

Material constancy in perception and working memory

Hiroyuki Tsuda

Keio Advanced Research Center, Keio University,
Tokyo, Japan



Munendo Fujimichi

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



Mikuho Yokoyama

Primate Research Institute, Kyoto University, Aichi, Japan



Jun Saiki

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



A key challenge for the visual system entails the extraction of constant properties of objects from sensory information that varies moment by moment due to changes in viewing conditions. Although successful performance in constancy tasks requires cooperation between perception and working memory, the function of the memory system has been under-represented in recent material perception literature. Here, we addressed the limits of material constancy by elucidating if and how working memory is involved in constancy tasks by using a variety of material stimuli, such as metals, glass, and translucent objects. We conducted experiments with a simultaneous and a successive matching-to-sample paradigm in which participants matched the perceived material properties of objects with or without a temporal delay under varying illumination contexts. The current study combined a detailed analysis of matching errors, data on the strategy use obtained via a self-report questionnaire, and the statistical image analysis of diagnostic image cues used for material discrimination. We found a comparable material constancy between simultaneous and successive matching conditions, and it was suggested that, in both matching conditions, participants used similar information processing strategies for the discrimination of materials. The study provides converging evidence on the critical role of working memory in material constancy, where working memory serves as a shared processing bottleneck that constrains both simultaneous and successive material constancy.

as glossiness, roughness, or transparency. Recently, perception and recognition of material properties or, more broadly, “shitsukan” (the sense of quality; Komatsu & Goda, 2018) is gaining increased interest in vision science. Although our understanding of material perception is advancing (see Fleming, 2017; Komatsu & Goda, 2018; Schmid & Doerschner, 2019 for recent reviews), it remains a challenging domain. The visual system must detect specific visual features that are diagnostic of particular materials (Fleming, 2014) to achieve perceptual stability, or constancy, in response to changes in viewing conditions. However, the degree of constancy varies depending on illumination, object shape, and viewing angle (e.g. see Chadwick & Kentridge, 2015 for a review of gloss constancy). How the human brain accomplishes (and sometimes fails in) material constancy remains a subject of debate that primarily focuses on the mechanism by which the visual system extracts perceptual properties of material information from low and mid-level image features (Anderson, 2020; Nishida, 2019). In contrast, the role of working memory in material constancy has so far been rarely investigated.

Successful performance in constancy tasks requires cooperation between perceptual and memory systems. For example, an individual engaging in foraging activity, and searches for food of superior quality (e.g. freshness or maturity) must compare items currently in view with items in memory, previously viewed, while also discounting differences in viewing conditions (e.g. variations in illumination). The involvement of working memory can be an influential source of error in such tasks because of the severe capacity limit (Luck & Vogel, 1997), susceptibility to viewing contexts (Allred & Flombaum, 2014), and stimulus-specific matching biases (Bae, Olkkonen, Allred, & Flombaum, 2015). We know little about the function of working memory in

Material constancy in perception and working memory

To illustrate the appearance of objects or goods seen around us, we commonly use descriptions, such

Citation: Tsuda, H., Fujimichi, M., Yokoyama, M., & Saiki, J. (2020). Material constancy in perception and working memory. *Journal of Vision*, 20(10):10, 1–16, <https://doi.org/10.1167/jov.20.10.10>.



constancy tasks because less attention has been focused on the role of working memory in material perception literature (but see a small number of studies that have considered the role of memory in color constancy tasks; Allred & Olkkonen, 2015; Jin & Shevell, 1996; Ling & Hurlbert, 2008; Uchikawa, Kuriki, & Tone, 1998). Recently, Tsuda & Saiki (2018) examined glossiness constancy against changes in illumination with a simultaneous (undelayed) and a successive (delayed) matching paradigms. They discovered that changes in the illumination context did not impair the precision of delayed matching that was evaluated by taking into account the perceptual baseline performance. This study suggests that glossiness information can be robustly recalled from working memory in the face of changes in the illumination context between study and test periods. Our motivation in the current research is to establish the role of working memory in material constancy in more detail; to this end, we use a more comprehensive range of materials and viewing conditions.

An obstacle in constructing experiments with several types of material dimension involves the difficulty in creating a large and standardized set of images due to a lack of technical expertise, computational resource, or rendering time. Recently, Sawayama and colleagues introduced a standard set of images for material recognition research (Sawayama, Dobashi, Okabe, Hosokawa, Koumura, Saarela, Olkkonen, & Nishida, 2019). This dataset contains images (computer graphics renderings) of a variety of materials with variations in conditions involving illumination context and object shape. The current study adopts a part of the image set. From six material dimensions recorded in the dataset, we selected three of them as experimental stimuli: (1) metallic silver versus glass dimension; (2) metallic gold versus plastic yellow dimension; and (3) opaque versus translucent dimension. The remaining dimensions in the dataset are related to glossiness and roughness; we did not include these material dimensions because working memory for glossiness and roughness has been investigated in our previous study (Tsuda & Saiki, 2018). A number of studies have investigated metal/glass perception (Kim & Marlow, 2016; Tamura et al., 2019; Todd & Norman, 2019), gold/yellow perception (Matsumoto, Fukuda, & Uchikawa, 2016; Okazawa, Koida, & Komatsu, 2011; Yang, Kanazawa, & Yamaguchi, 2013), and translucency perception (Chadwick et al., 2018; Fleming & Bühlhoff, 2005; Gkioulekas et al., 2013; Motoyoshi, 2010; Nagai et al., 2013; Xiao et al., 2014). These material properties are conceptually distinct from each other, and the visual system would be using different types of image cues to recognize them. By using these diverse types of materials, we can conduct a broader test of material constancy under the same experimental paradigm.

The current study was designed to investigate whether, and how, simultaneous and successive material

constancy varies in terms of matching performance and underlying information processing strategy in order to characterize the role of working memory in material constancy. Previous studies have shown that changes in illumination context does not impair either working memory of glossiness (Tsuda & Saiki, 2018) or long-term memory of surface color (Allred & Olkkonen, 2015), suggesting that successive material constancy can be well achieved, at least for some types of material dimension. Cognitive neuroscience research has shown that the same systems and representations that are engaged in perception are also recruited for the short-term retention of the sensory information (D'Esposito & Postle, 2015; Postle, 2016), and some researchers have proposed a view that any features represented in the perceptual system can also be maintained, at least briefly, in memory (Christophel, Klink, Spitzer, Roelfsema, & Haynes, 2017). In light of these findings, we can expect that any material information can be maintained in working memory well, as long as it can be stably perceived in corresponding viewing conditions.

However, it is also possible that performance in successive constancy tasks is dependent on material type. Previous research argues that material property estimation and material categorization are distinct aspects of material perception (Fleming, 2014; Fleming, 2017). The perception of glossiness and translucency involves the estimation and detailed comparison of material properties, whereas the recognition of silver/glass and gold/yellow is more similar to a categorization task. If different strategies (i.e. estimation versus categorization) are used for the evaluation of each material type, then this should affect how these materials are retained in memory. For example, if material information is encoded categorically, either by explicitly (e.g. verbal labeling) or implicitly, then detailed visual information will be lost; in turn, this would lead to increased errors and/or bias in recalling it later. Therefore, successive constancy can be less accurate than simultaneous constancy, due to shifts in an encoding strategy that may be intrinsic to each material type or task demands.

To address these issues, we conducted experiments with the matching-to-sample paradigm. Participants were presented with two sample stimuli and asked to choose which of the two is matched by a third, which was the test stimulus. In the simultaneous matching condition, sample and test stimuli were displayed together. In the successive matching condition, sample and test stimuli were separated by masking and a delay period of one second. In addition, two factors were manipulated. The first factor involved the degree of change in the illumination context. In the “same illumination” condition, all the three objects (two samples and a test) in a trial were rendered in an identical illumination environment (i.e. no constancy

is required). In the “near illumination” condition, objects were rendered in slightly different illumination environments (lower demand for constancy). In the “far illumination” condition, objects were rendered in widely different illumination environments (higher demand for constancy). The second factor was the discriminability of material property. That is, the magnitude of difference in material property varied between sample objects within a trial. Although subtle discrimination of material property is required in the low discriminability condition, coarse discrimination will suffice in the high discriminability condition.

By comparing performance between simultaneous and successive matching tasks in conditions involving varying demands for constancy and material discrimination, we aim to characterize the limits of material constancy and the underlying function of working memory. We also approached possible differences in the information processing strategy for material discrimination. Although estimation and categorization would not be mutually exclusive but complement each other in aiding material perception (see Figure 2 in Schmidt, 2019), there is ample evidence that categorization distorts working memory representations of color hues (Bae, Olkkonen, Allred, & Flombaum, 2015; Hardman et al., 2017; Persaud & Hemmer, 2016), orientations (Bae & Luck, 2018; Pratte, Park, Rademaker, & Tong, 2017), and spatial positions (Huttenlocher, Hedges, & Duncan, 1991). This suggests that a categorical encoding strategy can lead to bias in the recall of material information. We examined the use of strategy while performing the tasks via a self-report questionnaire, and conducted the statistical image analysis of diagnostic low-level image features for material discrimination, in order to gain further insight into the information processing strategy used in each task condition.

To summarize, understanding the role of working memory is key to characterizing the limits of material constancy. We conducted matching-to-sample experiments where each participant performed both simultaneous and successive matching of the material property of objects. Different material dimensions were tested by different groups of participants: Silver-glass (experiment 1), gold-yellow (experiment 2), and opaque-translucent (experiment 3). We aim to elucidate if and how the involvement of working memory will affect performance in the material constancy task by using behavioral data (accuracy and bias in matching), subjective reports of task strategy, and the statistical image analysis.

Method

We conducted three experiments. All shared the same experimental design and procedure. However, they

differed in that different stimuli images (i.e. material dimension) were tested for each experiment: metallic silver versus glass (experiment 1); metallic gold versus plastic yellow (experiment 2); and opaque versus translucent (experiment 3).

The data and analyses scripts are available on the Open Science Framework platform: <https://osf.io/cvhtx/> (DOI:10.17605/OSF.IO/CVHTX).

Participants

Seventy-two naive undergraduate and graduate students at Kyoto University participated in the experiment (24 women, mean age = 19.9 years, SD = 1.7). All participants had a normal color vision based on the Ishihara test (Ishihara, 2004), and had self-reported normal or corrected-to-normal vision acuity. Twenty-four participants were assigned to each experiment. All of them gave informed consent and 60 were given course credit and 12 were compensated monetarily (a book gift card of 1000 JPY). All experimental protocols were approved by the Institutional Review Board of Kyoto University and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Stimuli and apparatus

We used a subset of the standard image dataset for material recognition research (Sawayama et al., 2019). Of the six material dimensions recorded in the dataset, we used (1) metallic silver versus glass, (2) metallic gold versus plastic yellow, and (3) opaque versus translucent dimensions (Figure 1A). Each dimension is accompanied by five ranks of object images that have different material properties along that dimension (e.g. from pure silver to pure glass). For a detailed description of the stimulus images, see Sawayama et al. (2019). Here, we briefly describe the image rendering procedure. The images are divided into five ranks according to their material property. For example, the appearance of silver-glass images is gradually varied from pure silver to pure glass (see Figure 1A, top). In the silver-glass and gold-yellow dimensions, objects were rendered by blending the bidirectional reflectance distribution function of the two materials. In the opaque-translucent dimension, parameters of the absorption and scattering coefficients of the translucent medium (“whole milk” measured by Jensen et al., 2001) were used and the scale parameter of the scattering and absorption coefficients was varied to control translucency. All images were rendered by Mitsuba software (Jakob, 2010). We used three object shapes for each material task, which are shown in the three rows of Figure 1A. These shapes correspond to objects 1 to 3 in Figure 2b of Sawayama et al. (2019).

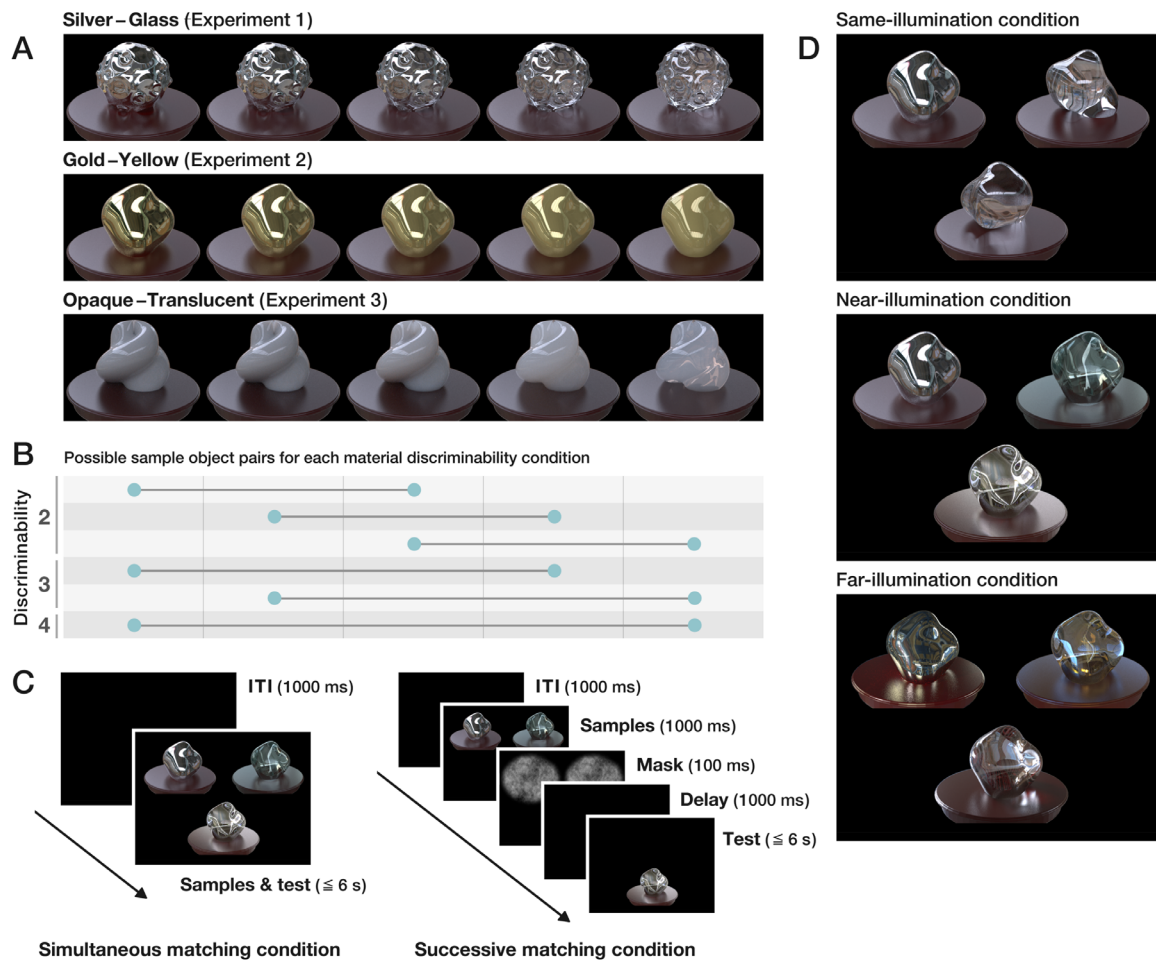


Figure 1. Stimuli and task procedure. (A) Different material dimensions were tested in each experiment. Metallic silver versus glass (Exp. 1), metallic gold versus plastic yellow (Exp. 2), and opaque versus translucent (Exp. 3). Three object shapes were used in each experiment. (B) Material discriminability, or the magnitude of difference in material rank of the two sample objects, was manipulated (discriminability was either 2, 3, or 4). Possible object pairs for each discriminability condition are illustrated. (Each dot position corresponds to each material rank in Figure 1A.) (C) A schematic depiction of a trial. On each trial, participants were presented with two sample objects (upper two) and a test object (lower one) simultaneously (in the simultaneous matching condition) or with masking and a delay period (successive condition). Participants were asked to choose which of the two sample stimuli was matched by the test stimulus in terms of material property. They were required to respond in six seconds, or the next trial began. (D) Illumination conditions. In the “same-illumination” condition, sample and test objects were rendered in an identical illumination environment. In the “near-illumination” condition, objects were rendered in slightly different illumination environments, and in the “far-illumination” condition, in widely different illumination environments. Note that on each panel, the sample stimulus on the right is matched by the test stimulus.

Stimuli were displayed on a CRT monitor (MITSUBISHI Diamondtron M² RDF223H, 1,792 × 1,344 pixels, 85-Hz) in a dark room at a distance of approximately 57 cm from participants. Each image subtended 9 degrees × 12 degrees, and the distances between images were 15 degrees. Experiments were controlled by MATLAB (The MathWorks, Inc., Natick, MA, USA) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Procedure and design

Figure 1C depicts the stimulus sequences for the matching-to-sample tasks. For a given trial, two sample objects (upper two) and a test object (lower one) were presented against a black background, and participants were asked to choose which of the two sample stimuli was matched by the test stimulus. In the simultaneous matching condition, all three objects

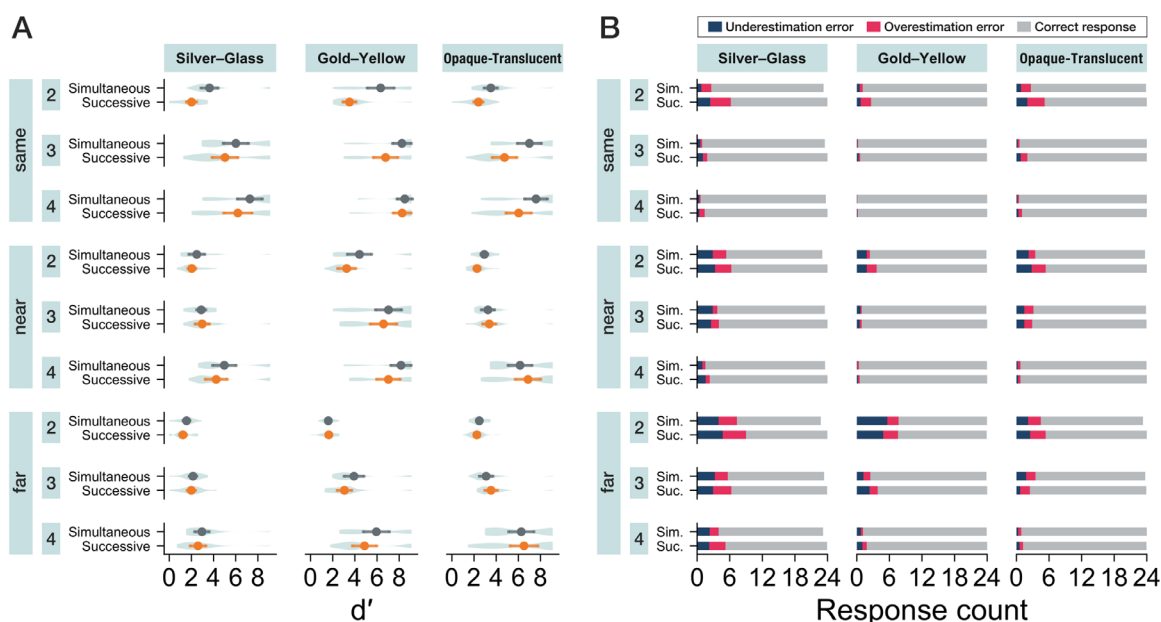


Figure 2. Matching performance (sensitivity and bias). (A) Sensitivity d' in each experimental condition (material, illumination, discriminability, and matching procedure) is shown as violin plots with mean (dots) and 95% CI (bars). (B) The proportion of response types (correct response, under-estimation error, and over-estimation error).

were presented together. In the successive matching condition, sample stimuli were followed by 100 ms noise masking (phase-scrambled images of gray-scaled sample stimuli) and a blank period of one second. In both simultaneous and successive conditions, the next trial began when participants made a response or six seconds had passed after the appearance of the test stimulus. The left/right position of the two sample objects was counterbalanced. The test image consisted of the identical material as the left sample in half of the trials and identical material as the right sample in the other half.

As briefly described in the introduction, two factors were manipulated along with the type of task (simultaneous/successive). The first factor was the illumination context (Figure 1D). In the “same-illumination” condition, sample and test objects were rendered in an identical illumination environment; and in the “near-illumination” condition, objects were rendered in slightly different illumination environments, and in the “far-illumination” condition, objects were rendered in widely different illumination environments. We used high-dynamic range light-probe images of real-world scenes from Bernhard Vogl’s database (<http://dativ.at/lightprobes/>) as the illuminations used for rendering. The degree of illumination difference was quantified by the multidimensional scaling analysis based on the pixel histogram similarities for each illumination probe (see Sawayama et al., 2019 for details). Note that in the same-illumination condition, all three objects had different poses (randomly assigned

from five pose options provided in the dataset), whereas in the near and far illumination conditions, all objects had the same pose (defined as 0 degrees angle in the dataset).¹ See Figure 1D for representative displays.

The second factor was material discriminability, or the magnitude of rank-distance in material property between two sample objects (Figure 1B). For example, when the rank distance is two, these objects are relatively similar to each other, and hence discriminability is low, while when the distance is four, objects are highly dissimilar to each other and discriminability is high. The discriminability of the two and three conditions involve more than one possible stimulus pair (3 and 2 options each; see Figure 1B), and a pairing option was randomly determined in each trial.

There were two matching procedures (simultaneous/successive), three illumination conditions (same/near/far), and three material discriminability conditions (2/3/4) in each experiment (silver-glass/gold-yellow/opaque-translucent). Each experiment took about 1 hour to complete and was composed of 12 blocks of 36 trials each, yielding 24 trials per combined task/illumination/discriminability condition. The matching procedure was fixed during a block and tested in random order; the illumination and discriminability conditions were tested in random order in each trial. Objects (two samples and a test) in a trial had the same shape, which was randomly assigned on each trial. At the beginning of each block, participants were informed of the procedure type (simultaneous or successive) of forthcoming trials. Before the formal

experiment, participants familiarized themselves with the task in a practice block. After participants completed the experiment, but before the debriefing, they were given a questionnaire designed to probe their use of strategy while performing the tasks. They were asked to write down how they performed the task in each matching procedure (simultaneous/successive).

Results

Matching performance

Any trials with response times exceeding 6 seconds were excluded from analysis, a total of 0.7% of all trials. We computed the sensitivity d' in each experimental condition based on matching accuracy (proportion correct). The d' for the matching-to-sample paradigm (ABX design; MacMillan & Creman, 1991) was calculated by using the “dprime.ABX” function in the “psyphy” package (version 0.1.9; Knoblauch, 2014) in R (version 3.6.1; R Core Team, 2016). Figure 2A shows the distribution of d' in each experimental condition: the x-axis represents the value of d' and the y-axis represents the experimental condition (illumination, discriminability, or task procedure). We can see how the matching performance was affected by the task procedure (simultaneous/successive) in each viewing condition (illumination \times discriminability) by comparing adjacent distributions of d' (gray and orange). The distributions of d' were highly overlapping between simultaneous and successive matching conditions in the near- and far-illumination condition, with the exception of the same-illumination condition, in which d' was lower in the successive matching condition. We conducted a 3 (material) \times 2 (task) \times 3 (illumination) \times 3 (discriminability) repeated measures ANOVAs with material as a between-participants factor and task, illumination, and discriminability as within-participants factors by using the “anovakun” function in R (Iseki, 2019). Results indicated that all four factors had significant main effects: material, $F(2, 69) = 28.4$, $p < 0.0001$, $\eta_p^2 = 0.45$; task, $F(1, 69) = 57.2$, $p < 0.0001$, $\eta_p^2 = 0.45$; illumination, $F(2, 138) = 194.0$, $p < 0.0001$, $\eta_p^2 = 0.74$; and discriminability, $F(2, 138) = 336.0$, $p < 0.0001$, $\eta_p^2 = 0.83$. Some factors had significant two-way interactions: task \times illumination, $F(2, 138) = 12.3$, $p < 0.0001$, $\eta_p^2 = 0.15$; material \times illumination, $F(4, 138) = 21.1$, $p < 0.0001$, $\eta_p^2 = 0.38$; material \times discriminability, $F(4, 138) = 9.13$, $p < 0.0001$, $\eta_p^2 = 0.21$; illumination \times discriminability, $F(4, 276) = 8.48$, $p < 0.0001$, $\eta_p^2 = 0.11$, and a significant three-way interaction: material \times illumination \times discriminability, $F(8, 276) = 5.02$, $p < 0.0001$,

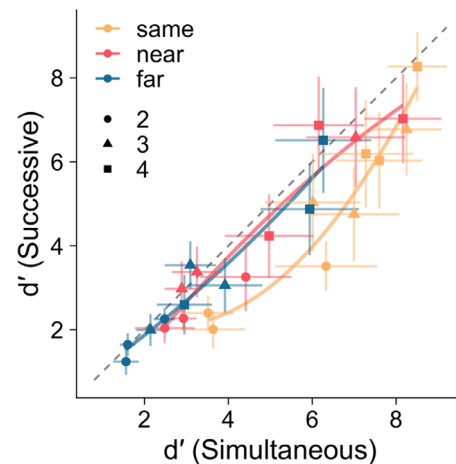


Figure 3. Effect of task procedure on matching sensitivity. Sensitivity d' in successive matching conditions are plotted against d' in the simultaneous matching conditions. Illumination and discriminability conditions are indicated in color and symbol shapes, respectively. Data were pooled across material conditions. Curves are quadratic regressions, and bars represent 95% CI.

$\eta_p^2 = 0.13$. None of the other interactions were statistically significant.

A multiple comparison (Shaffer’s modified sequentially rejective Bonferroni procedure, Shaffer, 1986) revealed a large effect of task procedure in the same-illumination condition, $F(1, 69) = 52.1$, $p < 0.0001$, $\eta_p^2 = 0.43$. However, the effect of task procedure was small in the near-illumination condition, $F(1, 69) = 4.88$, $p = 0.031$, $\eta_p^2 = 0.066$, and far-illumination condition, $F(1, 69) = 2.62$, $p = 0.11$, $\eta_p^2 = 0.037$. We plotted d' in the successive conditions against d' in the simultaneous conditions to illustrate the effect of task procedure or memory demands on matching sensitivity (Figure 3: data were pooled over the material dimension because the interaction between task procedure and material dimension was nonsignificant.) As shown above, the negative impact of memory demands on d' was evident in the same-illumination condition, but not in near- and far-illumination conditions. Data points in near- and far-illumination conditions were widely spread around the diagonal line, which suggests that observers were able to perform successive matching as accurately as simultaneous matching regardless of task difficulty in each viewing condition.

End point effect

When the material discriminability was two, one of the two samples could be picked from material ends (ranks 1 or 5). (As can be seen in Figure 1B, one of the two samples was always from ranks 1 or 5 when

discriminability was 3 or 4.) Matching might have been easier in trials that contained end point objects because end point objects could be easily classified or labeled as silver, glass, or gold, among others. In contrast, the ranks in the middle appear more ambiguous regarding their material category. We conducted an exploratory analysis of the effect of the presence or absence of endpoints on matching performance in order to test for the possible effect of categorization on material matching. Trials were split into endpoint-absent trials (a sample pair included ranks 2 and 4, which was 33% of all trials) and end point-present trials (a sample pair included ranks 1 and 3 or 3 and 5, which were 64% of all trials). Results indicated that the mean d' was higher in the endpoint-absent trial (4.6) compared to the end point-present trial (2.8), $t = -8.96$, $p < 0.0001$, which was independent of the material dimension and task procedure. An ANOVA indicated a nonsignificant interaction between the factors: material \times end point, $F(2, 69) = 1.99$, $p = 0.14$, $\eta_p^2 = 0.054$; task \times end point, $F(1, 69) = 0.027$, $p = 0.87$, $\eta_p^2 = 0.0004$. We have further discussed one possible explanation of higher performance in the end point-absent trial in the general discussion section.

Bias in errors

To further characterize the source of error in matching tasks, we identified two types of errors: underestimation and overestimation. Underestimation occurred when participants chose the wrong option (one of the sample stimuli) that had lower material rank than the other option (and the test stimulus). Overestimation occurred when participants chose the wrong option that had higher material rank than the other option/test. (Note that the direction of material dimension is arbitrarily defined such that for the silver-glass stimuli, silver-like objects are lower ranks and glass-like objects are higher ranks in the silver-glass dimension.)

The bias in errors (i.e. errors can be biased toward either underestimation or overestimation) may be present in both simultaneous and successive matching conditions. Bias in simultaneous matching is expected if the discriminability of material properties is not homogeneous (equidistant) along its material dimension, or if some stimuli are more accessible to be categorized as a specific type of material than others. Such biases may well emerge in the simultaneous matching condition because the material image dataset has not strictly controlled for these perceptual properties (Sawayama et al., 2019). Our interest here is to examine whether the bias in simultaneous matching, if it exists, will be amplified in the successive matching condition. If participants encoded material

information categorically and did not retain detailed visual information in working memory, then using such an encoding strategy should result in increased bias in errors, because it amplifies response bias that is intrinsic to perception.

We counted each response type (correct response, underestimation error, and overestimation error) in each condition; then, the counts were averaged across participants (Figure 2B). Visual inspection reveals that, in most cases, both types of errors were more or less equally likely, although, in some conditions, either underestimation or overestimation errors were more frequent. Importantly, these patterns were not much different between simultaneous and successive conditions. The statistical analysis confirmed these insights. For each material dimension, the count data (underestimation and overestimation) was fitted with a binomial generalized linear model with the logit link function (we confirmed that the choice of error structure was appropriate to the data by the Pearson dispersion parameters). The analysis of deviance (type III tests and likelihood ratio test) revealed that in all the experiments (i.e. material dimensions), the effect of task was not significant: silver-glass, $\chi^2 = 1.90$, $df = 1$, $p = 0.17$; gold-yellow, $\chi^2 = 2.99$, $df = 1$, $p = 0.083$; and opaque-translucent, $\chi^2 = 1.03$, $df = 1$, $p = 0.31$. This result is compatible with the view that memory demand did not amplify the bias in errors, in any material dimensions. Note that in all the experiments, the effect of illumination was significant: silver-glass, $\chi^2 = 13.1$, $df = 2$, $p = 0.0014$; gold-yellow, $\chi^2 = 8.18$, $df = 2$, $p = 0.017$; and opaque-translucent, $\chi^2 = 17.0$, $df = 2$, $p < 0.001$. In fact, in the same-illumination condition, overestimation is more frequent, whereas in the other illumination conditions, both types of errors were about equally likely on average.

Note that objects had different poses in the same-illumination condition. Could the rotation of objects have resulted in a change of material appearance? For example, could a highlight appear more pronounced in some poses than in others, which led to the response bias in the same-illumination condition? We conducted binomial GLM as described above with a subset of the data (i.e. trials in the same-illumination condition) to test for the effect of pose on response bias. The results indicated that the effect of pose was not significant: silver-glass, $\chi^2 = 2.10$, $df = 1$, $p = 0.15$; opaque-translucent, $\chi^2 = 3.23$, $df = 1$, $p = 0.072$. (Result of the gold-yellow condition is not shown because there were few bias counts in the gold-yellow condition, which made the model fitting unreliable.) Therefore, we concluded that the trend for overestimation in the same-illumination condition was not explained by object rotations.

In sum, the analysis of error bias type showed that although some biases in errors were observed depending on viewing conditions, the evidence that it was

	Estimation	Categorization	Miscellaneous	No strategy
Simultaneous condition				
Silver-glass	9	3	1	11
Gold-yellow	10	3	3	8
Opaque-translucent	14	0	1	9
Successive condition				
Silver-glass	7	6	3	8
Gold-yellow	13	2	4	5
Opaque-translucent	19	1	3	1
Aggregated				
Simultaneous condition	33	6	5	28
Successive condition	39	9	10	14

Table 1. Classification of self-reported task strategy.

affected by memory demand was weak. Such findings suggest that participants used similar discrimination criteria or strategy for material discrimination in both simultaneous and successive matching conditions. This point was further addressed with the analysis of questionnaire data, as described below.

Self-reported task strategy

Self-reported task strategy data was collected via a questionnaire. Participants were asked to report how they performed the task for each matching condition (simultaneous/successive). The reported strategies were categorized into four types: estimation strategy, categorization strategy, miscellaneous strategy, and no strategy. The first two of these strategies were identified based on the previous material perception literature (Fleming, 2014; Fleming, 2017). A response was categorized as the estimation strategy when it referred to visual properties of materials or surface qualities, with a focus on their intensity or subtle differences between objects. The use of visual mental imagery was also categorized as the estimation strategy. The categorization strategy was identified when a response referred to either categorization, coarse/discrete discrimination, or the use of verbal labeling. When a form was left blank or no specific strategy was mentioned (e.g. “nothing especially” or “just intuition”), it was categorized as the “no strategy.” The remaining responses were categorized as the miscellaneous strategy (e.g. “I made judgments based on inferred weights of objects”; “I judged how expensive or cheap the objects looked like”).

The results of the categorization of self-reported strategies are summarized in Table 1. Estimation was the most common strategy across materials and tasks (“no strategy” is not considered here). In addition, in each material dimension, the number of both estimation and categorization strategy responses did not differ between

simultaneous and successive conditions (all p values > 0.05 with the two sample Poisson exact test [Fay, 2010], conducted by using the “rateratio.test” package in R [Fay, 2014]). Two key findings by the analysis of subjective reports are: (1) most participants either used the estimation strategy or did not rely on a specific strategy, and the response of categorization strategy was relatively rare, and (2) strategy choice was not affected by task type (simultaneous/successive). Note that in the aggregated data, the number of the no strategy decreased from 28 in the simultaneous condition to 14 in the successive condition ($p = 0.043$). This result may reflect that, although there is an increased need of *some* strategy to perform memory-based matching, this need did not result in using a specific type of strategy. In sum, the categorical encoding was not an attractive strategy to perform the successive matching task, suggesting that participants did not merely encode material information by category, but retained detailed visual information of materials, in any material dimension conditions. We then examined what image cues were relevant for the discrimination of materials.

Image analysis

Sawayama et al. (2019) considered whether participants’ judgments in their oddity task could be predicted by simple image statistics, such as the mean of luminance and color. They examined the relationship between the sensitivity d' and the mean color difference between target and nontarget stimuli, and found some but not large correlation between them. We conducted the same analysis in our dataset. As in Sawayama et al. (2019), the mean color difference ΔE^*_{ab} was defined as follows:

$$\Delta E^*_{ab} = \left[(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2 \right]^{1/2}$$

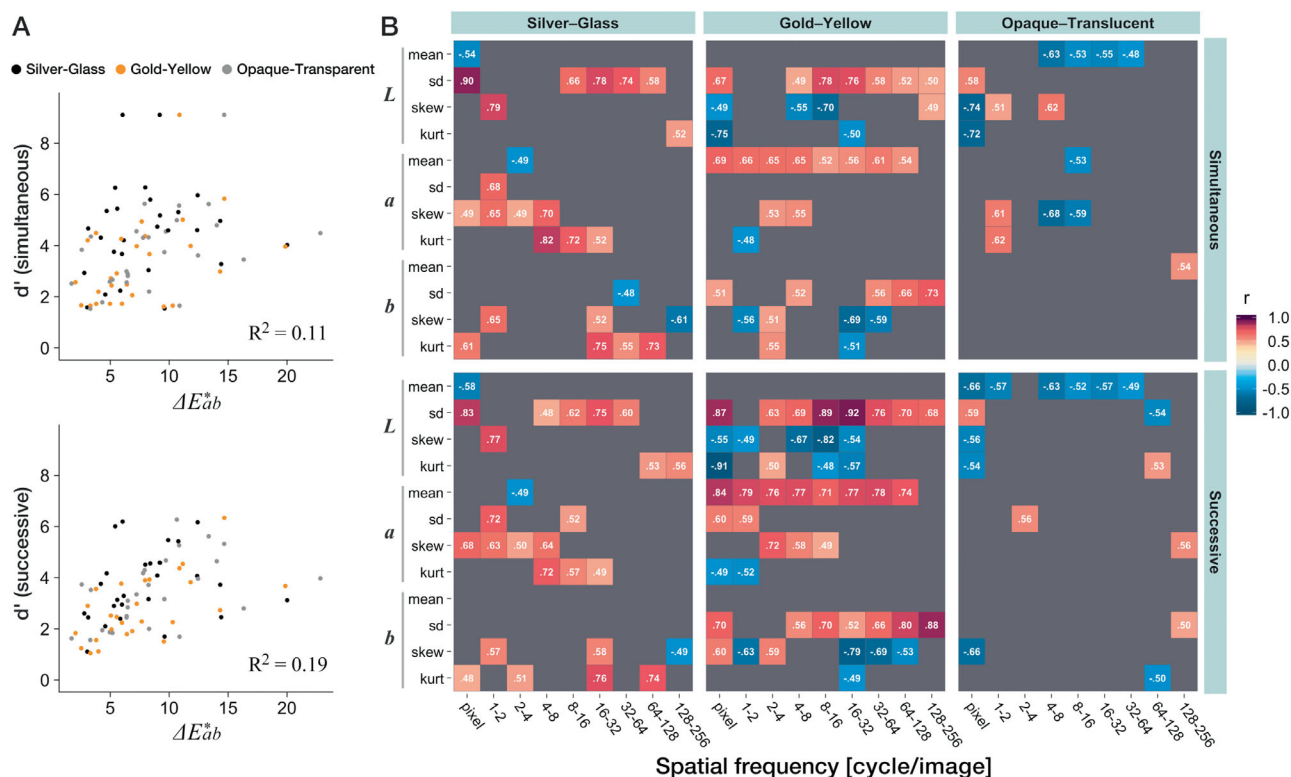


Figure 4. Image analysis. (A) Correlation between the sensitivity d' and the mean color difference ΔE^*_{ab} between test and distractor stimuli. (B) Correlation coefficients of low-level image features with d' . Only statistically significant correlations are shown.

where L_1^* , a_1^* , and b_1^* are the spatial mean CIELAB values averaged over all possible target stimuli (sample stimuli that had the same material property with the test stimulus), whereas L_2^* , a_2^* , and b_2^* are those averaged over all possible nontarget stimuli. Averaging was performed per combined material/shape/illumination/discriminability condition (yielding 81 groups or data points). Figure 4A shows the result as scatter plots. There was a significant relationship (Pearson correlation) between ΔE^*_{ab} and mean d' in both the simultaneous condition, $r = 0.33$, $p = 0.003$, and the successive condition, $r = 0.44$, $p < 0.001$. The mean color difference explained only 10% to 20% of the d' variance. These results suggest that participants did not simply rely on mean color differences to discriminate materials, which was also the case in the oddity task used in Sawayama et al. (2019).

We also conducted a more elaborate image analysis to examine how sensitivity d' was related to low-level image statistics, and examined if and how the relationships depend on the material dimension and memory demand. First, we extracted image features (moment statistics) from bandpass-filtered images. The original standard red, green, blue (sRGB) images in the dataset were converted to the CIELAB color space, and each color channel was decomposed into sub band images of eight different spatial scales by using the Butterworth bandpass filter (fourth-order) with

cutoff frequencies of 1 to 2, 2 to 4, 4 to 8, 8 to 16, 16 to 32, 32 to 64, 64 to 128, and 128 to 256 cycles/image. For each color channel, the moment statistics of the image (mean, SD, skewness, and kurtosis) were calculated for each sub band. Note that pixels only within the object region were used for the calculation of moments, and which is the reason we included the first moment (mean) in the analysis. (The first moment is meaningless for uncropped bandpass-filtered images because they are zero-centered.) Then, following the same procedure as described above, the mean difference in each moment statistic between target and nontarget stimuli was correlated with sensitivity d' . The results are shown in Figure 4B. The horizontal axis represents the spatial frequency band, and the vertical axis represents moment statistics for each color channel. (Note that the correlation analysis was also conducted with the original sRGB image, which is shown in the “pixel” column.) Only the statistically significant correlations ($p < 0.05$, multiple comparisons were corrected with the Beyer-Hardwick [BH] method) appear in this figure.

Some interesting patterns are suggested in the plot. Although the a-channel was predictive of sensitivity mostly at lower frequencies, the L and b channels were less dependent on frequency, and color channels were generally less predictive for the opaque-translucent discrimination, etc. However, we cannot commit to making such detailed interpretations about the

relationships between image features and d' because they are statistically unreliable: There were only 27 data points in each cell; the correlation coefficient derived from such small sample size is relatively inaccurate, and even the sign of correlation can be reversed (Schönbrodt & Perugini, 2013). Therefore, we will only consider the holistic pattern of correlations instead of focusing on the effects of individual image features.

First, in each material dimension, the set of diagnostic image features was similar between simultaneous and successive matching conditions. To measure the association, we calculated Kendall's tau-b statistic, a measure of rank correlation, between the correlation coefficient values of simultaneous and successive conditions, by using the “KendallTauB” function in the “DescTools” package in R (Signorell, 2020). The tau-b correlations were 0.75, 95% confidence interval (CI) = [0.59, 0.90] for silver-glass, 0.64, 95% CI = [0.50, 0.79] for gold-yellow, and 0.36, 95% CI = [0.11, 0.62] for opaque-translucent. High correlations indicate the similarity of diagnostic image features between simultaneous and successive matching conditions. (Note that the tau-b had relatively higher uncertainty for opaque-translucent dimension due to the small number of significant correlations.) Second, diagnostic image features were idiosyncratic to each material dimension, that is, they were dissimilar between different material dimensions. In fact, tau-b correlations were close to zero and the 95% CIs crossed zero when they were calculated for pairs with different materials. Taken together, the image analysis revealed idiosyncratic patterns of diagnostic low-level image cues to each material discrimination and, importantly, they were highly similar between simultaneous and successive matching conditions.

General discussion

Working memory is a fundamental system for a wide range of perceptual and cognitive activities (Tan, Lallee, & Mandal, 2017), yet the function of working memory in situations involving naturalistic stimuli and tasks remains unclear (Orhan & Jacobs, 2014), and material constancy is one such example. In three experiments with the matching-to-sample paradigm, we examined how material information is encoded in and recalled from working memory. First, we found that matching performance (measured by the sensitivity metric d') was comparable between simultaneous and successive matching conditions when there was a change in illumination context between objects. Second, although there were some biases in matching errors toward either underestimation or overestimation of material properties depending on the viewing condition, these biases were not affected by whether matching

was performed with or without a delay period. Combined with the analysis of self-reported strategy, our result suggests that participants performed both simultaneous and successive matching tasks with similar discrimination criteria or strategies; they did not merely encoded material information by category, but retained detailed visual information of materials to perform the tasks. Finally, we explored the association between low-level image features and matching performance. Here, results suggested that whereas idiosyncratic sets of diagnostic image features were observed for each material dimension, the diagnostic features were highly similar between simultaneous and successive matching conditions in each material dimension. Taken together, the set of results provide the basis for understanding the role of working memory in material constancy.

The role of working memory in material constancy

Previous studies have reported that constancy occurs in working memory for glossiness (Tsuda & Saiki, 2018) and long-term memory for color (Allred & Olkkonen, 2015; Jin & Shevell, 1996; Ling & Hurlbert, 2008; Uchikawa, Kuriki, & Tone, 1998). The results of the current study are in line with these observations, which further suggests that material constancy is comparable between simultaneous constancy (i.e. discounting of the spatial change in illumination) and successive constancy (i.e. discounting temporal changes in illumination), at least in the current paradigm and stimuli. The comparable performance between simultaneous and successive matching may be surprising, considering that it is generally assumed that visual working memory representation is the noisy version of perception (e.g. Bays, 2015; Sims, Jacobs, & Knill, 2012; van den Berg, Shin, Chou, Georgea, & Ma, 2012). When perceptual and memory matching performances were directly compared using the same experimental procedure and stimuli, studies revealed that working memory representation was indeed less accurate and more variable than perception (e.g. Bae et al., 2015; Tsuda & Saiki, 2018; Tsuda & Saiki, 2019).

However, in a color constancy study by Allen and Olkkonen (2015), a comparable precision of color matching was reported between simultaneous and successive matching conditions, where there was a change in illumination context between sample and test stimuli. The authors explained such results by theorizing that working memory is a common source of error in both simultaneous and successive matching conditions. That is, participants looked back and forth between stimuli in the simultaneous matching condition, which likely involved working memory. If both simultaneous and successive matching

conditions shared the same processing bottleneck (i.e. working memory capacity), it is unsurprising to observe a comparable performance between the two conditions. Other color constancy studies have likewise documented the involvement of the memory system in color constancy tasks (Jin & Shevell, 1996; Ling & Hurlbert, 2008; Uchikawa, Kuriki, & Tone, 1998).

Our results on material constancy can be similarly explained (i.e. working memory was a shared bottleneck in both simultaneous and successive matching conditions). In addition, more generally, it is possible that if a task requires constancy against changes in illumination, performance in simultaneous matching will tend to be closer to that of successive matching. (Although this would be dependent on how perceptual matching is performed. If a participant is given unlimited time for response, matching will be more precise in simultaneous than in the successive matching condition, as observed in a glossiness constancy study by Tsuda & Saiki, 2018.) In tasks that use simple visual features, such as color and orientation, in which constancy is totally an irrelevant factor, then it is no surprise to observe higher matching precision in the simultaneous matching condition because the exact image matching is possible. Likewise, in the same-illumination condition in our task, where constancy (against changes in illumination) is not required, the simultaneous matching condition should have (and actually) provided more accurate matching performance than the successive matching condition. (Note that the exact image matching was not possible in the same-illumination condition in a precise sense because of the variations in object's pose. Nevertheless, most low-level image properties would be preserved regardless of this manipulation.)

Regarding the result of image analysis, we found high similarities of diagnostic image features between simultaneous and successive matching conditions. According to the sensory recruitment theory of working memory (D'Esposito & Postle, 2015; Scimeca, Kiyonaga, & D'Esposito, 2018), the similarity may reflect the shared neural and representational basis of material information between perception and working memory systems. However, from the perspective of the discussion above, it may be more appropriate to interpret the diagnostic cue similarity as merely reflecting the involvement of working memory in both simultaneous and successive matching conditions. Although the image analysis provided some clues to the representational format of material information, it is still challenging to isolate the perceptual and memory contributions to behavioral performance data. Neuroimaging techniques such as functional magnetic resonance imaging (fMRI) would facilitate examining this issue more directly. Neurophysiological evidence provides a view that visual perception and categorization of materials are processed mainly

through a hierarchy of the ventral visual pathway (Hiramatsu, Goda, & Komatsu, 2011; Komatsu & Goda, 2018), and the involvement of the fusiform face area was suggested for visual working memory of material category information (Otsuka & Saiki, 2019). In order to fully elucidate the underlying mechanism of material constancy, future research should examine the neural basis of the estimation of material properties, how it differs from categorization, and whether the sensory recruitment theory is applicable not only to basic visual features, such as color but also to the mid-level visual information, such as material properties of objects.

Comparison to color/lightness constancy

The present study suggests that material constancy is comparable between simultaneous and successive matching conditions. However, previous studies on color (Olkkonen and Allred, 2014) and lightness perception (Olkkonen, Saarela, & Allred, 2016) have shown that adding a short retention interval decreases constancy. What caused the difference between this and previous studies? Are the memory effects in material perception different from that in color and lightness perception? Previous studies measured the effects of memory demand and illumination shift on perceived color/lightness, and examined the independent effects of memory and illumination on the appearance of colors and lightness. Those studies reported that these effects were not independent. Matching in the combined memory load and illumination shift conditions elicited a smaller bias than was predicted by the independent (and additive) effects of memory and context biases. This sub-additive property indicates less constancy in the successive matching condition for color and light (Olkkonen & Allred, 2014; Olkkonen, Saarela, & Allred, 2016).

The task procedure and the performance measure differ between those studies and the present one. Nevertheless, we found that a similar analysis could be conducted with our data. Previous studies used the point of subjective equality (PSE) of psychometric function as a measure of bias (appearance of color/lightness), whereas we quantified bias (appearance of materials) based on the frequency of under- and overestimation errors, such that the possibility of bias was suggested if either error type was more frequent than the other. Therefore, we defined the bias index as follows:

$$\text{Bias index} = P_{\text{overestimation}} - P_{\text{underestimation}},$$

where $P_{\text{overestimation}}$ is the proportion of overestimation error and $P_{\text{underestimation}}$ is the proportion of underestimation error. The index ranges between -1

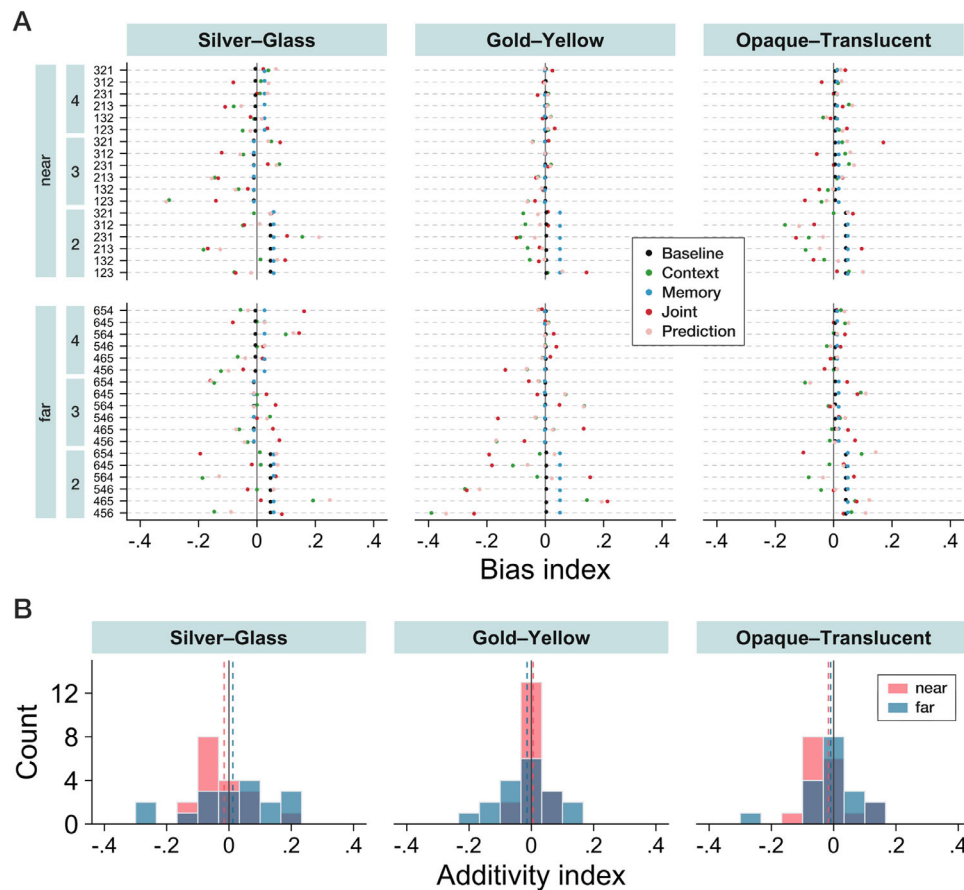


Figure 5. Analysis of constancy. (A) Bias index (see main text for the definition) was calculated for each viewing condition. The three-digit numbers on the left represent the illumination conditions. Each digit corresponds to the illumination ID of the target sample, the distractor sample, or the test object (e.g. “321” represents the condition in which the target sample had the illumination #3, the distractor sample had the illumination #2, and the test objects had the illumination #1). (B) Histogram of additivity indices calculated for each viewing condition. The vertical black line indicates the independence of memory and illumination biases. Vertical dashed color lines show the mean index.

and 1, in which 0 indicated no response bias. Then, we calculated the bias index for each condition following the procedure by [Olkkonen, Saarela, and Allred \(2016\)](#): baseline (simultaneous, same illumination), context (simultaneous, near/far illumination), memory (successive, same illumination), and joint context-memory (successive, near/far illumination), as shown in [Figure 5A](#). Finally, the additivity index (the difference between the observed and predicted bias index) was calculated for each material \times illumination condition ([Figure 5B](#)).

Previous studies on color and lightness perception have indicated a negative additivity index (i.e. subadditivity). In contrast, our data showed neither subadditivity nor super-additivity because the distribution of the additivity index was centered around zero in all the conditions, which is suggestive of material constancy in working memory. It is possible that these results cannot be taken at face value because our task was not designed for these types of

analyses. Nevertheless, our results suggest an interesting difference in constancy, and memory effects, among the studies.

How can we explain the different results between studies? The richness of visual information may be relevant to constancy. The stimuli used in the previous study ([Olkkonen & Allred, 2014](#); [Olkkonen, Saarela, & Allred, 2016](#)) were flat patches under artificial illumination (spatially constant single color illuminant). In contrast, the current study used objects with complex 3D shapes under real-world illumination. Material perception research has shown the importance of realistic viewing conditions for constancy. For example, gloss perception was close to the veridical when seen under complex real-world illumination, but not under simple/artificial illumination ([Dror, Willsky, & Adelson, 2004](#); [Fleming, Dror, & Adelson, 2003](#)). Moreover, gloss perception deteriorated when the object shape was changed from bumpy to uniform ([Marlow & Anderson, 2013](#)).

The findings of the current study suggest that the visual system took advantage of the rich visual information, or cues to the materials, in stimuli used in the present study, which led to the preserved successive constancy. This notion also suggests that it was not visual attributes per se (color/lightness/materials), but the naturalness of the stimuli and their complexity (object shape and illumination) that accounted for how the task affected constancy. Olkkonen, Saarela, and Allred (2016) argued that successive lightness constancy was weakened due to shifts in estimated illuminants in memory. However, such shifts might be weak under real-world illumination, because the visual system has tacit knowledge of the statistics of real-world illumination, or unsystematic, because of the complexity of natural illumination.

Limitations and future directions

Before concluding, some limitations of the current study and future directions should be mentioned. The first limitation concerns the variation of material constancy across material category and across time scales. Although we have found that optical properties, such as glossiness, translucency, and the type of materials (metal/glass/plastic), are well remembered and recalled in the presence of changes in illumination context, it remains to be determined whether the mechanical properties of materials, such as viscosity, elasticity, and stiffness, are also retained well in memory. How the dynamic aspects of material property are represented in memory is currently unknown. In addition, because we only considered material constancy in working memory, the effect of time scales remains unclear. Memory biases/distortions and categorical effects are more likely to occur in long-term memory (Persaud & Hemmer, 2016).

The second issue relates to the difference in the representation of material information between perception and memory; if such differences do exist, they might be more effectively identified with some higher-order image statistics because they can represent more abstract information that might be more suited to the modeling of memory representation. We focused on low-level image features because they have been frequently used for modeling material perception and recognition (Motoyoshi, Nishida, Sharan, & Adelson, 2007; Nishida, 2019; Nishida & Shinya, 1998). Such analysis can be easily applicable for any material dimension, and therefore useful for comparing results across different materials.

Finally, we would like to discuss the possible role of categorization in material constancy. The present results indicated that categorical coding was not a dominant strategy for performing material matching tasks. However, we found a possibly exciting result

of categorical effects in material perception. In the “end point effect subsection” of the results section, we described that matching performance was higher in the end point-absent trial than in the end point-present trial. We know that a categorical border between two sample colors accelerates their discrimination (Bornstein and Korda, 1984; Witzel, 2019; Witzel & Gegenfurtner, 2018). This result can be explained by the category border effect if end point objects (ranks 1 and 5) represent distinct categorical centers (e.g. silver and glass) and the middle rank (rank 3) is close to the categorical border of the two materials: Sample objects in the end point-absent trial (i.e. ranks 2 and 4) cross the categorical border and would be easier to discriminate. Conversely, both samples in the end point-present trial belong to the same material category (i.e. ranks 1 and 3 and ranks 3 and 5), and hence would not benefit from the category effect. Further research is needed to test these categorical effects in material perception.

Conclusions

In conclusion, we demonstrated comparable material constancy between simultaneous and successive matching conditions with diverse type of materials (metals, glass, plastic, and translucent objects). Combined with the self-reported task strategy and the analysis of diagnostic image features for material discrimination, converging evidence suggests that the set of results are best explained by a shared processing bottleneck (i.e. working memory) that constrains both simultaneous and successive material constancy. Although the role of the memory system has been under-represented in recent material perception literature, our study suggests that the capacity of working memory should be considered in characterizing the limits of material constancy.

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

Acknowledgments

The authors thank Yuki Takeda for coding support. Supported by JSPS KAKENHI Grant Numbers JP18H05006 and JP20H00107.

The data and analyses scripts are available on the Open Science Framework platform: <https://osf.io/cvhtx/> (DOI:10.17605/OSF.IO/CVHTX).

Commercial relationships: none.

Corresponding authors: Hiroyuki Tsuda; Jun Saiki.
Email: tsuda16k@gmail.com; saiki.jun.8e@kyoto-u.ac.jp.

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

Footnote

¹It would be more appropriate to vary object pose in the near-illumination and the far-illumination conditions in the same way as in the same-illumination condition to control for the effect of the pose across experimental conditions. However, this was not possible due to the structure of the image dataset: Variations in object pose was only provided for the same-illumination stimuli. We could have made the object pose constant in the same-illumination condition. However, it allows participants to use the exact image matching strategy in the simultaneous matching condition, which is inappropriate to investigate material (rather than image) matching processes. To avoid this, we varied object pose in the same-illumination condition. Moreover, this task setting is the intended use of the stimuli by the authors of the material image dataset.

References

- Allred, S. R., & Flombaum, J. I. (2014). Relating color working memory and color perception. *Trends in Cognitive Sciences*, 18(11), 562–565.
- Allred, S. R., & Olkkonen, M. (2015). The effect of memory and context changes on color matches to real objects. *Attention, Perception, & Psychophysics*, 77(5), 1608–1624.
- Anderson, B. L. (2020). Mid-level vision. *Current Biology*, 30(3), R105–R109.
- Bae, G., & Luck, S. J. (2018). What happens to an individual visual working memory representation when it is interrupted? *British Journal of Psychology*, 110(2), 268–287.
- 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(4), 744–763.
- Bays, P. M. (2015). Spikes not slots: noise in neural populations limits working memory. *Trends in Cognitive Sciences*, 19(8), 431–438.
- Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychological Research*, 46(3), 207–222.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- Chadwick, A. C., Cox, G., Smithson, H. E., & Kentridge, R. W. (2018). Beyond scattering and absorption: Perceptual unmixing of translucent liquids. *Journal of Vision*, 18(11), 18.
- Chadwick, A. C., & Kentridge, R. W. (2015). The perception of gloss: A review. *Vision Research*, 109, 221–235.
- Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, 21(2), 111–124.
- D’Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, 66(1), 115–142.
- Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-world illumination. *Journal of Vision*, 4(9), 11.
- Fay, M. P. (2010). Two-sided exact tests and matching confidence intervals for discrete data. *R Journal*, 2(1), 53–58.
- Fay, M. P. (2014). rateratio.test: Exact rate ratio test. R package version 1.0-2, <https://cran.r-project.org/package=rateratio.test>.
- Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, 94, 62–75.
- Fleming, R. W. (2017). Material Perception. *Annual Review of Vision Science*, 3(1), 365–388.
- Fleming, R. W., & Bülthoff, H. H. (2005). Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception*, 2(3), 346–382.
- Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5), 3.
- Gkioulekas, I., Xiao, B., Zhao, S., Adelson, E. H., Zickler, T., & Bala, K. (2013). Understanding the role of phase function in translucent appearance. *ACM Transactions on Graphics*, 32(5), 1–19.
- Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 30–54.
- Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *NeuroImage*, 57(2), 482–494.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98(3), 352–376.
- Iseki, R. (2019). anovakun. Retrieved from <http://riseki.php.xdomain.jp/index.php?ANOVA%E5%90%9B>.
- Ishihara, S. (2004). *Ishihara’s tests for colour deficiency*. Tokyo, Japan: Kanehara Trading Inc.

- Jakob, W. (2010). Mitsuba renderer, <http://www.mitsuba-renderer.org>.
- Jensen, H. W., Marschner, S. R., Levoy, M., & Hanrahan, P. (2001). A practical model for subsurface light transport. *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '01. the 28th annual conference*.
- Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A*, 13(10), 1981.
- Kim, J., & Marlow, P. J. (2016). Turning the world upside down to understand perceived transparency. *I-Perception*, 7(5), 204166951667156.
- Knoblauch, K. (2014). psyphy: Functions for analyzing psychophysical data in R. R package version 0.1-9, <https://cran.r-project.org/package=psyphy>.
- Komatsu, H., & Goda, N. (2018). Neural mechanisms of material perception: Quest on Shitsukan. *Neuroscience*, 392, 329–347.
- Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, 25(6), 1215.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, MA, Cambridge University Press.
- Marlow, P. J., & Anderson, B. L. (2013). Generative constraints on image cues for perceived gloss. *Journal of Vision*, 13(14), 2–2.
- Matsumoto, T., Fukuda, K., & Uchikawa, K. (2016). Appearance of gold, silver and copper colors of glossy object surface. *International Journal of Affective Engineering*, 15(3), 239–247.
- Motoyoshi, I. (2010). Highlight-shading relationship as a cue for the perception of translucent and transparent materials. *Journal of Vision*, 10(9), 6–6.
- Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, 447(7141), 206–209.
- Nagai, T., Ono, Y., Tani, Y., Koida, K., Kitazaki, M., & Nakauchi, S. (2013). Image regions contributing to perceptual translucency: A psychophysical reverse-correlation study. *I-Perception*, 4(6), 407–428.
- Nishida, S. (2019). Image statistics for material perception. *Current Opinion in Behavioral Sciences*, 30, 94–99.
- 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.
- Okazawa, G., Koida, K., & Komatsu, H. (2011). Categorical properties of the color term “GOLD.” *Journal of Vision*, 11(8), 4.
- Olkkonen, M., & Allred, S. R. (2014). Short-Term Memory Affects Color Perception in Context. *PLoS ONE*, 9(1), e86488.
- 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.
- Orhan, A. E., & Jacobs, R. A. (2014). Toward ecologically realistic theories in visual short-term memory research. *Attention, Perception, & Psychophysics*, 76(7), 2158–2170.
- Otsuka, S., & Saiki, J. (2019). Neural correlates of visual short-term memory for objects with material categories. *Heliyon*, 5(12), e03032.
- Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology*, 88, 1–21.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Postle, B. R. (2016). How does the brain keep information “in mind”? *Current Directions in Psychological Science*, 25(3), 151–156.
- Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 6–17.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <https://www.R-project.org/>.
- Sawayama, M., Dobashi, Y., Okabe, M., Hosokawa, K., Koumura, T., Saarela, T., Olkkonen, M., . . . Nishida, S. (2019). Visual discrimination of optical material properties: a large-scale study. *bioRxiv*. Published online October 10, 2019.
- Schmidt, F. (2019). The art of shaping materials. *Art and Perception*, 1–27, <https://doi.org/10.1163/22134913-20191116>.
- Schmid, A. C., & Doerschner, K. (2019). Representing stuff in the human brain. *Current Opinion in Behavioral Sciences*, 30, 178–185.
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609–612.

- Scimeca, J. M., Kiyonaga, A., & D'Esposito, M. (2018). Reaffirming the sensory recruitment account of working memory. *Trends in Cognitive Sciences*, 22(3), 190–192.
- Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of American Statistical Association*, 81, 826–831.
- Signorell, A. (2020). DescTools: Tools for descriptive statistics. R package version 0.99.32, <https://cran.r-project.org/package=DescTools>.
- Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, 119(4), 807–830.
- Tamura, H., Prokott, K. E., & Fleming, R. W. (2019). Distinguishing mirror from glass: A 'big data' approach to material perception. *arXiv preprint arXiv:1903.01671*.
- Tan, C., Lallee, S., & Mandal, B. (2017). Vision and memory: Looking beyond immediate visual perception. In Q. Zhao (Eds.), *Computational and Cognitive Neuroscience of Vision* (pp. 195–219). Singapore: Springer, https://doi.org/10.1007/978-981-10-0213-7_9.
- Tsuda, H., & Saiki, J. (2018). Constancy of visual working memory of glossiness under real-world illuminations. *Journal of Vision*, 18(8):14, 1–16.
- Tsuda, H., & Saiki, J. (2019). Gradual formation of visual working memory representations of motion directions. *Attention, Perception, & Psychophysics*, 81(1), 296–309.
- Todd, J. T., & Norman, J. F. (2019). Reflections on glass. *Journal of Vision*, 19(4), 26.
- Uchikawa, K., Kuriki, I., & Tone, Y. (1998). Measurement of color constancy by color memory matching. *Optical Review*, 5(1), 59–63.
- van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109(22), 8780–8785.
- Witzel, C. (2019). Misconceptions about colour categories. *Review of Philosophy and Psychology*, 10(3), 499–540.
- Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: Objects, constancy, and categories. *Annual Review of Vision Science*, 4(1), 475–499.
- Xiao, B., Walter, B., Gkioulekas, I., Zickler, T., Adelson, E., & Bala, K. (2014). Looking against the light: How perception of translucency depends on lighting direction. *Journal of Vision*, 14(3), 17–17.
- Yang, J., Kanazawa, S., & Yamaguchi, M. K. (2013). Can Infants Tell the Difference between Gold and Yellow? *PLoS One*, 8(6), e67535.