove this, they used a matching task, which presented the images of two spheres, generated under different illumination conditions. Participants adjusted the contrast gloss and gloss sharpness of the probe sphere to match the appearance of the reference sphere. Their matches were close to veridical under real-world illuminations (photographically captured from real-world scenes), but not under artificial illuminations (such as multiple point lights or Gaussian noise). Fleming et al. have proposed that the visual system takes advantage of the characteristics of natural illumination fields (such as statistical regularities in natural scenes), making it possible to reject unlikely image interpretations and leading to a reliable and accurate estimation of glossiness. Although glossiness constancy is not perfect and deteriorates in some situations, even under real-world illuminations (Motoyoshi & Matoba, 2012; Olkkonen & Brainard, 2010), this inconsistency may be a natural consequence of the computational difficulty of separating surface properties from other components. If so, it parallels numerous other studies that have documented the limits of color and lightness constancy (Foster, 2011; Gilchrist, 2006).

The primary aim of the current study is to ascertain whether glossiness constancy, as demonstrated by previous studies, can be extended to a situation in which matching is based on a stored representation in VWM. This is a challenging task because the appearance of objects changes dramatically, depending on illumination conditions. It is therefore necessary to form and maintain an illumination-independent representation in memory, successfully retrieve it, and then make a match that counterbalances the discrepancy in illumination between study and test. In previous studies of glossiness constancy, the memory load of each participant has been minimized by presenting the reference and test images simultaneously (Fleming et al., 2003; Olkkonen & Brainard, 2010) or by alternating images repeatedly until the participant responds (Motoyoshi & Matoba, 2012). Since the involvement of

VWM in glossiness matching has not been explicitly addressed in the literature, the extent to which perceptual and memory matching of glossiness may differ remains unclear. This study also addresses the issue of glossiness memory and the independence of glossiness dimensions. Although basic visual features, such as orientation and color, can be maintained independently in VWM (Bays, Wu, & Husain, 2011), while the contrast and sharpness dimensions of glossiness are perceptually independent (Ferwerda et al., 2001; Fleming et al. 2003), it does not necessarily follow that both dimensions of glossiness are independent in VWM. Since it must be difficult for VWM to retain all of the glossiness cues available in online processing, the VWM glossiness-encoding process is likely to create a more compact representation by selecting and transforming perceptual cues. During that process, contrast and sharpness could contaminate each other. Such contamination might also occur during the decay and retrieval processes. Therefore, in contrast to perceptual matching, the two dimensions may not be independent in memory matching.

To address these issues, we have employed a continuous report (feature matching) paradigm, commonly used in recent VWM studies, to measure the precision of memory representations (Zhang & Luck, 2008). The paradigm requires participants to adjust a test probe feature so that it matches, as closely as possible, a studied sample. The procedure is basically the matching paradigm used in glossiness perception studies. However, there are two critical differences between perceptual matching tasks and our memory task: (a) our (memory) task depends on memory, because the matching is carried out after the offset of a study stimulus, followed by a 1-s delay—the study stimulus is also masked, to eliminate the contribution of sensory memory; (b) we used many more images as matching options in order to obtain fine-grained estimates of matching precision. In previous studies, the number of levels of stimulus glossiness (or the number of matching options) has been relatively low, at 10 to 11 levels (Fleming et al., 2003) or eight (Motoyoshi & Matoba, 2012; but see Olkkonen & Brainard, 2011, for a more fine-grained measurement with a staircase procedure). This means that the differences between glossiness levels at each step were relatively large, potentially causing some thresholding effects in relation to the matching. In the current study, we have obtained matching precision in relation to both perception and memory; the degree of memory constancy has been evaluated using perceptual baselines. Using a more fine-grained approach to measure matching precision has been critical to the aims of the present study. For this reason, our tasks involved images representing 60 levels of glossiness.

Experiment 1 began by examining how matching performance (precision and bias) differed for perceptual and memory matching. Each trial required participants to match either contrast gloss or gloss sharpness. In some trials, the lighting conditions were the same for both reference and test stimuli; in others, the form of illumination changed. To evaluate the robustness of VWM to the shift in illumination, the precision of the memory matching was normalized using the corresponding perceptual baseline. If we were to observe glossiness constancy in memory, such a measure will not be affected by changes of illumination (see details below). In Experiment 2, we further explored how glossiness is represented in VWM by testing the independence of contrast gloss and gloss sharpness. Although independence has been suggested for perceptual matching, memory matching may not exhibit independence, due to the limitations of storage capacity or the retrieval process. If both dimensions can be independently retrieved from memory, it follows that recall of a feature in one dimension does not depend on the other dimension; there should thus be no correlation of errors between contrast and sharpness.

Experiment 1

Matching precision and bias were analyzed using perceptual (simultaneous) and memory (delayed) matching tasks. In addition to the main experiment (Experiment 1a), we also ran control experiments, using a slightly different procedure (Experiment 1b through d). The method and results of Experiment 1a are described below.

Method

Participants

A total of fifty-four Kyoto University students (18–22 years old) participated in the experiment for monetary compensation (1,500 Japanese Yen, or approximately 14 US dollars). Three groups of fourteen participants were included in Experiments 1a through c, and twelve participants were assigned to Experiment 1d. All of the participants reported normal color vision and had normal or corrected-to-normal visual acuity. Each participant provided written informed consent before participating in the experiment. All experiments were conducted in accordance with guidelines issued by the Ethics Committee of Kyoto University and the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Stimuli and apparatus

The stimulus images were three-dimensional spheres, computer-rendered using RADIANCE software (Ward, 1994). The outcome of the renderings is shown in Figure 1. Spheres were spatially uniform in their surface properties (a uniform gray color across the surface, without texture, and a smooth surface without bumps). To achieve different levels of contrast gloss and gloss sharpness, the software used the isotropic Ward model (Ward, 1992). This model represents surface reflectance properties using three parameters: The ρ_d controls the diffuse component (corresponding to the base color or “albedo” of a surface); ρ_s controls the specular reflectance component (contrast gloss); and α controls the microscopic surface roughness (gloss sharpness). As we were interested in glossiness, surface color was held constant by fixing the diffuse parameter ρ_d at red = 0.043, green = 0.037, and blue = 0.028, which yielded a gray color.

Spheres with different levels of contrast gloss were rendered by varying the specular reflectance parameter (ρ_s) from 0.000813 to 0.1 in 60 steps, with the sharpness parameter (α) held constant at 0.02. Similarly, spheres with different levels of gloss sharpness were rendered by varying the sharpness parameter (α) from 0.00167 to 0.1 in 60 steps, with the specular reflectance parameter (ρ_s) held constant at 0.02. The parameters were stretched nonlinearly to make each step perceptually equal, in accordance with the psychophysically based model of glossiness perception proposed by Ferwerda et al. (2001).¹ The range of parameters was comparable with previous studies on glossiness constancy (Doerschner, Boyaci, & Maloney, 2010; Ferwerda et al., 2001; Olkkonen & Brainard, 2010).

The spheres were rendered under two different real-world illuminations. We used indoor and outdoor high-dynamic range photographs as light fields to represent two distinct illumination contexts, which were derived from a database designed by Debevec (1998). The light fields were called “Kitchen” and “Overcast Breezeway” in the database. The light field images were treated as panoramic environmental maps to generate images of a glossy sphere reflecting the surrounding scene. As seen in Figure 1, although the color of the spheres was constant, spheres under the Breezeway illumination (upper row) were lighter than those under the Kitchen illumination (lower row), due to a difference in the distribution of pixel intensities of the two illuminations (i.e., the Kitchen scene was darker than the Breezeway scene). We also note that since images were not matched in mean luminance, the current task could not be performed by mere lightness matching, and the matching of glossiness was necessary. Backgrounds of the rendered images (the Kitchen or Breezeway scene behind the sphere) were cropped, since the role of context is usually weak for the perceived glossiness

(Fleming et al., 2003; but see Hansmann-Roth & Mamassian, 2017). In the task, the spheres were presented alone against a dark gray background.

A total of 240 images were created, half of which were rendered under Kitchen illumination and the other half under Breezeway illumination. Of the 120 images under each illumination condition, the contrast gloss varied in 60 images and the gloss sharpness varied in the other 60 images (Figure 1). The stimuli were displayed on a CRT monitor with a 120-Hz refresh rate, using MATLAB Psychtoolbox (Brainard, 1997; Pelli, 1997). Participants were seated approximately 57 cm from the monitor in a dark room, with their heads stabilized using a chin-rest.

Design and procedure

Participants performed a memory task and a perception task (Figure 2). In the memory task, each trial began by presenting a central white cross (0.62° diameter) for 500 ms against a dark background. A study stimulus was then presented for 1000 ms at the center of the display, followed by a mask (random block noise) for 200 ms and a blank interval of 1000 ms. After the blank, a test probe was presented, and participants adjusted a feature of the probe sphere to match the preceding sample, making it appear to be the same object as the sample sphere. Pressing the left or right arrow key decreased or increased the level of contrast gloss or gloss sharpness, depending on the block (we note that a dimension was adjusted on each trial in Experiment 1, but both dimensions were adjusted on each trial in Experiment 2). Participants pressed the spacebar to proceed to the next trial. In the perception task, after a central cross was presented for 500 ms, the reference and test spheres were presented simultaneously, side-by-side (with the distance between the centers of the spheres at 12.3°). Matching was then carried out, as in the memory task. Stimulus images subtended $7.9^\circ \times 7.9^\circ$ on the display. Participants were encouraged to respond as accurately as possible.

There were two illumination conditions. In the same-illumination condition, the reference and test spheres were rendered under the same illumination (Kitchen or Breezeway). In the different-illumination condition, the reference and test spheres had different illumination (either Kitchen to Breezeway or Breezeway to Kitchen, counterbalanced). Each matching dimension (contrast/sharpness) was tested in separate blocks of trials, and the illumination conditions (same/different) were randomized on a trial-by-trial basis. There were 50 trials in each condition, yielding a total of 200 trials in a task (perception/memory task). Of the 60 levels of contrast gloss or gloss sharpness, 50 were randomly selected and each was presented once in a randomized order. During the response period, the initial level of contrast gloss or

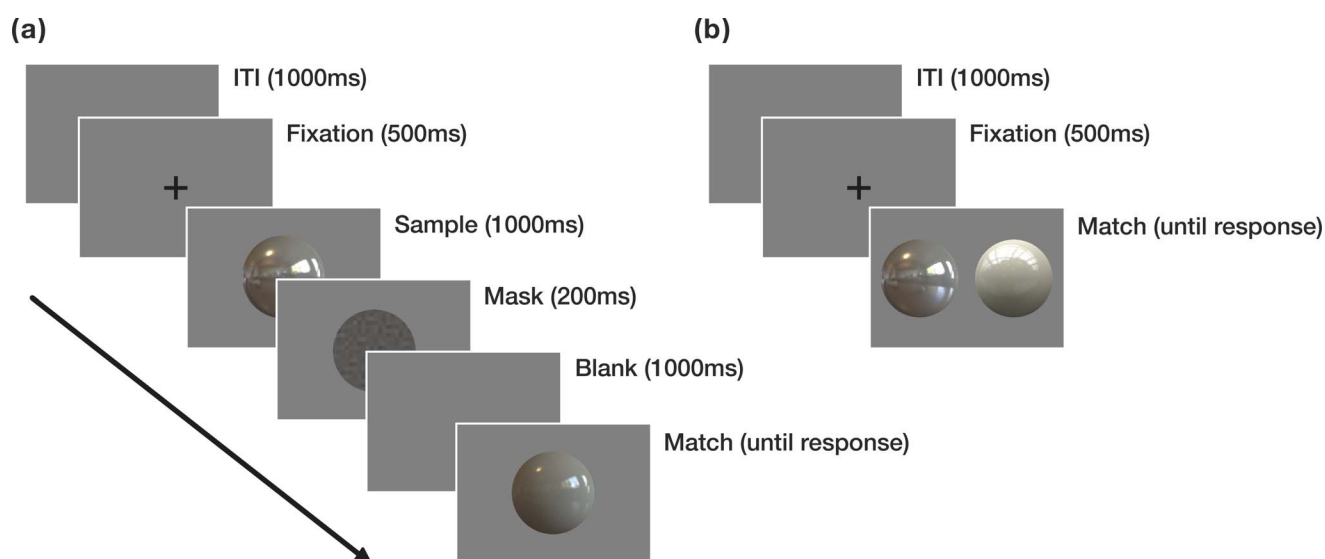


Figure 2. A schematic illustration of a single trial from Experiments 1 and 2; (a) In the memory task, after the presentation of a study stimulus, followed by a mask and a delay period, a test probe appeared and participants adjusted either the contrast gloss or gloss sharpness (Experiment 1a–d) or both dimensions (Experiment 2) using an unspeeded key press; (b) In the perception task, reference and test spheres appeared side-by-side. In the same-illumination condition, both spheres were rendered with the same illumination. In the different-illumination condition, the two spheres were illuminated differently (ITI = intertrial interval).

gloss sharpness of the probe sphere was set randomly in each trial. Half of the participants performed the perception task first, followed by the memory task. The order was reversed for the other participants.

Results and discussion

See Figure 3 for observed matching errors, organized by experimental conditions, in which results from all participants (and of all trials) are shown (see the figure caption for details). Performance was then assessed in terms of the precision and bias of the matching errors. Matching errors were first binned, according to the feature value of the to-be-tested items (higher bins corresponded to a glossier appearance for the contrast match, and a rougher appearance for the sharpness match). In each bin, precision (as measured by a standard deviation of error) and bias (as measured by the mean of errors) were calculated. Since there was no systematic difference in the distribution of errors within the same illumination conditions (B to B and K to K) and within the different illumination conditions (B to K and K to B), the data from B to B and K to K conditions were pooled when calculating the precision and bias measures (the same goes for different illumination conditions, B to K and K to B). The results are summarized in Figures 4 (precision) and in Figure 5 (bias). Tables 1 and 2 summarize the results of a 2 (Task: perception, memory) \times 2 (Illumination: same, different) \times 5 (Bin: 5 stimulus levels) repeated

measures ANOVA on precision or bias. The results of control experiments (Experiments 1b through d) and Experiment 2 are also included in the figures and tables. The discussion below first describes the results of Experiment 1a, and then those of Experiments 1b through d, which basically replicated the main experiment.

Matching precision

The matching precision of Experiment 1a is shown in the leftmost column in Figure 4. A larger *SD* corresponds to imprecise matching. The Task (perception/memory), Illumination (same/different), and Bin had main effects on the *SD* in both the contrast and sharpness matches (see Table 1). In the sharpness match, the interaction of Task and Illumination was also significant. The main effects of Task and Illumination (unsurprisingly) indicated that memory matches were less precise than perceptual matches, and also that matching was less precise when the reference and test had different forms of illumination.

The effect of Bin, namely that participants were less sensitive to higher contrast or sharpness, was unexpected, since we applied a reparameterization of the physical parameters of Ward's model (based on equations proposed by Ferwerda et al., 2001) to make each glossiness step perceptually uniform. Interestingly, similar patterns in the increase of errors can be seen in a previous study with experimental settings that were very similar to ours (Fleming et al. 2003). The shift in

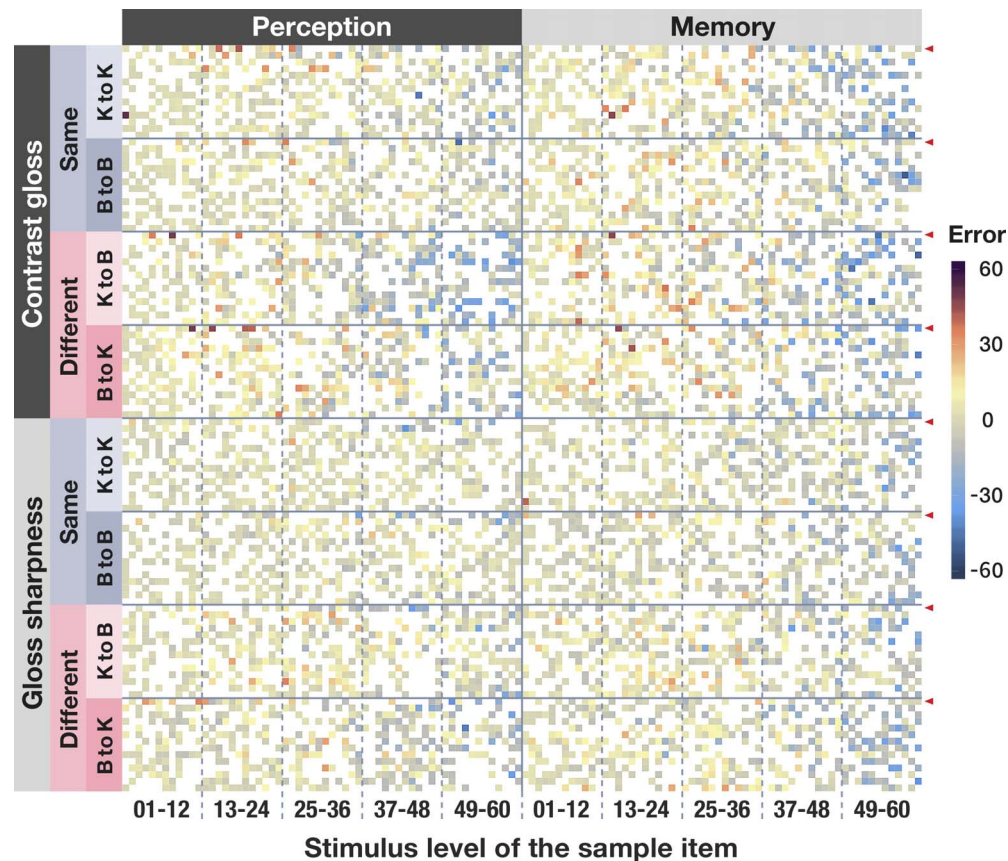


Figure 3. Matching errors in Experiment 1a. Each dot represents a trial of a participant, and the matching error in the trial is represented by color. Errors were calculated by subtracting the correct value from the participant's response. Results from all participants are shown in the figure, and the red arrows in the right indicate a group of rows, which is the result of a single participant. Positive errors (redder cells) mean that the adjusted glossiness of a probe sphere was shinier (than the sample stimulus) in contrast gloss matching, or rougher in gloss sharpness matching. Negative errors (bluer cells) mean that the adjusted glossiness of a probe sphere was more matte in contrast gloss matching, or smoother in gloss sharpness matching. Blank (white) cells are conditions not tested for that participant. Vertical dotted lines represent bin borders (see main text). Same = same illumination condition, Different = different illumination condition, B = breezeway illumination, K = kitchen illumination.

sensitivity could be attributed to a nonspecific response bias, but may also reflect a difference in measurement method (a magnitude estimate in Ferwerda et al., 2001, and perceptual matching in Fleming et al., 2003, and the present study). However, that discussion goes well beyond the scope of this article and should be an objective for future research.

The interaction of Task and Illumination on *SD* shows that the difference in matching precision between perception and memory was less salient when there was a change of illumination (Figure 4: black and gray dotted lines are relatively close to each other). This result supports the notion that memory matching is relatively robust to a shift in illumination (i.e., memory of glossiness was not impaired/interfered by the shift of illumination). We will address this issue in more detail below, in the Relative loss of memory precision section.

Matching bias

The matching bias of Experiment 1a is shown in the leftmost column in Figure 5 (also see Table 2 for the result of the statistical tests). The results clearly show a matching bias toward the mean of possible feature values. The central tendency bias is not surprising and may be a natural consequence of the procedure (the range of possible error changes, with the location of the sample feature value). More interesting here is the interaction between Bin and Task/Illumination. Whereas glossiness matching was relatively close to veridical in the perception-same condition, it was prone to bias in memory conditions, and also in the perception-different condition. This interaction can be explained by considering the increased variability of response in the perception-different condition and memory conditions, which will lead to the increase in the central tendency. Another possibility is that in the perception-same condition, glossiness cues that were

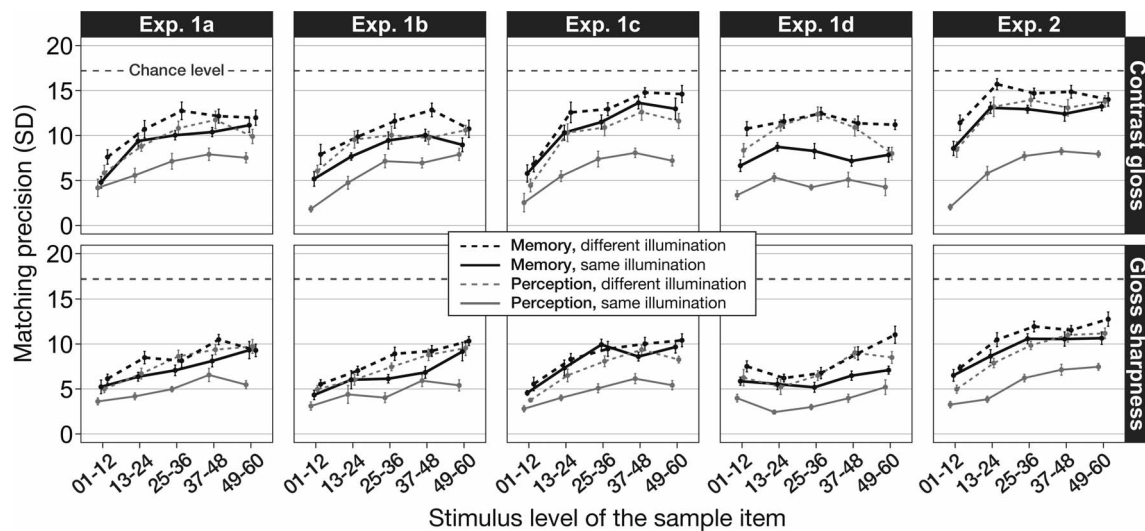


Figure 4. Summary of matching precision (standard deviation of error) of Experiments 1a–d and 2. Performance was binned according to the feature value of the to-be-tested item; larger bins corresponded to trials with a glossier or rougher sample. The gray lines indicate a perceptual match and the black lines indicate a memory match. The solid lines indicate the same illumination conditions, and dotted lines represent different illumination conditions. The horizontal dotted lines at the upper part of each plot indicate the chance level. The error bars are ± 1 SEMs.

specific to one form of illumination were available for veridical matching, which may have contributed to the higher matching accuracy in that condition.

Relative loss of memory precision

In order to quantify the robustness of glossiness memory with a shift in illumination, a precision of memory match was normalized using the corre-

sponding perceptual baseline. As described above, the precision of the perceptual match (baseline sensitivity) depended on whether there was a shift in illumination, and also on the location of the feature value of the reference item. To fairly compare the precision of memory matching between conditions, the shift in baseline performance had to be taken into account. The normalized memory score, termed “relative loss of memory precision,” was defined as

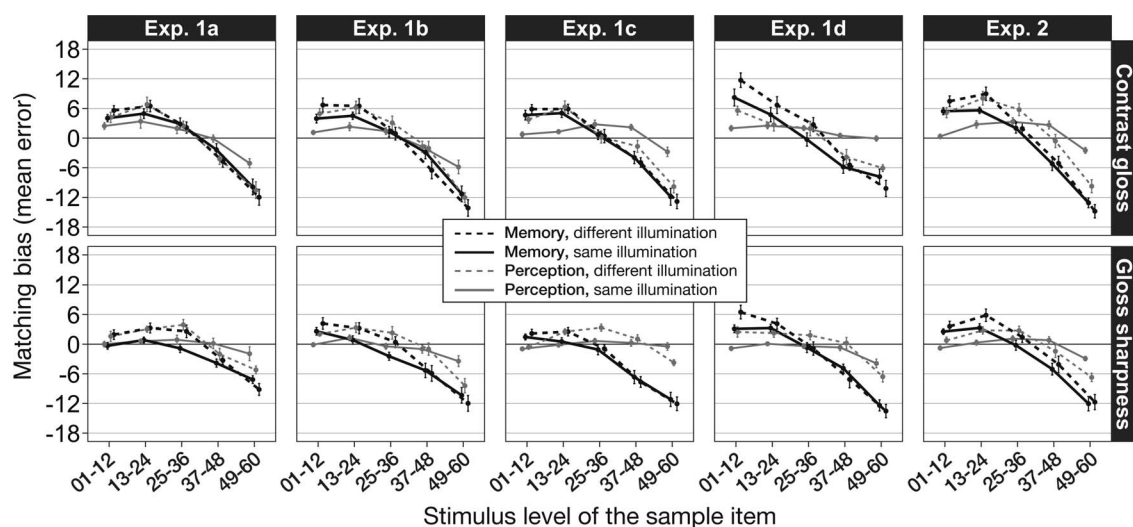


Figure 5. Summary of the matching bias (mean of errors) of Experiments 1a–d and 2. Performance was binned according to the feature value of the to-be-tested item; larger bins corresponded to trials with glossier or rougher samples. The gray lines indicate a perceptual match and the black lines indicate a memory match. The solid lines indicate the same illumination conditions, and the dotted lines indicate different illumination conditions. The error bars are ± 1 SEMs.

Source	Contrast gloss			Gloss sharpness		
	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Experiment 1a						
Task	13.8	0.0026	0.51	8.05	0.014	0.38
Illumination	36.7	< 0.001	0.74	34.9	< 0.001	0.73
Bin	42.0	< 0.001	0.76	24.5	< 0.001	0.65
Task × illumination	1.79	0.20	0.12	9.15	0.0010	0.41
Task × bin	1.03	0.40	0.07	1.22	0.32	0.09
Illumination × bin	0.82	0.52	0.06	1.28	0.29	0.09
Task × illumination × bin	0.62	0.65	0.05	2.01	0.11	0.13
Experiment 1b (trial order controlled)						
Task	13.8	< 0.001	0.52	10.1	0.0072	0.44
Illumination	72.9	< 0.001	0.85	26.6	< 0.001	0.67
Bin	30.4	< 0.001	0.70	28.6	< 0.001	0.69
Task × illumination	2.56	0.13	0.16	1.64	0.22	0.11
Task × bin	1.48	0.22	0.10	1.17	0.34	0.08
Illumination × bin	0.61	0.65	0.05	2.13	0.09	0.14
Task × illumination × bin	0.69	0.61	0.05	1.09	0.37	0.08
Experiment 1c (avoid image-based match)						
Task	40.5	< 0.001	0.76	76.9	< 0.001	0.86
Illumination	52.8	< 0.001	0.80	50.4	< 0.001	0.80
Bin	44.4	< 0.001	0.77	31.4	< 0.001	0.71
Task × Illumination	7.87	0.015	0.38	31.6	< 0.001	0.71
Task × bin	0.65	0.63	0.05	1.92	0.12	0.13
Illumination × bin	0.94	0.45	0.07	0.63	0.64	0.05
Task × illumination × bin	0.35	0.85	0.03	1.50	0.22	0.10
Experiment 1d (use of other light probes)						
Task	43.0	< 0.001	0.80	80.1	< 0.001	0.88
Illumination	162	< 0.001	0.94	116	< 0.001	0.91
Bin	5.76	< 0.001	0.34	15.4	< 0.001	0.58
Task × illumination	3.72	0.080	0.25	6.63	0.026	0.38
Task × bin	2.33	0.071	0.17	0.56	0.69	0.05
Illumination × bin	2.65	0.046	0.19	3.18	0.022	0.22
Task × illumination × bin	1.33	0.27	0.11	1.09	0.37	0.09
Experiment 2 (simultaneous match)						
Task	86.3	< 0.001	0.84	70.7	< 0.001	0.81
Illumination	136	< 0.001	0.89	82.3	< 0.001	0.83
Bin	26.7	< 0.001	0.61	39.8	< 0.001	0.70
Task × Illumination	61.7	< 0.001	0.78	24.0	< 0.001	0.59
Task × bin	3.26	0.017	0.16	1.22	0.31	0.07
Illumination × bin	1.23	0.30	0.07	1.80	0.14	0.09
Task × illumination × bin	0.79	0.54	0.04	0.57	0.68	0.03

Table 1. Summary of ANOVA results on matching precision (*SD*) in Experiments 1a–d and 2. *Notes:* Significant (< 0.05) *p* values are written in bold.

follows:

Relative loss of memory precision

$$= 1 - \frac{\|SD_{\text{memory}} - SD_{\text{chance}}\|}{\|SD_{\text{perception}} - SD_{\text{chance}}\|}$$

where SD_{memory} and $SD_{\text{perception}}$ are the standard deviation of error obtained in the memory or perception task. $\|\cdot\|$ denotes taking the absolute. Chance level (SD_{chance}) was derived from a numerical simulation, in which a simulated observer had a

uniform function to respond to each trial (random guessing). We ran the simulation 10,000 times and obtained $SD_{\text{chance}} = 17.17$ by the maximum likelihood estimate of the output. The index was 0 when there was no difference in matching precision between perception and memory (perfect memory); the index increases as that difference increases, reaching 1 when memory precision is equal to chance (a complete loss of memory). The index therefore provides a reasonable and interpretable measure of memory perfor-

Source	Contrast gloss			Gloss sharpness		
	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Experiment 1a						
Task	0.35	0.56	0.03	9.38	0.0091	0.41
Illumination	1.04	0.33	0.07	3.12	0.10	0.19
Bin	77.6	< 0.001	0.86	55.5	< 0.001	0.81
Task × illumination	0.09	0.76	0.01	1.11	0.31	0.07
Task × bin	2.52	0.052	0.16	3.06	0.024	0.19
Illumination × bin	9.22	< 0.001	0.42	13.6	< 0.001	0.51
Task × illumination × bin	1.52	0.21	0.11	0.69	0.59	0.05
Experiment 1b (trial order controlled)						
Task	1.00	0.33	0.07	4.42	0.055	0.25
Illumination	0.01	0.91	0.00	2.90	0.11	0.18
Bin	79.0	< 0.001	0.85	67.1	< 0.001	0
Task × illumination	1.12	0.31	0.07	0.45	0.51	0.03
Task × bin	4.27	< 0.001	0.24	5.40	0.001	0.29
Illumination × bin	16.7	< 0.001	0.56	12.7	< 0.001	0.49
Task × illumination × bin	2.10	0.094	0.13	1.50	0.21	0.10
Experiment 1c (avoid image-based match)						
Task	4.46	0.054	0.25	46.9	< 0.001	0.78
Illumination	2.43	0.143	0.15	1.53	0.23	0.10
Bin	86.0	< 0.001	0.86	53.3	< 0.001	0.80
Task × illumination	1.25	0.28	0.08	0.70	0.41	0.05
Task × bin	18.4	< 0.001	0.58	60.4	< 0.001	0.82
Illumination × bin	11.0	< 0.001	0.45	4.61	< 0.001	0.26
Task × illumination × bin	5.84	< 0.001	0.31	2.44	0.058	0.15
Experiment 1d (use of other light probes)						
Task	0.17	0.68	0.01	8.36	0.014	0.43
Illumination	0.00	0.97	0.00	2.56	0.13	0.18
Bin	66.4	< 0.001	0.85	90.2	< 0.001	0.89
Task × illumination	6.25	0.029	0.36	3.26	0.098	0.22
Task × bin	24.1	< 0.001	0.68	45.8	< 0.001	0.80
Illumination × bin	15.5	< 0.001	0.58	7.98	< 0.001	0.42
Task × illumination × bin	0.90	0.47	0.07	2.07	0.10	0.15
Experiment 2 (simultaneous match)						
Task	14.2	0.0015	0.45	2.28	0.15	0.11
Illumination	1.75	0.20	0.09	3.31	0.08	0.16
Bin	108	< 0.001	0.86	61.5	< 0.001	0.78
Task × illumination	0.14	0.70	0.00	5.65	0.029	0.24
Task × bin	32.3	< 0.001	0.65	32.2	< 0.001	0.65
Illumination × bin	15.4	< 0.001	0.47	7.16	< 0.001	0.29
Task × illumination × bin	4.68	0.0021	0.21	2.17	0.08	0.11

Table 2. Summary of ANOVA results on matching bias in Experiments 1a–d and 2. *Notes:* Significant (< 0.05) *p* values are written in bold.

mance, controlling the shift in perceptual sensitivities. The index is also useful for comparing memory performance across experiments that employ a different type of stimulus or procedure, since it cancels out the difference in the absolute levels of matching precision that can vary depending on the stimulus/task.

The results of the index are summarized in Figure 6 (see also Table 3 for the results of the statistical tests). The effect of stimulus level was not significant,

suggesting that the index was independent of the feature value of the to-be-tested item. In addition, the effect of Illumination on the index was not significant, which suggests that memory matching was not impaired by the shift in illumination. This result supports the view that there were few, if any, additional costs introduced by the shift of illumination in terms of how well glossiness information could be recalled from memory, suggesting that memory of glossiness was robust against changes in illumination.

Source	Contrast gloss			Gloss sharpness		
	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Experiment 1a						
Illumination	0.33	0.57	0.02	2.45	0.14	0.15
Bin	2.34	0.06	0.15	2.39	0.06	0.15
Illumination \times bin	0.45	0.76	0.03	1.94	0.11	.13
Experiment 1b (trial order controlled)						
Illumination	0.36	0.55	0.02	2.51	0.13	0.16
Bin	1.11	0.36	0.07	2.73	0.0038	0.17
Illumination \times bin	0.72	0.58	0.05	1.19	0.32	0.08
Experiment 1c (avoid image-based match)						
Illumination	0.86	0.36	0.06	10.2	0.0071	0.43
Bin	4.07	0.0060	0.23	3.60	0.015	0.21
Illumination \times bin	0.13	0.97	0.01	0.65	0.62	0.04
Experiment 1d (use of other light probes)						
Illumination	0.00	0.97	0.00	0.10	0.75	0.00
Bin	1.44	0.23	0.11	1.13	0.35	0.09
Illumination \times bin	1.58	0.19	0.12	1.21	0.32	0.09
Experiment 2 (simultaneous match)						
Illumination	28.1	< 0.001	0.62	2.17	0.15	0.11
Bin	1.12	0.35	0.06	1.52	0.20	0.08
Illumination \times bin	1.18	0.32	0.06	0.10	0.98	0.00

Table 3. Summary of ANOVA results on relative loss of memory precision in Experiments 1a–d and 2. Notes: Significant (< 0.05) *p* values are written in bold.

Experiments 1b, 1c, and 1d

In Experiments 1b through d, we have addressed some possible artifacts of Experiment 1a that could affect the generalizability of the findings. We have first described the purpose and method of each control experiment, and then discussed the results.

Experiment 1b

We addressed the effect of trial order. In Experiment 1a, the same and different illumination conditions were tested in randomized orders, so that participants could not predict whether the illumination would change in each trial. The fact that participants were required to prepare for both conditions forced them to use the same strategy to encode glossiness. Although retaining glossiness information in memory could be relatively easy when the matching was performed in the same illumination conditions, our procedure could have artificially induced the same level of matching performance across conditions (casting doubt on the validity of the view that the relative loss of memory performance was not impaired by the shift in illumination). In such a case, it would follow that, when upcoming illumination was predictive (with the same and different conditions tested in separate blocks of trials), the advantage, if it existed, of performing the match with the same type of illumination (or being disadvantaged by a shift in illumination) would be observed. The experimental procedure was identical to that of Experiment 1a, except that the trial order was altered so that the same and different illumination conditions were tested in separate blocks of trials.

Experiment 1c

We addressed possible confounds using image matching. In Experiment 1a, the participants' matching performance was most precise and accurate in the perception-same condition, which could have resulted from exact image matching. In an extreme case, participants might have performed the match based on the intensity of a local spot, without taking into

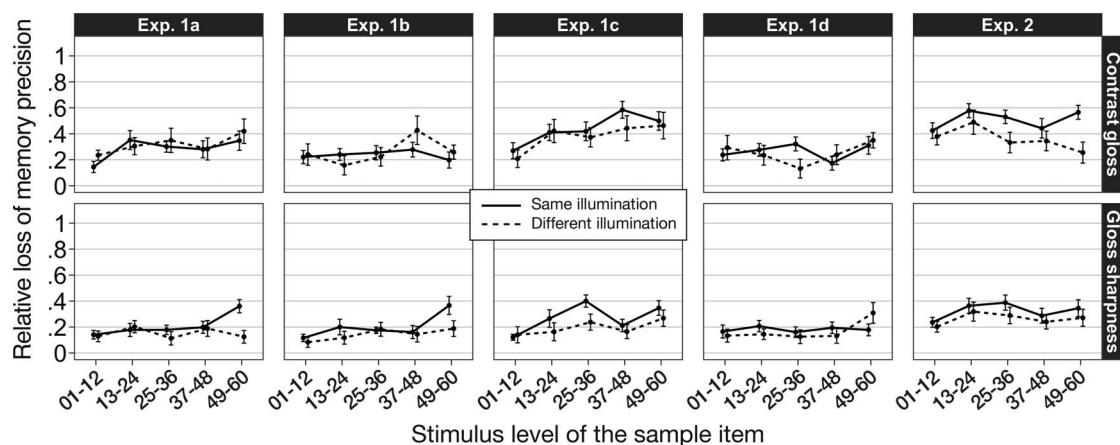


Figure 6. Summary of the relative loss of memory precision in Experiments 1a–d and 2 (see the main text for a definition of the index). Performance was binned according to the feature value of the to-be-tested item; larger bins corresponded to trials with glossier or rougher samples. Solid lines indicate the same illumination condition, and dotted lines indicate different illumination conditions. The error bars are ± 1 SEMs.

consideration the glossiness itself. The participants' matching performance in the perception-same condition could have been distorted by the use of such a strategy. To prevent this possibility, images were presented in a different way in Experiment 1c. Specifically, one of the items (study or test) was flipped horizontally and also slightly rotated (4°) to diminish the utility of the exact image matching strategy. The procedure was identical to Experiment 1a except for this manipulation.

Experiment 1d

We also asked whether the findings of Experiment 1a could be reproduced when more than two lighting conditions were used. A potential concern in Experiment 1a was that, when there were only two kinds of lightings, participants could acquire and rely on some approximate mappings of glossiness from one illumination to the other. In Experiment 1d, we used six illuminations that were all different from those in Experiment 1a. Since the numbers of lighting combinations were increased from two to fifteen, the response-mapping strategy became cumbersome and less likely to be relied on. As it would have taken a long time to render hundreds of additional stimuli using RADIANCE software, we used Unreal Engine (<https://www.unrealengine.com>) to enable the real-time rendering of the stimulus, so that it would no longer be necessary to prerender the images. Although the engine adopts a different lighting model—a more physically based BRDF (Karis, 2013)—it can control surface reflectance properties (contrast and sharpness) that are similar to Ward's model; the output of the rendering is also visually very similar. The procedure was identical to that in Experiment 1a, apart from using six different light probes (HDR photographs of three indoor and three outdoor scenes, obtained from <http://www.hdrlabs.com/sibl/archive.html>).

The results of Experiments 1b through d are summarized in Figures 4 through 6 and Tables 1 through 3. These results basically agree with the findings of Experiment 1a, although with some differences in detail. For matching precision, the main effects of Task and Illumination were replicated across experiments, and the interaction between Task and Illumination was also observed in some experiments and conditions. In Experiment 1d, the effect of Bin was relatively weak for the contrast match, which may reflect the difference in the lighting model (it could be perceptually more uniform). The pattern of bias was also qualitatively very similar across experiments. Finally, across the experiments, although there were some random fluctuations, the relative loss of memory precision index was not significantly worse when the illumination was different from when the illumination

was the same. Interestingly, in Experiment 1c, there was a main effect of Illumination on the index for the sharpness match (i.e., the index was larger for the same rather than different illumination conditions). This difference may have been caused by a shift in baseline performance, memory precision, or both (although there was no significant difference in matching precision across Experiments 1a through d). We will discuss the implications of this finding later, in the General discussion section.

Experiment 2

In Experiment 2, we tested the independence of the contrast and sharpness matches. Instead of performing each match separately for contrast and sharpness, as in Experiment 1, participants were required to match both dimensions in each trial.

Method

Participants, stimuli, and apparatus

Eighteen Kyoto University students (18–22 years old) participated in the experiment for monetary compensation (1,500 Japanese Yen, or approximately 14 US dollars). All participants reported normal color vision and had normal or corrected-to-normal visual acuity. Each participant gave written informed consent to participate in the experiment. All procedures were preapproved by the Ethics Committee of Kyoto University. The apparatus and viewing conditions were the same as in Experiment 1. To achieve two-dimensional matching, a set of images with 60 levels of contrast gloss and 60 levels of gloss sharpness, rendered under two types of illumination, was required (7,200 images in total). These were generated using RADIANCE, as in Experiment 1. Illuminations and the range of parameters of contrast gloss and gloss sharpness were the same as in Experiment 1.

Design and procedure

Participants performed a memory task and a perception task. Both tasks were modified so that participants were asked to match both contrast gloss and gloss sharpness in each trial by pressing left-right or up-down keys (key mapping was counter-balanced). Out of the stimulus images, 64 images were chosen as sample stimuli; by combining eight levels of contrast gloss and gloss sharpness, they covered the whole stimulus space at even intervals, tested in a randomized order. The illumination condition order (the same or different) was random in each trial, as in Experiment

1a. Participants performed 64 trials for each illumination condition, yielding a total of 128 trials for each task. Half of the participants performed the perception task first and the memory task second. The order was reversed for the other participants.

Results and discussion

When it came to matching precision, the experimental effects of Task and Illumination on *SD* were qualitatively similar to those in Experiment 1 (see Figure 4 and Table 1). However, when precision was compared across experiments 1a through d and 2, a significant difference in the *SD* was observed between Experiment 2 and Experiment 1. We conducted a 5 (Experiment) \times 2 (Task: perception, memory) \times 2 (Illumination: same, different) \times 5 (Bin: 5 stimulus levels) repeated measures ANOVA on *SD*, and found that there was a main effect of the Experiment on *SD*: both for the contrast match, $F(4, 68) = 9.98$, $p < .0001$, $\eta_p^2 = 0.37$; and for the sharpness match, $F(4, 68) = 7.85$, $p < 0.0001$, $\eta_p^2 = 0.32$. The interaction between the Experiment and the Task was also significant: both for the contrast match, $F(4, 68) = 3.27$, $p = 0.016$, $\eta_p^2 = 0.16$; and for the sharpness match, $F(4, 68) = 4.07$, $p = 0.0051$, $\eta_p^2 = 0.19$. A multiple comparison for the Experiment revealed that, whereas there was no significant difference in perceptual *SD*, the memory *SD* in Experiment 2 was significantly larger than those in Experiment 1a, 1b, and 1d (but not 1c), and this was true for both contrast and sharpness matching. When it came to the relative loss of memory performance index, there was a main effect of Illumination for contrast matching, but not for sharpness matching (Table 3). The matching bias results for Experiment 2 were qualitatively very similar to Experiment 1, and will therefore not be discussed here.

The significant decrease in memory precision in Experiment 2 suggests that it was costly to maintain both features simultaneously in memory. Perceptual precision was not significantly different from Experiment 1. Therefore, the decrease in matching precision was not an artifact associated with the difficulty of making a simultaneous response, but a reflection of the limits of the memory resource (Bays, Catalao, & Husain, 2009; Zhang & Luck, 2008). Another possibility is that memory was more prone to interference and decay in Experiment 2 because it took longer to match both features. When we calculated the correlation between RT and *SD*, however, the correlation coefficients ranged between -0.068 and 0.0066 , and the Bayes factors (BF_{01}) ranged between 3.3 and 3.4 (evidence of the absence of a correlation was calculated based on Wagenmakers, Verhagen, & Ly, 2016),

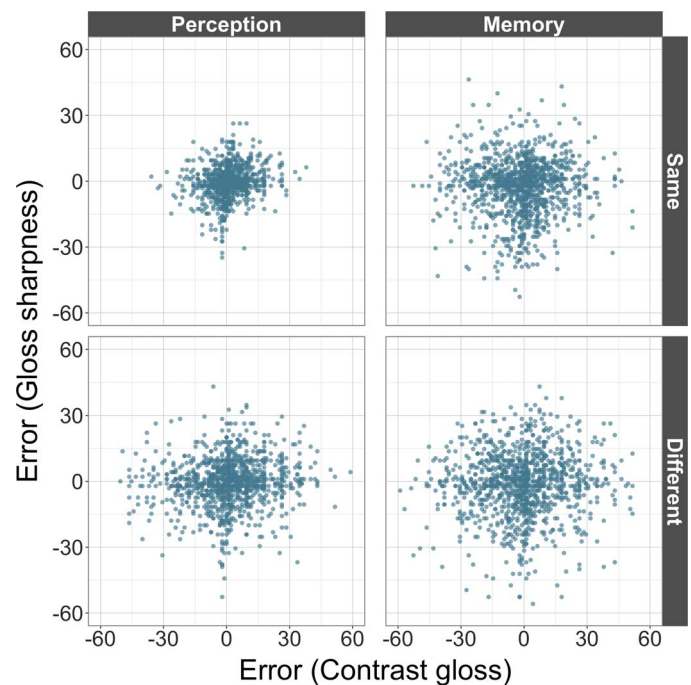


Figure 7. Distribution of matching errors (pooled across subjects) in Experiment 2. Each dot represents a trial.

suggesting that there was little or no effect of response time on matching precision.

We then calculated the correlation between matching errors related to contrast and sharpness to see if there may be any systematic influence of one dimension on the other in making the response. If, for example, reporting of a high contrast gloss (shiny stimuli) bias the evaluation of gloss sharpness toward a smoother surface, then there will be a negative correlation between the errors. Figure 7 shows the distribution of matching errors (each dot represents a trial, pooled across participants). The correlation coefficients and Bayes factors (BF_{01}) for each condition were as follows: Perception-Same, $r = 0.175$ ($SD = 0.122$), $BF_{01} = 2.74$; Perception-Different, $r = 0.090$ ($SD = 0.159$), $BF_{01} = 3.24$; Memory-Same, $r = 0.00242$ ($SD = 0.144$), $BF_{01} = 3.43$; Memory-Diff, $r = 0.0395$ ($SD = 0.128$), $BF_{01} = 3.39$. Across conditions, the current data favor the absence of a correlation. Note that the correlation was somewhat larger in the perception-same condition. This correlation, if it exists, could be an experimental artifact related to the light probes used in the experiment (e.g., some types of illumination induce larger matching errors than do others; Motoyoshi & Matoba, 2012). The correlation for perceptual matching in the different-illumination condition may have been weaker because the two bias directions canceled each other out. Importantly, the results of the memory matching indicated that participants were able to retrieve each feature without any systematic influence

from the other, suggesting that the independence of the dimensions was preserved in memory.

General discussion

The current study has been inspired by recent advances in material perception research, which show that human vision has a notable ability to estimate the surface properties of objects (such as glossiness) from variable viewing conditions (e.g., illumination), despite the computational difficulty involved (Chadwick & Kentridge, 2015; Fleming, 2014). This research has extended previous findings relating to glossiness constancy to the domain of working memory by comparing the extent to which matching performance differs between perception and memory. Our results have shown that the precision and bias of memory matching is close—and at times comparable—to its perceptual counterpart when there is an illumination shift between the reference and test stimuli. By comparing the normalized memory score by the perceptual baseline, or the relative loss of memory precision index, we observed that the memory performance in the different-illumination condition was not worse than that in the same-illumination condition. The fact that introducing an illumination shift did not impair participants' memory matching performance suggests that the memory of glossiness was robust to changes in illumination. These results were consistently observed across several experimental conditions, when the forthcoming illumination was and was not predictable (Experiment 1a and 1b), when the exact matching of images was prevented by flipping and rotating the stimulus (Experiment 1c), and when the number of lighting conditions was increased to confirm the generalizability of results and prevent a possible confound caused by participants learning to carry out a simple response mapping between illuminations (Experiment 1d). We also investigated the independence of matching performance in relation to contrast gloss and gloss sharpness by requiring participants to match both features simultaneously. This showed that there was little correlation of errors between features (Experiment 2), suggesting that both dimensions were independently processed, as has been suggested in relation to the perception of glossiness (Ferwerda et al., 2001; Fleming et al., 2003). In addition, memory precision in Experiment 2 decreased from that seen in Experiment 1, where only a single feature was maintained, suggesting that there was a cost to the simultaneous storage of contrast gloss and gloss sharpness in visual working memory.

Glossiness constancy in perception and memory

Memory plays an important role in constancy: To make a successive match, one needs to compare current sensory input with a stored representation seen previously (Allen, Beilock, & Shevell, 2011; Jin & Shevell, 1996). Although some degree of color constancy in memory has been reported using simple colored patches (Allen et al., 2011; Jin & Shevell, 1996; Olkkonen & Allred, 2014), it was somewhat surprising to discover an illumination-robust working memory of glossiness, given the visual complexity of the stimulus used in our task (physically based renderings of photo-realistic images using real-world illumination). The representation of glossiness is multidimensional in nature, meaning that no single feature can fully explain the perception of glossiness (for a review, see Chadwick & Kentridge, 2015). Although there are several different kinds of visual cues that can contribute to glossiness perception, including lightness, contrast, color, highlights, and specularly reflected mirror images (Chadwick & Kentridge, 2015; Hunter, 1937; Wendt, Faul, Ekroll, & Mausfeld, 2010), it is difficult to retain all online perceptual cues in working memory, given its limited capacity. Instead, the visual system may store some approximate measures of glossiness that remain roughly invariant across changes of illumination. Image statistics, such as the marginal statistics (variance, skew, and kurtosis) of luminance distribution, are a plausible candidate in this regard (Fleming & Bühlhoff, 2005; Motoyoshi & Matoba, 2012; Motoyoshi, Nishida, Sharan, & Adelson, 2007). Although marginal statistics may not fully account for the perception of glossiness (Anderson & Kim, 2009; Fleming, 2014; Landy, 2007; Olkkonen & Brainard, 2011), such an image heuristic approach may be more appealing to the memory domain because it allows efficient coding and storage of surface properties of an object in the capacity-limited system. Thinking about the difference between perception and memory in terms of the availability of cues and processing capacity, it seems worth investigating what type of cue is stored in memory to better characterize glossiness constancy.

We note some implications for the characteristics of stored representations of glossiness. While, in most cases, there was no significant difference in the relative loss of memory precision index between within- and across-illumination matching, in some conditions (sharpness matching in Experiment 1c and contrast matching in Experiment 2), the index was significantly larger in the same-illumination condition (see Figure 6 and Table 3). Little difference in the index between within- and across-illumination matching would be expected if glossiness memory were robust to the shift of illumination. How can we explain the increases in the index of memory performance in same-illumination

conditions, if they were not false positives? In the same-illumination matching, some extra glossiness cues that were specific to the specific type of illumination were available for matching, in addition to illumination-independent glossiness cues, while in the across-illumination matching, only the latter (illumination-independent cues) could be used to make a match. In fact, matching precision was better in the same-illumination condition than in the different-illumination condition for both perceptual and memory tasks, suggesting that illumination-specific glossiness cues contribute to making a match in same-illumination conditions. If both types of cues had the same rate of decay in memory, there would be no difference in the index under same and different illumination conditions. However, if we assume that illumination-specific information is more prone to decay (because, although it may contain rich information relating to the lower level visual properties of glossiness, it can be harder to retain in memory), an increase in the index in the same-illumination condition will be predicted, given that illumination-specific cues, which were available in the perceptual matching, were less accessible or lost in the memory matching, leading to an increase in the matching error. Although this assumption provides a plausible explanation of the results, we are not confident that the deviations in memory performance observed in some conditions reflect a genuine effect. If they do, it also remains unclear why a difference in performance was not consistently observed across experiments. Future work may be needed to clarify this issue.

Matching bias

Across experiments, there was a central tendency bias toward the middle of the possible feature values (Figure 5). Although the central tendency seems to be a natural consequence of the experimental procedure (the response was bounded in a certain range), the mere response bias alone cannot fully explain the data because, under certain conditions, matching was less susceptible to bias (perceptual matching was relatively close to veridical in the same-illumination condition). A possible explanation based on the discussion above is that the participants used illumination-specific cues to achieve a precise matching; these cues were not useful in different-illumination conditions (for either perceptual or memory matching). They were also less accessible for memory matching in the same-illumination condition. However, we note that the increase in the central tendency bias can also be explained by the increase in response variability, or another factor as described below, so that it is difficult to specify which is the most reasonable account for the results.

Matching bias was also reported in previous studies of glossiness perception (Doerschner et al., 2010; Fleming et al., 2003; Olkkonen & Brainard, 2010), where it was thought that a nonspecific response bias might have been more pronounced in matching tasks than in forced-choice tasks (Doerschner et al., 2010; Olkkonen & Brainard, 2010). However, bias in perception and memory may reflect an adaptive mechanism of visual cognition (Sims, Jacobs, & Knill, 2012; Wei & Stocker, 2017). In terms of the Bayesian framework of visual cognition, another interpretation of the results should also be considered. When the reliability of the sensory or memory signal was relatively weak (due to a shift in illumination or the decay of a stored representation), and if the prior distribution of glossiness values was not uniform but centered on the middle of the feature space, the response would be more biased to the mean, as shown in our experiments (see similar discussions on color memory in Olkkonen & Allred, 2014 and on visual working memory in general, Sims et al., 2012). Future work would benefit from more systematic investigations of matching bias in the perception and memory of glossiness.

Limitations and future directions

Before concluding, we would like to discuss some limitations of the current study. Some parameters that we did not manipulate in our experiments could affect the generalizability of our findings; these include the set size of the memory array, the duration of the blank interval, and the surface geometry of the objects. First, it remains unclear whether glossiness constancy will be stable when the memory load (set size) is increased. The fact that memory precision decreased when both features were maintained simultaneously in Experiment 2 suggests that increasing the number of objects puts more demands on the working memory. Given reports that the capacity of working memory affects color constancy (Allen et al., 2011), there may be some similar interaction between memory load and glossiness constancy. Second, we examined working memory over a short time span of a few seconds; it will be interesting to investigate the effects of longer durations on the durability of glossiness memory. It also may help to explain how different types of glossiness cues decay differently. Finally, it remains unclear whether our findings can be generalized to arbitrary shapes other than spheres. It is known that the surface geometry of objects affects glossiness perception (Marlow, Kim, & Anderson, 2012; Nishida & Shinya, 1998; Vangorp, Laurijssen, & Dutré, 2007). Shape also interacts with illumination in affecting glossiness perception (Olkkonen & Brainard, 2011). Therefore,

glossiness constancy in memory may sometimes be more challenging, depending on the surface geometry of the objects used.

Conclusion

While it is essential for the visual system to extract perceptual cues that have invariance to environmental variability, retaining that information in memory is also important for adaptive behaviors. The findings of the current study support the view that the visual working memory of glossiness is robust to shifts in illumination. Although it remains an open question how glossiness information is represented in memory and how it may differ from perception, the present study highlights the capacity of visual working memory by showing consistent results for glossiness constancy across the range of experimental conditions we have explored.

Keywords: visual working memory, material perception, gloss, roughness, illumination, constancy

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP25135719 and JP16H01672.

Research conducted by Hiroyuki Tsuda and Jun Saiki, Graduate School of Human and Environmental Studies, Kyoto University.

Commercial relationships: none.

Corresponding authors: Hiroyuki Tsuda; Jun Saiki.
Email: tsudahiroyuqi@gmail.com; saiki@cv.jinkan.kyoto-u.ac.jp.

Address: Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan.

Footnote

¹ Reparameterization was obtained by converting perceptually linear c – d parameters (Ferwerda et al., 2001) to Ward's physical parameters. Specifically, specular reflectance was computed such that $\rho_s = (c + \sqrt[3]{\rho_d/2})^3$. The value for ρ_d was obtained by mapping the RGB channels of ρ_d to the CIE Y dimension, as in a previous study (Fleming et al., 2003). Surface roughness was computed such that $\alpha = 1 - d$. Values of c and d were chosen to yield both ρ_s and ρ_d ranging from 0 to 0.1.

References

- Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2016). Touch influences perceived gloss. *Scientific Reports*, 6(1), 1–12, <https://doi.org/10.1038/srep21866>.
- 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(4), 1014–1021, <https://doi.org/10.1037/a0023257>.
- Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision*, 9(11):10, 1–17, <https://doi.org/10.1167/9.11.10>. [PubMed] [Article]
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10):7, 1–11, <https://doi.org/10.1167/9.10.7>. [PubMed] [Article]
- Bays, P. M., Wu, E. Y., Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, 49, 1622–1631, <https://doi.org/10.1016/j.neuropsychologia.2010.12.023>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 443–446.
- Chadwick, A. C., & Kentridge, R. W. (2015). The perception of gloss: A review. *Vision Research*, 109, 221–235, <https://doi.org/10.1016/j.visres.2014.10.026>.
- Debevec, P. E. (1998). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proceedings of SIGGRAPH*, 1998, 189–198.
- Doerschner, K., Boyaci, H., & Maloney, L. T. (2010). Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *Journal of Vision*, 10(4):8, 1–9, <https://doi.org/10.1167/10.4.8>. [PubMed] [Article]
- Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-world illumination. *Journal of Vision*, 4(9):11, 821–837, <https://doi.org/10.1167/4.9.11>. [PubMed] [Article]
- Ferwerda, J. A., Pellacini, F., & Greenberg, D. P. (2001). A psychophysically based model of surface gloss perception. *SPIE 4299, Human Vision and Electronic Imaging VI*, 4299, 291–301, <https://doi.org/10.1117/12.429501>.
- Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, 94, 62–75, <https://doi.org/10.1016/j.visres.2013.11.004>.
- Fleming, R. W., & Bülthoff, H. H. (2005). Low-level

- image cues in the perception of translucent materials. *ACM Transactions on Applied Perception*, 2(3), 346–382, <https://doi.org/10.1145/1077399.1077409>.
- Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5): 3, 347–368, <https://doi.org/10.1167/3.5.3>. [PubMed] [Article]
- Foster, D. H. (2011). Color constancy. *Vision Research*, 51, 674–700, <https://doi.org/10.1016/j.visres.2010.09.006>.
- Gilchrist, A. L. (2006). Seeing black and white. Oxford, UK: Oxford University Press.
- Hansmann-Roth, S., & Mamassian, P. (2017). A glossy simultaneous contrast: Conjoint measurements of gloss and lightness. *I-Perception*, 8(1): 204166951668777, 1–16, <https://doi.org/10.1177/2041669516687770>.
- Hunter, R. S. (1937). Methods of determining gloss. *Journal of Research of the National Bureau of Standards*, 18(1), 19–41.
- Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A*, 13(10), 1981–1991, <https://doi.org/10.1364/josaa.13.001981>.
- Karis, B. (2013). Real shading in unreal engine 4. *SIGGRAPH Course Notes: Physically Based Shading in Theory and Practice*, <https://doi.org/10.1.1.372.5001>.
- Landy, M. S. (2007, May 10). Visual perception: A gloss on surface properties. *Nature*, 447(7141), 158–159, <https://doi.org/10.1038/nature05714>.
- Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, 22(20), 1909–1913, <https://doi.org/10.1016/j.cub.2012.08.009>.
- Motoyoshi, I., & Matoba, H. (2012). Variability in constancy of the perceived surface reflectance across different illumination statistics. *Vision Research*, 53(1), 30–39, <https://doi.org/10.1016/j.visres.2011.11.010>.
- Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007, May 10). Image statistics and the perception of surface qualities. *Nature*, 447(7141), 206–209, <https://doi.org/10.1038/nature05724>.
- Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A*, 15(12), 2951–2965, <https://doi.org/10.1364/josaa.15.002951>.
- Olkkonen, M., & Allred, S. R. (2014). Short-Term Memory Affects Color Perception in Context. *PLoS One*, 9(1), e86488, <https://doi.org/10.1371/journal.pone.0086488>.
- Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision*, 10(9):5, 1–19, <https://doi.org/10.1167/10.9.5>. [PubMed] [Article]
- Olkkonen, M., & Brainard, D. H. (2011). Joint Effects of Illumination Geometry and Object Shape in the Perception of Surface Reflectance. *I-Perception*, 2(9), 1014–1034, <https://doi.org/10.1068/i0480>.
- Orhan, A. E., & Jacobs, R. A. (2014). Toward ecologically realistic theories in visual short-term memory research. *Attention, Perception, & Psychophysics*, 76(7), 2158–2170, <https://doi.org/10.3758/s13414-014-0649-8>.
- Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, 119(4), 807–830, <https://doi.org/10.1037/a0029856>.
- Vangorp, P., Laurijssen, J., & Dutré, P. (2007). The influence of shape on the perception of material reflectance. *ACM Transactions on Graphics*, 26(3), 77, <https://doi.org/10.1145/1276377.1276473>.
- Wagenmakers, E.-J., Verhagen, J., & Ly, A. (2016). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods*, 48(2), 413–426, <https://doi.org/10.3758/s13428-015-0593-0>.
- Ward, G. J. (1992). Measuring and modeling anisotropic reflection. *Computer Graphics*, 26(2), 265–72, <https://doi.org/10.1145/142920.134078>.
- Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. *Computer Graphics*, 28(2), 459–472, <https://doi.org/10.1145/192161.192286>.
- Wei, X.-X., & Stocker, A. A. (2017). Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences, USA*, 114(38), 10244–10249, <https://doi.org/10.1073/pnas.1619153114>.
- Wendt, G., Faul, F., Ekroll, V., & Mausfeld, R. (2010). Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision*, 10(9):7, 7–17, <https://doi.org/10.1167/10.9.7>. [PubMed] [Article]
- Zhang, W., & Luck, S. J. (2008, May 8). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235, <https://doi.org/10.1038/nature06860>.