

# The effect of memory and context changes on color matches to real objects

Sarah R. Allred · Maria Olkkonen

Published online: 31 March 2015  
© The Psychonomic Society, Inc. 2015

**Abstract** Real-world color identification tasks often require matching the color of objects between contexts and after a temporal delay, thus placing demands on both perceptual and memory processes. Although the mechanisms of matching colors between different contexts have been widely studied under the rubric of color constancy, little research has investigated the role of long-term memory in such tasks or how memory interacts with color constancy. To investigate this relationship, observers made color matches to real study objects that spanned color space, and we independently manipulated the illumination impinging on the objects, the surfaces in which objects were embedded, and the delay between seeing the study object and selecting its color match. Adding a 10-min delay increased both the bias and variability of color matches compared to a baseline condition. These memory errors were well accounted for by modeling memory as a noisy but unbiased version of perception constrained by the matching methods. Surprisingly, we did not observe significant increases in errors when illumination and surround changes were added to the 10-minute delay, although the context changes alone did elicit significant errors.

**Keywords** Color constancy · Memory · Color memory

---

S. R. Allred (✉)  
Department of Psychology, Rutgers, The State University of  
New Jersey, 311 N. Fifth Street, Camden, NJ 08102, USA  
e-mail: srallred@scarletmail.rutgers.edu

M. Olkkonen  
Department of Psychology, University of Pennsylvania,  
Philadelphia, USA  
e-mail: mariaol@sas.upenn.edu

## Introduction

Recognizing and identifying objects in the real world requires both perception and memory. Consider shopping for an item, such as a scarf, to match an item already at home, such as a pair of shoes (Fig. 1). Performing this task requires the cooperation of several systems: the shoes must be held in long-term memory, and then each potential scarf must be perceived and held in working memory while a judgment is made as to the best match. Upon reaching home, we may discover that the selected scarf and shoes do not match (Fig. 1, bottom right). Errors in such tasks are often attributed to memory, but the perceptual and memory processes involved in real-world tasks have rarely been studied jointly.

To understand why perceptual processes might introduce errors to real-world color memory tasks, consider again Fig. 1. Items at the store are viewed under different illumination and surrounded by different surfaces than they are at home. Thus, the light that reaches the eye from the scarf in the two locations is also different (two circles in Fig. 1). Maintaining stability of color appearance across such changes in the illumination and surrounding surfaces, an ability known as color constancy, poses a challenge for the perceptual system. Although failures of color constancy are rarely salient features of visual experience, a large body of empirical research demonstrates that color constancy is imperfect (see Brainard and Radonjic (2014), Foster (2011), Shevell and Kingdom (2008), and Smithson (2005), for review). The pattern of errors made by observers is well understood: observers typically fail to compensate entirely for the physical change in chromaticity of the light reflected by an object that is caused by a change in the illuminant.

Here we sought to address two questions: (1) How does long-term memory affect the representation (both bias and



**Fig. 1** Illustration of a real-world color memory task. Shopping for a scarf at the store (*top left*) to match shoes at home (*bottom left*) requires matching a perceptual representation in one scene to a memory

representation in another scene. The light reaching the eye from an object (*two circles, top right*) depends on the light source

variability) of color? (2) How does changing the context between encoding and retrieval—that is, adding the perceptual demands associated with color constancy—affect that long-term memory representation? To address these questions, we asked observers to identify the paint color of real, three-dimensional study objects as we independently manipulated the delay between seeing the study object and identifying its paint color, the illumination under which the matching paint color was identified and the color of the surface surrounding the study object.

#### Predictions for long-term memory

Consistent with intuition, there is consensus that memory introduces errors in color tasks, although there is considerable disagreement about the type and magnitude of errors introduced by memory for color. This disagreement about memory fidelity has arisen in part because memory for color is studied in disparate fields. Researchers motivated by understanding the structure of visual working memory have used color as a convenient stimulus with which to elucidate general principles of memory that are independent

of specific representational content. In contrast, researchers motivated by understanding the representation of color have used memory to make inferences about the representation of color at different levels of the visual system (cf. Allred & Flombaum, 2014; Brady et al., 2011; Olivers & Schreij, 2014). Although elaborating on the differences in approach is beyond the scope of this paper, we here outline some main conclusions.

There is a general consensus that the variability of color representation (1) increases over the few seconds of iconic and working memory; and (2) reaches a plateau with no further degradation in long-term memory. The first proposition is a robust finding that generalizes across the variety of different methodologies employed by both working memory research and color representation research (Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014; Brady et al., 2011; Fougny et al., 2012; Nemes et al., 2012; Olkkonen & Allred, 2014; Olkkonen et al., 2014; Uchikawa & Ikeda, 1981). We note, however, that the exact form of the decay and its dependence on memory load is highly contested within the visual working memory community. That variability increases is also consistent with work in other

visual dimensions (cf. Magnussen & Greenlee, 1999). The second proposition is more speculative, since relatively little research has focused on the fidelity of color memory through longer durations. However, Brady, Konkle, Gill, Oliva, and Alvarez (2013) recently reported that although objects were forgotten more frequently with long durations, those that were remembered at all were remembered just as precisely as those in working memory. We thus expect that adding a delay between sample and match will increase the variability of responses compared to a low-delay condition.

Although there is agreement that memory increases variability in the color representation, there is active debate within the color representation community about whether memory introduces systematic *biases* to the color representation. (On this point, visual working memory research is silent, since these paradigms typically have a built-in assumption that memory is unbiased). Given the number and contradictory nature of claims about color memory, it is difficult to make one clear prediction about the effect of memory on bias in color matches. For example, there are claims that color memory is systematically biased in the direction of greater saturation (Bartleson, 1960; Burnham & Clark, 1955; de Fez et al., 2001; Siple & Springer, 1983), or towards focal or prototypical colors (Bartleson, 1960; Collins, 1932; de Fez et al., 2001). Other work suggests that a number of performance benefits accrue to focal colors in memory (Boynton, Fargo, Olson, & Smallman, 1989; Heider, 1972; Uchikawa & Shinoda, 1996). However, other studies have failed to find any systematic biases in color memory (Hamwi & Landis, 1955; Hedrich et al., 2009; Jin & Shevell, 1996; Ling & Hurlbert, 2008; Nilsson & Nelson, 1981). In addition, most of this work focuses on working memory, and little work has examined biases in long-term memory.

We note that a separate body of research has investigated the distinct phenomenon of memory color (as opposed to color memory). In memory color, the typical color of an object (e.g., the yellow of a banana) influences either the memory or perception of an object's color (Duncker, 1939; Hansen et al., 2006). Here we focus primarily on color memory for objects without a typical color, and thus focus exclusively on literature relevant to color memory, as opposed to memory color.

Given the contradictory nature of claims about bias in color memory, but the overwhelming evidence in support of increased variability in memory, we developed an ideal observer model that assumes memory is a noisy but unbiased version of perception, and uses gamut characteristics to constrain responses. Such constraints can in principle lead to biases, if an observer's desired response is outside the palette gamut on some trials, but not on others. This can be termed a simplest possible model for memory against which other models could be compared.

## Changing the context between encoding and retrieval

Given that the effect of illumination changes on color perception is relatively well understood, there are three possibilities for how changing the illumination will affect color matches in memory. First, the sources of error in long-term memory could be independent of errors introduced by a change in the illumination. In that case, we would expect the errors to add. Second, changes in illumination could interfere with normal memory processes, causing errors that are even worse than one would expect from independent sources of error. Third, long-term memory and constancy processes could share common sources of error. In its most extreme instantiation, where all sources of error are common, adding a change in illumination would have no effect on long-term memory matches.

Despite the ubiquity of everyday tasks involving both memory and context changes, they are rarely studied together. In reduced-cue contexts, there is some evidence for the second alternative that memory interacts with context changes to reduce overall constancy (Olkkonen & Allred, 2014). Perhaps surprisingly, however, the limited data available in realistic scenes are consistent with the third hypothesis. Two studies that measured color constancy with or without a delay concluded that memory did not decrease color constancy (Jin & Shevell, 1996; Uchikawa et al., 1998).

To summarize, we sought here to examine long-term memory and its relationship to color constancy; that is, how long-term memory for color relates to perception of color, and how memory is affected by changes in context between encoding and retrieval.

## Materials and methods

### Participants

Participants were 122 college students participating for course credit, with normal color vision as assessed by the Ishihara Color Plates. Procedures were approved by the Rutgers University IRB (Protocol E10-410).

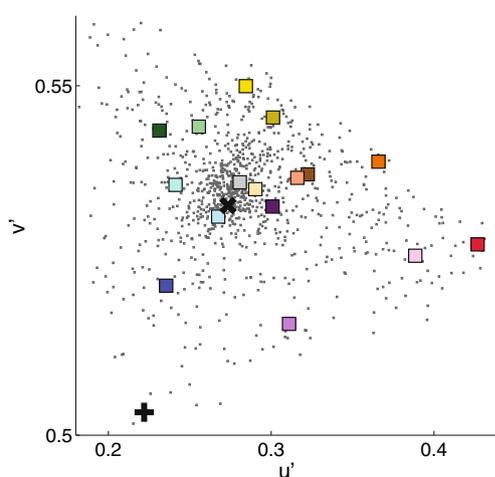
### Apparatus and Stimuli

Observers viewed two adjacent  $4' \times 4' \times 4'$  grey flat matte booths. Illumination in the room was provided separately for each booth (chromaticity in CIE  $Lu'v'$  space; Booth A:  $u' = 0.274, v' = 0.533$ ; Booth B:  $u' = 0.222, v' = 0.503$ ). There was no other source of illumination in the room, and we verified that the booths did not illuminate each other. Mounted 4.5 feet from the front of each booth was a book of 1,022 paint chips

(Sherwin-Williams, 2010) that served as a matching palette (Fig. 3). Sixteen  $3'' \times 3'' \times 3''$  cubes (subtending  $4.5^\circ - 6.5^\circ$  at usual viewing distances), painted with different colors of flat matte paint chosen to approximately span color space (see Fig. 2) served as stimuli. In addition, eight painted plywood shapes (see Fig. 3) served as backgrounds for cubes on half of the trials. Cube, background, and palette chromaticity measures were made using a Spectrascan PR-655 spectroradiometer (Photo Research Inc, Chatsworth, CA) and are reported in Allred and Olkkonen (2013). Luminance information was discarded. The illumination across each booth was non-uniform, and in combination with the three-dimensional nature of the stimuli, this meant that the amount of light reflected by a surface depended on its location within the booth. This can be seen clearly in Fig. 3, where the top surface of each cube reflects more light to the observer than either side. Radiometer measurements verified that although luminance varied across cubes, chromaticity remained approximately constant. Each cube was always presented in the same location within a booth.

## Design

Participants made color matches to a total of 16 cubes by manually searching through the palette and indicating



**Fig. 2** Chromaticity in CIE  $u'v'$  space of all 1,022 paint chips (black dots), cube stimuli (colored squares), and booth illumination (Booth A: black cross; Booth B: black plus symbol). Chromaticity measures for all chips and stimuli are under (Booth A) illumination. Chromaticity measures were taken with a PR-655 spectroradiometer, and luminance information was discarded. Radiometer measurements of each cube were taken on the top face of the cube, in the corner closest to the observer, while the cube was in its experimental location. Radiometer measurements of paint chips were taken in the center of each chip, with the palette in its mounted location. Square color indicates the cube's apparent color. The black line surrounding the squares is to increase figure visibility; no such border surrounded cubes in the study. The relatively small range of  $v'$  chromaticity results from the narrow bands in the illuminants; paint chips and cubes spanned a large reflectance range

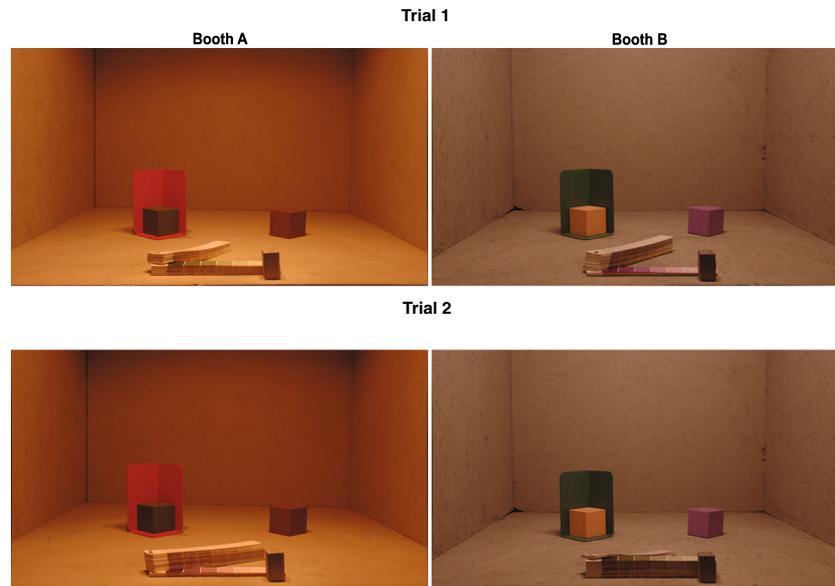
the chip that best matched the cube's paint. Participants matched two cubes each in the eight different conditions (see Table 1) resulting from a  $2 \times 2 \times 2$  factorial design with background, illumination, and delay as factors. On each of four trials, participants made four color matches (see Table 2 for example). To do so, observers viewed four cubes (see Fig. 3), two in each booth. One cube in each booth was embedded in a background (left cubes each panel of Fig. 3). Thus, both levels of the background factor (background/no background) were present on each trial. The four trials varied the other factors as described in Table 2. In one trial, observers simultaneously viewed cubes and matching palette in the same booth (Fig. 3 top panel, baseline, and background conditions). In a second trial, observers looked back and forth between booths while viewing cubes in one booth and matching palette in the other booth (Fig. 3, bottom panel, illumination and illumination + background conditions). A third trial was the same as the first, except that participants had 1 min to memorize cube colors before a 10-min delay during which the cubes were removed from the booth and observers completed a survey or homework to prevent them from rehearsing verbal labels for the colors. Participants then returned to the room and made the color match in the absence of the cubes (memory and memory + background conditions). A fourth trial added the 10-min delay to the change in the illumination (illumination + memory, and illumination + memory + background, conditions). The conditions and trials are summarized in Table 1 and Table 2.

The data for baseline, illumination, background, and illumination + background conditions were described elsewhere (Allred & Olkkonen, 2013). The purpose of that paper was to characterize color constancy of real, 3D objects across changes in the background and changes in illumination. In that paper, we discuss motivations for choosing particular stimuli, and the color-specific effects of different backgrounds. Here we focus on the effect of memory itself and how it interacts with constancy. Because of the complexity involved, we felt the analysis and interpretation of the constancy data itself was more appropriately addressed in a distinct paper.

To prevent sequence effects, we used a between-observers design. Thus, participants viewed each cube only once, and performed one trial for each of the eight conditions. Between-observers color matches for each cube were made in all eight conditions. The order of the trials was counterbalanced between observers.

## Procedure

On each trial, observers entered the experimental room and were seated in front of the two booths on a rolling chair. The door to the room was closed, and the booth lights provided



**Fig. 3** Photograph of experimental setup for a two-example trials. On each trial, observers viewed four cubes, two cubes each in (Booth A) (*left images*) and (Booth B) (*right images*) that were separately illuminated. On each trial, one cube in each booth was embedded in a 3D background (for this condition, left cubes in each image). The matching palette (booklet in the front of each booth) contained 1,022 paint chips. The palette in each booth rotated freely on a long screw mounted into palette, and the wooden stopper prevented observers from pulling palette strips out of the booth. Observers were permitted to flip freely through the book, but were instructed to choose a match only when the palette strip was aligned with the stopper. On the trial labeled 1,

(baseline and background conditions) observers chose color matches from the palette mounted in the same booth as the cubes. To illustrate this, the palette is open to the green section (Trial 1, Booth A) and the purple section (Trial 1, Booth B). On Trial 2 (illumination and illumination+background conditions) observers chose color matches for a cube from the palette mounted in the other booth. As illustrated, the color match for the green cube (Booth A) was selected from the palette in (Booth B), and the color match for the purple cube (Booth B) was selected from the palette in (Booth A). On trials with a delay (not shown), no cubes or backgrounds were present while observers chose color matches

the only source of illumination. On each trial, observers adapted to the room illumination while the experimenter described the task. This typically took several minutes. Observers were free to move about on the chair, but were instructed to remain outside of the booths. To make a match for a cube in each condition, observers searched through the book of paint chips, chose the paint chip that best matched the paint of the cube, and wrote the number of that paint chip

on a response sheet. To reduce illumination variation across the samples, observers were instructed to make their choice while the strip of paint chips was positioned against the wooden stopper. No specific instructions were given about what area of the cube should guide their choices.

On each trial, participants made matches from the palette samples under each illuminant. Thus, it is unlikely that observers were completely adapted to either illuminant.

**Table 1** Description of the eight different conditions resulting from a  $2 \times 2 \times 2$  complete factorial design with background, illumination, and delay as factors

Condition	Background	Cube / Palette Illumination	Delay
Baseline	None	Same (A/A or B/B)	None
Memory	None	Same (A/A or B/B)	10 minutes
Background	Yes	Same (A/A or B/B)	None
Background + Memory	Yes	Same (A/A or B/B)	10 minutes
Illumination	None	Different (A/B or B/A)	None
Illumination + Memory	None	Different (A/B or B/A)	10 minutes
Background + Illumination	Yes	Different (A/B or B/A)	None
Background + Illumination + Memory	Yes	Different (A/B or B/A)	10 minutes

The first column is the condition name. The second, third, and fourth describe the level of the factors of background, illumination, and memory, respectively. The illumination factor refers to whether observers chose the matching paint chip in the same booth as the cube (Same) or in the booth other than the one in which they viewed the cube (Different). A or B designations refer to booth identity

**Table 2** Example trial order and conditions in each trial for one observer

Order 1		Booth A(1)	Booth A(2)	Booth B(1)	Booth B(2)
Trial 1	Cube	1	2	3	4
	Condition	Baseline	Background	Baseline	Background
Trial 2	Cube	5	6	7	8
	Condition	Memory	Memory + Back	Memory	Memory + Back
Trial 3	Cube	9	10	11	12
	Condition	Illum	Illum + Back	Illum	Illum + Back
Trial 4	Cube	13	14	15	16
	Condition	Mem + Illum	Mem + Illum + Back	Mem + Illum	Mem + Illum + Back

We counterbalanced to ensure that conditions were experienced in different orders, and this led to eight different orders of trials. The *left column* indicates the trial number. For each trial, the *top row* indicates the cube number, and the *second row* indicates the condition for that cube. Columns 3–6 indicate the Booth (A or B) and the location within the booth (leftmost = 1, rightmost = 2) where the cube was presented). Across orders, cube location remained the same

### Data and analysis

We excluded seven of 122 observers for not recording a response for at least 10/16 cubes or systematically recording cube color in the incorrect location. From the remaining 1,840 trials (115 observers  $\times$  16 matches per observer), we discarded trials for the following reasons: failure to record a decipherable response (184/1,840, 10 %), errors in palette chip chromaticity measures (32/1,840, about 2 %), and color match of a clearly different, non-adjacent color category to the cube (89/1,840, 5 %).

To calculate the average color match in each condition, we discarded luminance information and computed arithmetic means in both chromatic dimensions. We completed analysis in both CIE  $u'$  and  $v'$ , and CIE  $a^*b^*$ . CIE  $L^*a^*b^*$  values were computed from XYZ tristimulus coordinates using colorimetric routines in the Psychophysics toolbox, and using the measured booth illuminant as the white point. For each condition, variability was defined as the average distance in CIE  $u'v'$  (or  $a^*, b^*$ ) space between each individual match and the average match for that condition. Because the transformation between color spaces is non-linear, analysis of data in the two spaces gives slightly different results, but the patterns are broadly similar. For the most part, we report  $u', v'$  values; exceptions are for analyses involving hue angle and saturation. The color space used in an analysis is clear in the text and figure axes and captions.

We compared color matches between conditions using an atheoretic bias calculation. In separate analyses, we defined the correct color match in two ways: (1) Correct color was defined as the cube's measured chromaticity in the booth where the match was made; or (2) Correct color was defined as the average color match in the baseline condition. The motivation for using the average match in the baseline condition as an operational definition of "correct" is as follows: Here, we are primarily interested in how

experimental manipulations alter color matches. In an ideal world, matches in the baseline condition would be identical to the physical chromaticity, and the two definitions of correct would be identical. However, because the booths themselves contained some variation in illumination, this second definition allowed for the possibility of some non-veridical matches in the baseline condition resulting from luminance heterogeneities or because the surfaces surrounding each paint chip differ from the surfaces surrounding the cube. Without this second definition, we could confound illumination effects in the baseline condition with experimental manipulations such as memory or context changes between encoding and retrieval. Fortunately, the pattern of results was very similar with both measures of correct responses. Measures are clearly indicated in the text. In conditions with an illumination shift, the correct chromaticity was calculated by taking the chips chosen in the baseline condition, but averaging their chromaticity measured under the illumination shift.

For a more complete discussion of the apparatus and methodology used in this experiment, including choice of illumination, equations for analysis, stimulus characterization, stimulus-specific background effects, matching palette specification, and controls for the effect of palette non-uniformity, see Allred and Olkkonen (2013).

All data analyses were performed using Matlab. Except where noted,  $p$  values are uncorrected results from two-tailed paired  $t$  tests. All conversions between color spaces were made from the radiometer measurements using colorimetry routines in the PsychToolbox (Brainard, 1997).

### Ideal observer model

To investigate possible sources of error in memory matches, we developed an ideal observer model that operationalized the idea that memory is a more variable but unbiased

version of perception. The model is not meant to implement the neurophysiology of color perception or memory. Indeed, the neural representation of color varies depending on brain area and is not likely to match well to any particular representation of color space (for discussion, see Gegenfurtner (2003) and Hansen et al. (2007)). Instead, this model is meant as a conceptual preliminary investigation of the relationship between perception and memory.

First, we assumed that color is represented as a multivariate Gaussian distribution in CIE  $u'v'$  space. Next, we assumed that  $u'$  and  $v'$  are independent of each other, so  $(u, v)$  can be characterized as  $u \sim N(\mu_u, \sigma_u^2)$  and  $v \sim N(\mu_v, \sigma_v^2)$ . We define  $\mu_{baseline}$  as the chromaticity of the average match in the baseline condition. If the chromaticity of the  $i^{th}$  observer's match is represented as  $(u_i, v_i)$ , then the average chromaticity for  $u$  and  $v$  are the respective means across the sets  $\{u_1, u_2, \dots, u_n\}$  and  $\{v_1, v_2, \dots, v_n\}$  in the baseline condition, where  $n$  is the total number of observers. Similarly,  $\sigma_{baseline}$  for  $u$  and  $v$  dimensions is calculated as the standard deviation of the set  $\{u_1, u_2, \dots, u_n\}$  and  $\{v_1, v_2, \dots, v_n\}$  in the baseline condition, respectively. Thus, for each cube we have  $\mu(u, v)$  and  $\sigma(u, v)$  in the baseline condition.

To model the effect of memory, we assumed that the memory representation had the same mean chromaticity as the baseline, so that  $\mu_{memory} = \mu_{baseline}$ . We modeled memory of each cube as increasing the variability of the baseline representation for that cube. Specifically, for each cube, we assumed that in both  $u$  and  $v$  dimensions, the standard deviation increased by a scalar factor  $k$ , such that  $\sigma_{memory} = k * \sigma_{baseline}$ .

To link the model to human performance, we ran 25-trial simulations for each cube. The memory chromaticity on each trial was a draw from the distributions defined above, and the ideal observer chose the chip that minimized the Euclidean distance between the chip and the chromaticity of the draw on that trial. Average chromaticity for the 25 ideal observer choices was calculated in the same way as for the behavioral data.

### Measuring prototypical hues

After observers completed the main experiment, we asked them to flip through the paint palette and write down the number of the paint chip that corresponded to the best example of red, blue green, yellow, orange, pink, purple and brown. Observers were randomly assigned to Booth A or Booth B. We found no effect of booth on the chips chosen for a given color name, and thus we collapsed across booths. Hue angle for prototypical colors was determined by using polar coordinates of CIE  $L^*a^*b^*$  values converted from CIE  $Lu'v'$  space. The measured background of the booth was used as the white point in the conversion

**Table 3** Names (*first column*) and hue angle of baseline matches (*second column*) for all 16 cubes in the order shown in Fig. 7

Cube name	Cube hue	Prototypical color category	Prototypical hue angle (average)
Light green	124°	Green	144°
Orange	62°	Orange	59°
Brown	62°	Brown	57°
Dark green	149°	Green	144°
Purple	318°	Purple	313°
Red	38°	Red	37°
Yellow	88°	Yellow	86°
Dark blue	242°	Blue	247°
Pink	20°	Pink	14°
Ice Blue	140°	Green	144°
Plum	24°	Pink	14°
Peach	55°	Brown	57°
Gold	76°	Yellow	86°
Aqua	205°	Blue	247°
Gray	80°	Yellow	86°
Doeskin	53°	Brown	57°

The prototypical color category (*third column*) and average hue angle (*fourth column*) corresponding to each of the cubes are also shown. For cubes above the horizontal space, the closest prototypical hue was straightforward; for those following, each cube was assigned to the nearest prototypical hue in terms of angular distance

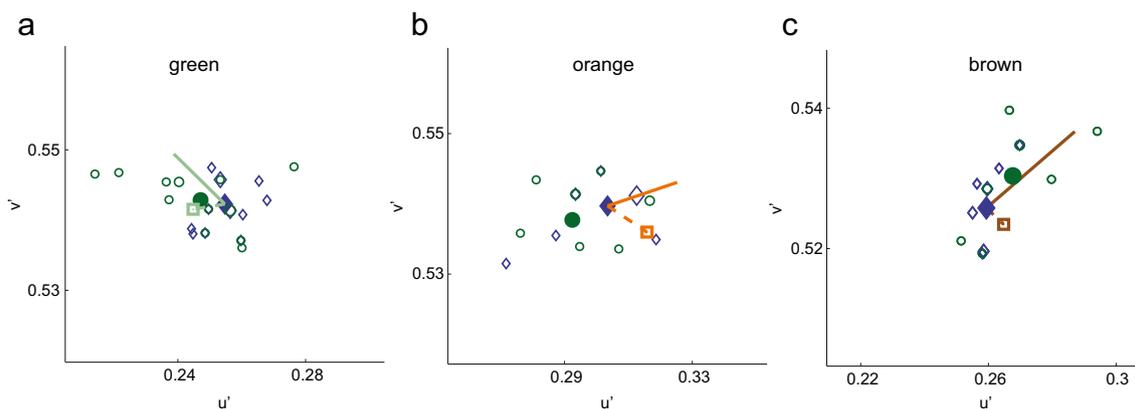
calculations. For this calculation, saturation values were discarded. Hue angles of prototypical colors are reported in Table 3.

## Results

There were two main goals of this paper: (1) Investigate the role of memory in color matching tasks; (2) Measure the impact of context changes on that role. To these ends, we first compare errors elicited by memory demands to those in the baseline condition and explore potential causes of memory errors. Next, we describe the pattern of errors elicited when perceptual demands were added to memory demands.

### Memory errors

All participants' color matches for the dull green, orange and brown cubes in the baseline and memory conditions are shown in Fig. 4. In the baseline condition (unfilled blue diamonds), participants chose many different paint chips. This range of color matches in the baseline condition was a common feature across all cubes (median number of paint chips chosen in baseline condition = 7, min = 4, max = 11; median number of observers per cube = 12; data



**Fig. 4** Chromaticity of individual (*open symbols*) and average (*filled symbols*) color matches for the dull green (**a**), orange (**b**), and brown (**c**) cubes in the baseline (*blue diamonds*) and memory (*green circles*) conditions. The *open square* in each plot is the average of the prototypical color matches (see Methods) for that cube. The *solid lines* represent the direction of increased saturation, which was defined in

CIE  $L^*a^*b^*$  as increasing distance from the origin at the hue angle of the average match. A point along this line was chosen, and these CIE  $a^*b^*$  coordinates were transformed into CIE  $u'v'$ . The size of individual data points is proportional to the number of observers choosing that paint chip. In (**a**), (**b**), and (**c**), the x- and y-axis ranges are identical, though the starting point shifts to accommodate the data

not shown). The baseline condition, where observers view cubes and make matches simultaneously and under the same illuminant, represents the best performance we could expect to achieve in this task. Given this, the trend of variability in baseline color matches is reassuring and provides a useful sanity check: the paint palette was sufficiently discretized to provide a reasonable estimate of between-observers variability in color perception. The region of space spanned by color matches also varied by cube.

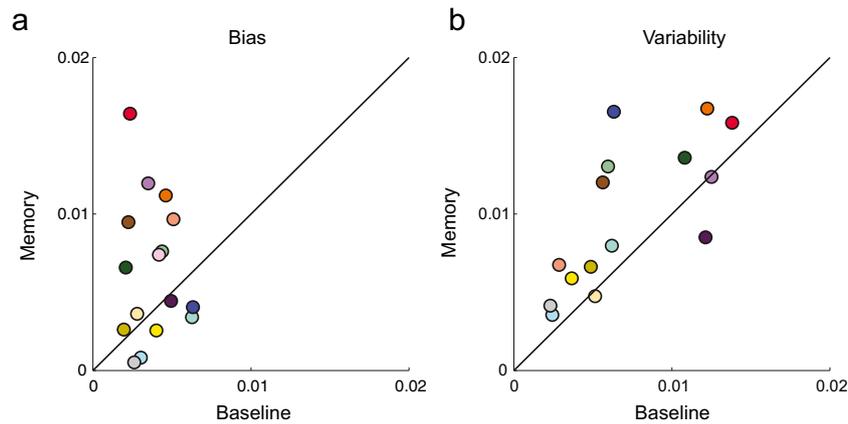
To understand how memory affected color matches, we compared observers' memory matches to baseline matches in light of possible alternative outcomes (Fig. 4, green circles). In each panel, the solid line represents increasing saturation: if memory were saturated compared to perception, then the average memory match (solid green symbol) would fall somewhere along this line. In each panel, the small square represents the focal color associated with the cube: if memory were biased toward focal colors, then the average memory match (green circle) should fall along the dashed lines connecting the baseline match (solid blue diamond) to the focal color (small empty square). If, on the other hand, memory did not introduce bias but was more variable, then the memory average (solid green circle) would be very similar to the baseline average (solid blue diamond), but individual memory matches (unfilled green circles) would cover a larger region of color space than the individual baseline matches (unfilled blue diamonds).

For these individual cubes, memory elicited a small bias (Fig. 4), shown here as the distance in color space between the chromaticity of the average memory match and the chromaticity of the baseline match. Biases were not consistently in the direction of greater saturation (solid

lines) nor toward focal colors (dashed lines). Unsurprisingly, memory matches for these cubes did seem to be more variable than baseline matches, as evidenced by the green circles spanning a larger range of color space than the blue diamonds.

The trends evident for the individual cubes continued when data were aggregated across cubes (Fig. 5, left panel). To gauge the size of the memory bias, we compared it to the split-half baseline bias, calculated by randomly dividing baseline color matches into two groups and calculating the distance between the group averages (x-axis, Fig. 5, left panel). On a cube-by-cube basis, the memory bias was on average 1.96 times the baseline error ( $p < 0.05$ , two-tailed paired t-test.) Unsurprisingly, memory matches were also consistently more variable across cubes (Fig. 5, right panel,  $p < 0.005$ ). On a cube-by-cube basis, memory matches were 1.57 times more variable than matches in the baseline condition.

There is clearly a great deal of variability in the size of the memory bias. Some cubes, such as red and purple, exhibit large memory biases, while other cubes, such as gray and doeskin, exhibit small biases. To gain more insight into the relative sizes of these biases, we performed permutations testing. To do so, we created a vector with all baseline and memory matches. Next, we randomly assigned each match to either the memory or baseline conditions, with the constraint of keeping the original number of trials per condition. Next, we computed the bias between these two groups of data, and we repeated this process 100 times. This provides a sense of how often various biases could arise by chance given the response palette. If the observed bias is larger than most of these permuted biases, it suggests the bias we

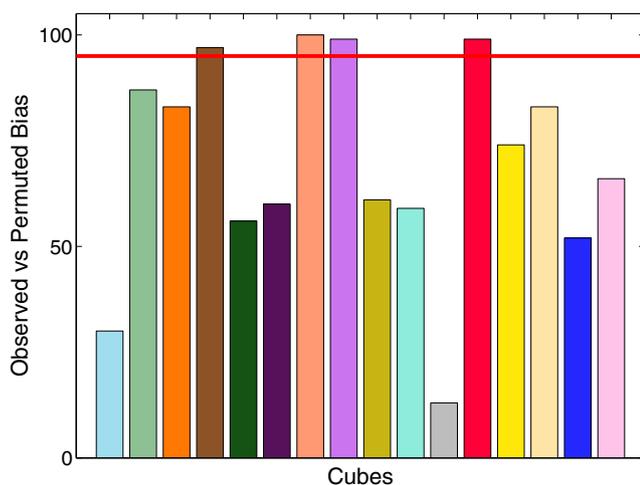


**Fig. 5** Memory increases bias (a) and variability (b) of color matches. a) Bias for each cube was calculated as the average of the Euclidean distance between the ( $u'$ ,  $v'$ ) coordinates of the measured cube chromaticity and the baseline matches (x-axis) and the memory matches (y-axis). b) Variability for each cube was calculated by

summing the distance between each color match and the average color match in the baseline (x-axis) and memory (y-axis) conditions and dividing by the total number of matches. In both a) and b) the solid black line represents unity; each color represents the apparent color of a different cube

observed is unlikely to have arisen by chance. Such permutations testing is particularly useful in a data set such as ours, where the underlying response space is known to violate assumptions of normality. This should be viewed as a conservative test. The results of the permutations tests are plotted in Fig. 6. For 4/16 cubes, the measured bias was larger than the permuted bias 95 of 100 times, which can be thought of as significant at a  $p = 0.05$  level.

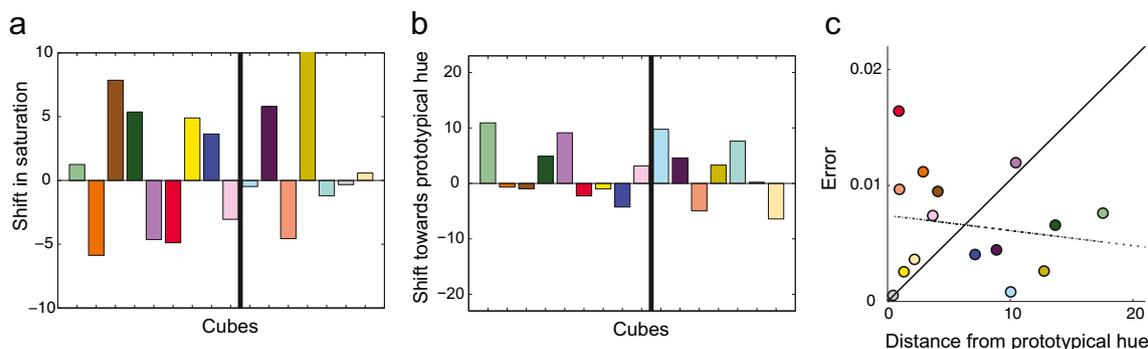
Thus far we have established the size and significance of the memory bias. Next we examine sources of errors in memory.



**Fig. 6** Relative significance of the memory bias. The y-axis represents the number of times that the observed memory bias was greater than the permuted memory bias. Permuted bias was calculated as described in the text. Bar color represents cube color. Horizontal red line is  $n = 95$ , reflecting a significance level of  $p = 0.05$ . The significance of the permuted bias depends on the number of measurements, the size of the observed bias, and the discretization of the matching palette at the cube's chromaticity

First, we evaluated whether memory matches are more saturated or greater in chroma than baseline matches. It is convenient to use CIE  $L^*a^*b^*$  space for this analysis, since the distance of a point ( $a_1$ ,  $b_1$ ) from the origin represents the saturation of that point. To do so, we first converted average matches from CIE  $Lu'v'$  space to CIE  $L^*a^*b^*$  space using the measured booth chromaticity as the white point for the conversion (see Methods). In Fig. 7a, we plot the difference in saturation between memory matches and baseline matches. If memory matches were more saturated than baseline matches, we would expect each cube to be above the zero line. Instead, on average, there was no significant difference between saturation of memory and baseline matches (mean difference = 0.91;  $p = 0.65$ ).

A second extant claim about color memory is that observers are biased towards prototypical hues. To evaluate this claim, we compared hue of memory matches to both baseline hues and prototypical hues. Prototypical hues for each observer were determined after the main experiment, as described in Methods. In Fig. 7b, we plot the shift in memory towards prototypical hue for each cube. Values above zero indicate that memory matches were closer in hue to the prototypical color than baseline matches were. For each cube, the comparison prototypical hue was defined as the hue angle closest to the baseline match (see Table 3). For the nine cubes to the left of the vertical line in Fig. 7b, the assignment of cube to prototypical color seemed natural. For the rest of the cubes, the assignment seemed less natural; for example, the cube we called aqua could be labeled as either blue or green, and the cube we called gold does not seem to be closely aligned with any of the prototypical colors. If memory of a hue were systematically shifted toward the relevant prototypical hue, the bars should be consistently above zero in Fig. 7b. Instead, averages were close to zero both for



**Fig. 7** Potential sources of memory bias. *Bar/symbol colors* represent apparent cube colors. **a)** Difference between saturation of matches in memory and baseline conditions for all cubes. For each cube, the average match in CIE  $L^*u^*v^*$  space was converted to CIE  $L^*a^*b^*$  using the chromaticity of the booth as the white point (see Methods). Lightness information was discarded, and saturation for each point was defined as its distance from the origin. *Positive values* indicate that memory matches were more saturated than baseline matches. **b)** Angular shift towards prototypical hue of memory matches compared to baseline matches. Prototypical hues were determined for eight color categories (*blue, yellow, green, red, orange, brown, pink, purple*) as described in the Methods. Hue angle of color matches was determined by converting  $a^*b^*$  values into polar coordinates. Y-axis values were determined

by subtracting the distance between memory matches and the nearest prototypical hue from the distance between baseline matches and the nearest prototypical hue. *Positive values* indicate that memory matches were closer to prototypical hue than baseline matches. Cubes were divided between those that were plausibly related to a prototypical hue (*bars left of solid vertical line*) and those that were not (*bars right of solid vertical line*). See Table 3 for hue angle values of cubes and prototypical colors. **c)** Relationship between memory bias and distance from prototypical hue. X-axis represents angular distance of average baseline match from the nearest prototypical hue as defined in (b). Y-axis is memory bias as reported in Fig. 5a. *Solid black line* is the identity. *Dashed black line* is best fit line through data

cubes to the left of the vertical line ( $mean = 2.10^\circ, p = 0.27$ ) and for all cubes ( $mean = 2.06^\circ, p = 0.15$ ).

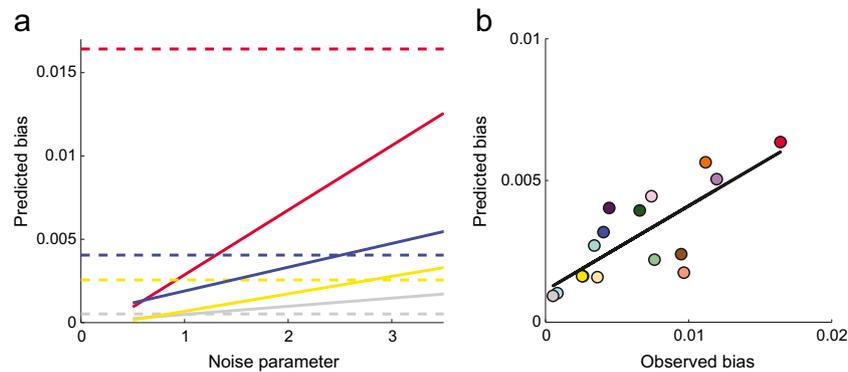
Although we found no systematic bias of memory toward prototypical hue, we note that many of the cube stimuli were very similar in hue to the measured prototypical colors. Furthermore, cube stimuli varied in saturation as well as hue.

A third claim, related to the second, is that the fidelity of memory for a color might depend on its distance from a prototypical color. To evaluate this claim, we computed the distance of each cube’s hue from the closest prototypical hue, and compared that to the memory bias for that cube (Fig. 7C). The slope of the best fit line is shown. A positive slope would mean that cubes closer to a prototypical hue were remembered more accurately; we found a small but non-significant correlation ( $r = -0.22, p = 0.41$ ).

The previous analyses showed no systematic relationship between the fidelity of memory and the saturation or hue of cubes. Next, we investigated whether the increased variability of matches in memory (Fig. 5, right panel) combined with the discretization of the palette could account for the memory biases of the sort we observed in the data. To motivate this idea, consider the red cube, for which we measured a relatively large memory bias (Fig. 5, left panel). The chromaticity of the red cube is near the right-most edge of the palette gamut, as seen in Fig. 2, where color space is relatively sparsely sampled by the paint chips. Suppose that, instead of being biased, memory is simply more

variable than perception. The remembered chromaticity of a given chip on individual trials will deviate from the average, and on trials where that deviation is to the right of the red cube, no paint chips exist with a similar chromaticity. If, on those trials, observers choose the paint sample closest to the chromaticity in memory, their match will be shifted leftward from their actual memory. However, no such shift will occur on trials where the deviation is to the left of the red cube, because color space is more densely sampled there. In principle, then, increasing the variability could shift the average match for the red cube to the left, even if the mean of the memory representation is the same as the perceptual representation.

We developed an ideal observer model to operationalize this idea, the details of which are described in the methods. Briefly, we assumed colors to be represented as noisy multivariate Gaussian distributions. We further assumed that the means of the memory and perceptual representations are identical, but that the standard deviation of the distribution in memory varies as a multiplicative factor  $k$  of the perceptual distribution. On each simulated trial, the ideal observer chooses the paint chip with the chromaticity closest to the memory representation, which is a random draw from the memory distribution. We then computed the size of the simulated bias as the difference between the mean of the simulated memory trials and the actual baseline match. This procedure was repeated for 35 different values of  $k$ , resulting in simulated bias as a function of increasing variability



**Fig. 8** Ideal observer model. **a)** Relationship between variability parameter  $k$  and memory bias predicted by the ideal observer model (as described in Methods) for the four indicated cubes. *Horizontal dashed lines* indicate the memory bias observed in the data. Lines are best fits to model predictions for values of  $k$  that ranged from 0.5 to

3.5 in 0.1 increments. **b)** For  $k = 1.6$ , relationship between observed memory bias (calculated as in Fig. 5a) and bias predicted by the ideal observer model for all 16 cubes. Points are extracted from fits shown in **a)**. Black line is best fit through data. Note that x- and y-axis ranges differ

for each cube. The simulated biases were well-fit by lines, which we plot for four example cubes in Fig. 8a. For reference, the dashed lines represent the observed memory bias for each cube. There are two salient features. First, simulated bias increases as variability increases. If there were no relationship between variability and bias, we would expect the lines to be flat. Second, the slope of the lines is dependent on cube. Variability affects bias for the red cube, which is close to the edge of the palette gamut in Fig. 2, more than it affects bias for the gray cube, which is close to the center of the gamut in Fig. 2. For these cubes, the magnitude of the simulated bias seems well correlated with the measured bias.

In Fig. 8b, we plot simulated bias as a function of observed bias for all 16 cubes, for one value of  $k$ , which is the same as the average factor with which the standard deviation in memory increased compared to baseline ( $k$  and observed increase = 1.6). Simulated bias values

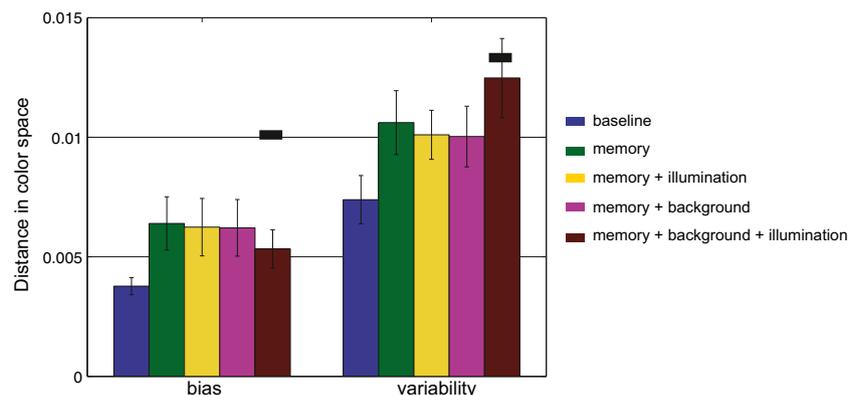
were extracted from the lines fit in Fig. 8a. The correlation between simulated and measured bias was significant ( $r = 0.75$ ;  $p < 0.001$ ).

To summarize, an ideal observer model that characterizes memory as a noisier but unbiased version of perception results in a pattern of biases that is similar to that observed in the data.

#### Adding context changes

Thus far, we have examined the effect of memory on color matches. Next, we investigated the effect of adding illumination and background to memory.

A change of context between encoding and retrieval seemed to have little effect on either bias or variability of color matches (Fig. 9). In Fig. 9, we plot the average bias and variability in each condition employing memory, as well as the baseline condition for reference. If



**Fig. 9** Adding context changes has little effect on bias in memory matches. Bias (*left bars*) and variability (*right bars*) for memory alone (*green bars* in each condition) and memory with the addition of context changes of illumination (*yellow bars*), background

(*magenta bars*), or illumination and background together (*brown bars*). For reference, *blue bars* indicate error (*left: split-half error; right: variability*) in the baseline condition. *Error bars* are s.e.m. across cubes

changing the illumination, embedding the object in a surround, or doing both increased bias or variability, then the green bars representing the memory error in Fig. 9 would be lower than the others. Instead, they are approximately the same height, with overlapping error bars. Paired two-tailed *t* tests revealed no significant differences between bias or variability in the memory condition compared to memory along with any context change (i.e., illumination, background, or illumination + background). For reference, we also include the errors in the baseline condition (blue bars).

One trivial explanation for such a null result is if changes of context alone produced no errors in observers' matches, that is, if color constancy is perfect. However, as we previously reported when we analyzed the no-delay data separately, a change in the illumination *without* the 10 minute delay did significantly increase bias and variability compared to baseline (Figs. 6 and 7 of Allred and Olkkonen (2013)). In fact, as calculated here, the illumination change without the 10-min delay produced errors roughly comparable in magnitude to the error elicited by the 10-min delay alone (bias: *illumination* = 0.0075, *memory* = 0.0064; variability: *illumination* = 0.011, *memory* = 0.010, data not shown). Thus, the 10-min delay appears to interact with the illumination change; errors in the delay + illumination condition are smaller than would be expected if they added. The additive prediction is shown in Fig. 9 as the solid black bar, with the assumption that baseline error (blue bars) is common to all predictions.

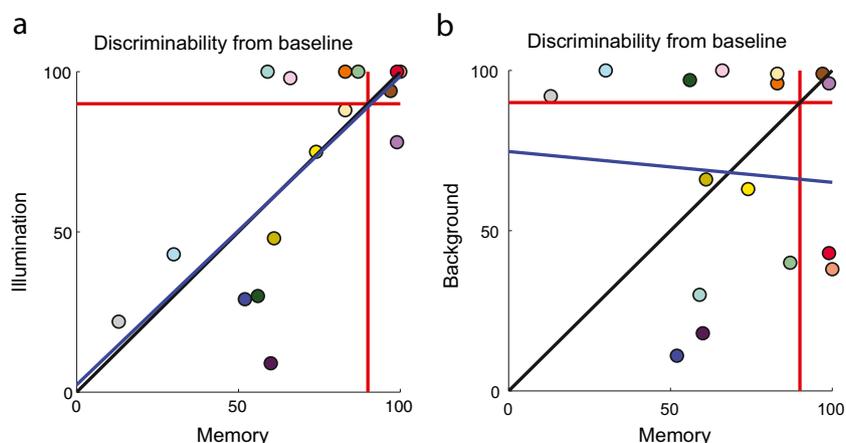
An additional indication that context changes without a delay can elicit errors comes from permutations analysis

of the matches in the illumination condition and the background condition. As a reminder, the axis values represent the number of times that the observed bias was greater than the permuted bias; large values (i.e., > 90, here) represent biases that are significant at the  $p = 0.10$  level. We compared the relative significance of the observed bias in the illumination (Fig. 10a) and background (Fig. 10b) conditions to the relative size of the memory bias. Similar numbers of cubes elicited significant biases in the illumination ( $n = 7$ ) and background ( $n = 8$ ) conditions as in the memory condition ( $n = 4$ ). Interestingly, the correlation between illumination and memory biases was strong ( $r = 0.73$ ,  $p < 0.005$ ), whereas there was very little relationship between background and memory biases ( $r = -0.07$ ,  $p = 0.79$ ).

To summarize, adding context changes to a 10-min delay does not significantly alter the bias or variability of color matches.

Although adding context changes to the demands of memory did not appreciably alter the bias of color matches, we completed the same analysis about the source of errors that we did for color matches in memory. We continued to find no effect of memory on saturation of matches. A four-way ANOVA with cube as a random effect, and memory, illumination and background as fixed effects showed no main effect of memory ( $F(1,105) = 0.41$ ;  $p = 0.52$ ) on saturation.

We also examined the shift toward prototypical hue from baseline hue for all conditions. Again, a four-way ANOVA showed no main effect of memory ( $F(1,105) = 0.39$ ;  $p = 0.53$ ) on hue shift.



**Fig. 10** Discriminability between baseline and experimental matches determined with permutations testing. Axis values represent the number of permuted biases that were smaller than the observed bias in the data. Larger axis values indicate that the color matches in the labeled condition (memory, x-axis; illumination: y-axis in (a); background: y-axis in (b)) were more discriminable from color matches in

the baseline condition. *Red horizontal lines* indicate values of 90, so data points above or to the right of this line indicate that color matches in the two conditions were discriminable at the  $p = 0.10$  level. *Circle color* is an approximation of cube's apparent color. *Black lines* are identity lines, and *blue lines* are best fit to data in a linear least squares sense

Similarly, we again found no significant correlations between distance to prototypical hue and memory bias for any condition involving memory (illumination + memory:  $r = 0.06$ ,  $p = 0.81$ ; background + memory:  $r = -0.02$ ,  $p = 0.94$ ; background + illumination + memory:  $r = 0.07$ ;  $p = 0.78$ ).

Next, we investigated whether the ideal observer model outlined in Fig. 8a and b continued to explain the effect of memory when perceptual demands were added. To do so, we repeated the ideal observer process, but we substituted the relevant perceptual demand for the baseline condition. Thus, to predict performance in the memory + illumination condition, we assumed that  $\mu_{\text{memory} + \text{illumination}} = \mu_{\text{illumination}}$ . This allows for the likelihood that the illumination significantly shifts average color matches (as would be expected from the incomplete color constancy). We continued to model  $\sigma$  in each memory condition as a multiplicative factor of  $\sigma$  in the baseline condition, so that  $\sigma_{\text{memory} + \text{illumination}} = k^* \sigma_{\text{baseline}}$ . In these calculations, bias was calculated as the distance in color space between matches in the baseline condition and matches in the relevant experimental condition. As with the original model, we then correlated the predicted and observed biases for each cube as a function of  $k$ .

In Fig. 11, we plot the correlation between observed and predicted bias when memory is added to baseline, (blue

line), background (green), illumination (brown) and illumination + background (yellow), and heavy weight lines indicate significant correlations. Although correlations were highest for baseline, they were also significant when memory was added to the other conditions. Thus, modeling the effect of memory as an unbiased but more variable version of a perceptual representation continues to explain part of the effect of memory.

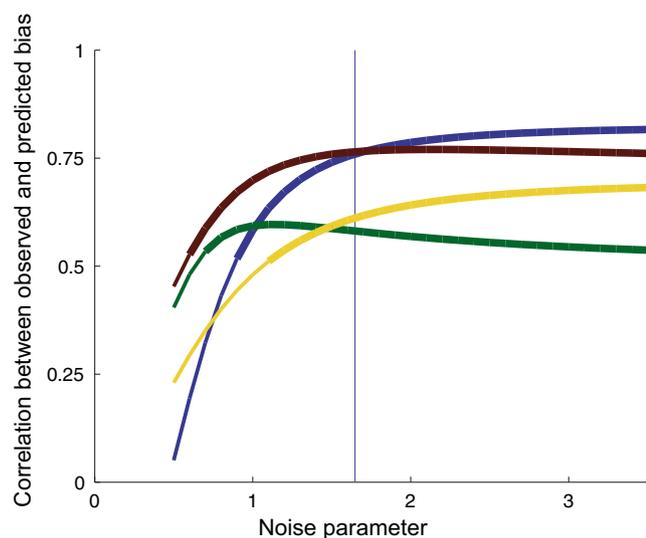
## Discussion

Here we measured the effect of memory on color matches by adding a 10 minute delay between seeing a color stimulus and selecting its match. We also examined how addition of a context change affected memory matches. We found that this delay elicited errors in average matches that are approximately twice the size of errors in the baseline condition, and that the variability among memory matches was also elevated compared to baseline. The pattern of errors was well accounted for by an ideal observer model that assumes memory is a noisy, unbiased version of perception. Perhaps surprisingly, we also found that changing the context between encoding and retrieval did not add significant bias or variability to the responses made in with a delay alone. Below, we discuss both results in more detail.

### Implications for color memory

Color matches made in memory exhibited significant biases and were significantly more variable than those made in the baseline condition (Fig. 5). Although the competing claims about memory fidelity made it difficult to form a clear hypothesis about memory bias, we evaluated the biases we measured in light of three of these hypotheses. First, we examined whether memory increased the saturation of color matches (Bartleson, 1960; de Fez et al., 2001; Newhall et al., 1957; Siple & Springer, 1983). Second, we tested whether memory matches were biased toward prototypical or focal hues (Bartleson, 1960; Collins, 1932). Third, we examined whether the fidelity of memory for a color was related to its proximity to a prototypical hue (Bartleson, 1960; Boynton et al., 1989; Uchikawa & Shinoda, 1996). In no condition did we find evidence that memory had any of these three effects (see Fig. 7 and reported ANOVA values).

The existence of biases related to color categories is somewhat controversial (Ling & Hurlbert, 2008), and our results are consistent with the notion that such biases are not a salient feature of memory across a large range of color space. It remains possible that these biases exist only for a subset of color stimuli. For example, previously reported category-related biases have been measured in reduced-cue conditions such as color patches on a monitor



**Fig. 11** Ideal observer predictions for the effect of adding a 10-min delay to the other four conditions: baseline (blue line, reported in Fig. 8), background (green), illumination (brown) and background + illumination (yellow). Y-axis values are the correlation between the observed bias in a condition and the bias predicted for that condition by the ideal observer model, as described in the text. X-axis values are variability parameter  $k$  as described in Fig. 8. Thicker lines indicate correlations with significance  $p < 0.05$ . Vertical blue line is the average  $k$  observed in the data for the memory-baseline comparison

(de Fez et al., 2001). The biases may be small enough that they do not survive the further cues to surface reflectance that arise in the complexity of real-world scenes. Alternatively, our study stimuli spanned a very large range of color space, but it did not do so finely. Several of our stimuli were very close in hue to prototypical colors: our red cube was  $1^\circ$  different from our average measure of prototypical red and our yellow cube was  $2^\circ$  different than our average measure of prototypical yellow. Perhaps stimuli with an  $8^\circ$  distance from a focal color would have exhibited a focal color bias. We also note that our measure of memory bias is relative to the baseline of perceptual matches. Thus, it is insensitive to category effects that arise in perception but are not exacerbated in memory.

A much better account of the magnitude of memory biases was found by modeling memory as a more variable but unbiased version of perception. The model contains two parts. The first contains an assumption about memory representation—that memory on each trial is a draw from a noisy multivariate Gaussian distribution. The second contains an assumption about the decision process, and this rests on our particular choice of palette. In the model, the ideal observer chooses the paint chip closest in chromaticity to the random draw. Thus, the discretization and gamut of the palette are integral to the model. This simple model was motivated by several empirical and theoretical observations.

First, there are strong empirical grounds for the central assumption of the model, that memory is more variable than perception. We reported such increased variability (Fig. 5). The factor of increase we observed ( $\sigma_{memory} = 1.6\sigma_{baseline}$ ) is similar in magnitude to the increase in thresholds found in other studies (1.5–2.5 in Nemes et al., (2010) and 1.6 in Olkkonen and Allred (2014)), although the two tasks and variability measures are quite different from each other. That memory is more variable (less precise) than perception is reported in traditional studies of color memory (Nemes et al., 2010; Olkkonen & Allred, 2014; Olkkonen et al., 2014), in studies of working memory using color as the stimulus of interest (Bae et al., 2014; Brady et al., 2013), and in other visual dimensions (Brady et al., 2011; Magnussen & Greenlee, 1999).

Second, linking variability of color matches to the average bias of color matches is a key feature of this model. This general idea of relating variability in a visual stimulus dimension to the appearance of that stimulus dimension has received considerable attention recently in both color (Hillis & Brainard, 2005; Olkkonen & Allred, 2014; Olkkonen et al., 2014) and other stimulus dimensions such as line length (Ashourian & Loewenstein, 2011; Crawford et al., 2006), time perception (Jazayeri &

Shadlen, 2010) and speed perception (Stocker & Simoncelli, 2006). In many such studies, biases are thought to arise because a noisy sensory representation is combined with prior beliefs about the stimulus collection. In these cases, biases are often toward a particular stimulus value, such as the average of the presented stimuli (Crawford, Huttenlocher, & Hedges, 2006; Olkkonen & Allred, 2014; Olkkonen et al., 2014). Here, we add to this body of literature by showing that even without an explicit prior, biases can arise through combining a variable memory representation with characteristics of the matching palette.

Given the ubiquity of evidence that memory representations are more *variable* than perceptual representations, and the contradictory nature of claims about *bias* in memory, the idea that increases in variability can masquerade as bias seems important to evaluate. We combined this assumption about memory representation with the specific features of our matching palette. Although the particular discretization or gamut of matching palette may vary widely from study to study, the general goal of applying palette constraints to understand the relationship between bias and variability is likely to remain important. As one example, consider that in most studies of color that span a hue circle, that circle is chosen to maximize the saturation of the stimuli. Because of this, many stimuli are likely to fall near the edge of the palette gamut, irrespective of whether that palette consists of real objects (such as Munsell papers, NCS papers, or OSA papers) or patches presented on a color monitor. Indeed, Speigle and Brainard (1996) noted that on light-emitting displays generally, many experimental stimuli fall near the edge of the monitor gamut. And any time the representational variability of color overlaps the palette gamut, color matches can exhibit biases, even if the palette is nominally perceptually uniform. Thus, both parts of our model are likely to generalize to other experimental designs, although the particulars of the palette may differ from experiment to experiment. The idea that increased variability in memory can explain apparent biases is a simple one, and we suggest that future studies should evaluate biases in light of some implementation of this idea before invoking more convoluted explanations.

It is conceptually simple to minimize the likelihood that variability could masquerade as bias. To do so, the matching palette should be uniform, finely sampled with respect to memory variability, and the variability of memory representation should not overlap the gamut of available matches. Practically, however, various other experimental desiderata conflict with this design concept. For example, to avoid the edge of the gamut, study stimuli must be relatively desaturated. In addition, many psychophysical techniques vary stimuli in only one dimension (e.g., hue, saturation,

or luminance). If the underlying memory representation varies in more than the one dimension of possible matches, similar distortions can occur. Thus, experimental design may prohibit a choice of stimuli and palette that would prevent variability from masquerading as bias. Fortunately, the ideal observer model we have described here is relatively simple to implement. Such implementation can provide a reference for the size of bias expected if only variability is increasing. Biases systematically different or systematically larger than these could then reasonably be attributed to something other than increased variability.

#### Relating memory and context changes

In one canonical task used to measure color constancy, a study stimulus is viewed under one illumination, and observers adjust the chromaticity of a test stimulus viewed under a second illumination until it appears to match the color of the study stimulus. Observers typically adjust their matches to compensate for the change in illuminant chromaticity. Matches may be made simultaneously or successively, with a brief delay between study and test. In realistic scenes, color constancy indices are between 60 % and 90 %; in other words, a change in illumination elicits a bias on the order of 10–40 % (Shevell & Kingdom, 2008). In real world tasks such as recalling an object at home to match the color of an object at the store, color matches across contexts are often made after much longer delays. Thus we sought to examine the effect of changing context on memory matches.

Adding a change in context to a 10-min delay did not significantly alter color matches. If color constancy in our experimental context were perfect, this would be a trivial explanation: in other words, if a change in illumination added no error *without* a 10-min delay, it might not be expected to add error *with* a 10-min delay. However, we previously reported that color constancy in this stimulus set without a delay was 88 %, on average (Allred & Olkkonen 2013). This degree of constancy is within the range typically reported in color constancy studies with realistic stimuli (Bäumel, 1994; Kraft & Brainard, 1999; Olkkonen et al., 2010).

It is surprising that a change in context did not add a significant degree of additional bias or variability to memory matches. To understand this result, consider that it is likely that the processes of working memory are a common source of error in *both* the illumination condition and the long-term memory condition. First, the illumination condition likely involved working memory. In our task specifically, observers looked back and forth between booths, and the time of a saccade is sufficient to demonstrate effects of working memory (Nilsson & Nelson, 1981).

Empirically, the permutations analysis showed a strong correlation between the discriminability of cube matches from baseline in the illumination condition and the discriminability of cubes in the memory condition, suggesting that stimulus-specific effects share a common source of error. In contrast, cube discriminability in the background condition was not correlated with either the illumination or memory condition.

More generally, we and others have noted that although color constancy is often referred to as a purely perceptual phenomenon, it is both practically and theoretically difficult to isolate the perceptual and memory contributions to matching surfaces between illuminants (Allen, Beilock, & Shevell, 2012; Allred & Flombaum, 2014; Jin & Shevell, 1996; Ling & Hurlbert, 2008; Olkkonen & Allred, 2014). Thus, it should not be surprising that the illumination and delay conditions likely share working memory as a common source of error. This common source of error may be responsible for the interaction observed in Fig. 9, where the overall error is less than expected by adding illumination error alone to the delay error alone.

Clearly, the effect of adding context changes to the delay was less than the additivity prediction (Fig. 9), but it is difficult to draw a firm conclusion about whether adding context to the delay had any noticeable effect. In aggregate, context changes appear to add no additional bias. However, evaluating this claim statistically is difficult: rejecting the research hypothesis that illumination adds significant bias does not permit one to accept the null hypothesis that adding the illumination had *no* effect on delay-only bias. Thus, the data do not admit a firm conclusion. On the one hand, if the null hypothesis were true, it might suggest that the common source of error, working memory, is the *only* source of error in the joint delay and illumination task. Consistent with this view, increasing the delay past more than the few seconds of working memory adds little error (Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Jin & Shevell, 1996; Nilsson & Nelson, 1981; Nemes et al., 2010). Additionally, although it seems counterintuitive, there is some evidence that color constancy may improve with a delay. Jin and Shevell (1996) reported that in some conditions, color constancy was higher after a 10-min delay than a 10-s delay. If this is the case, then the illumination-specific error present in the illumination only condition may become vanishingly small in the joint condition that combines an illumination shift and long term memory, leaving only the working memory error.

Still, a simpler alternative explanation for the failure to measure significant additional error in the joint condition may be that the between-subjects variability was large enough to mask a real effect, resulting in a Type 2 error. Future research should examine further the interaction of long-term memory and context changes.

## Conclusions

In conclusion, we have demonstrated that memory alone elicits errors in color matches that are best accounted for by a simple model that assumes memory is an unbiased by noisy version of perception. We suggest that this null model should be considered in explanations of color biases as a comparison to more complicated theoretical frameworks. Second, adding context changes to the 10-min delay does not significantly alter color matches. These empirical observations lead us to propose that working memory is a source of error common to both color constancy and long-term memory. Further research should examine this proposition more directly.

**Author note** The authors would like to acknowledge Jeremy Bell, Patrice McCarthy, and Jessica McLaren for their help in collecting data.

**Conflict of interests** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Funding** This research was funded by NSF-BCS 0954749 to Sarah Allred.

## References

- Allen, E. C., Beilock, S. L., & Shevell, S. K. (2012). Individual differences in simultaneous color constancy are related to working memory. *Journal of the Optical Society of America A*, 29(2), A52–A59.
- Allred, S. R., & Flombaum, J. I. (2014). Relating color working memory and color perception. *Trends in Cognitive Sciences*, 18(11), 562–565. doi:10.1016/j.tics.2014.06.002
- Allred, S. R., & Olkkonen, M. (2013). The effect of background and illumination on color identification of real, 3D objects. *Frontiers in Psychology*, 4, 821. doi:10.3389/fpsyg.2013.00821
- Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One*, 6(5), e19551.
- Bae, G. Y., Olkkonen, M., Allred, S., Wilson, C., & Flombaum, J. (2014). Stimulus specific variability in color working memory with delayed estimation. *Journal of Vision*, 14(4), 1–23.
- Bartleson, C. J. (1960). Memory colors of familiar objects. *Journal of the Optical Society of America*, 50, 73–77.
- Bäumli, K. H. (1994). Color appearance: effects of illuminant changes under different surface collections. *Journal of the Optical Society of America A*, 11(2), 531–542.
- Boynton, R. M., Fargo, L., Olson, C. X., & Smallman, H. S. (1989). Category effects in color memory. *Color Research & Application*, 14(5), 229–234.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11, 1–34.
- Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, 24(6), 981–990.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Brainard, D. H., & Radonjic, A. (2014). Color constancy. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 545–556). Cambridge, MA: MIT Press.
- Burnham, R. W., & Clark, J. R. (1955). A test of hue memory. *The Journal of Applied Psychology*, 39(3), 164–172.
- Collins, M. (1932). Some observations on immediate colour memory. *British Journal of Psychology*, 4, 344–352.
- Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2006). Within-category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin & Review*, 13(2), 245–250.
- de Fez, M. D., Capilla, P., Luque, M. J., Prez-Carpinell, J., & del Pozo, J. C. (2001). Asymmetric colour matching: Memory matching versus simultaneous matching. *Color Research & Application*, 26(6), 458–468.
- Duncker, K. (1939). The influence of past experience upon perceptual properties. *The American Journal of Psychology*, 52(2), 255–265.
- Foster, D. H. (2011). Color constancy. *Vision Research*, 51(7), 674–700.
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1229.
- Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*, 4(7), 563–572.
- Hamwi, V., & Landis, C. (1955). Memory for color. *The Journal of Psychology*, 39(183–194).
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, 9(11), 1367–8.
- Hansen, T., Walter, S., & Gegenfurtner, K. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision*, 7, 1–15.
- Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2009). Color constancy improves for real 3D objects. *Journal of Vision*, 9, 1–16.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1), 10–20.
- Hillis, J. M., & Brainard, D. H. (2005). Do common mechanisms of adaptation mediate color discrimination and appearance? Uniform backgrounds. *Journal of the Optical Society of America A*, 22(10), 2090–2106.
- Jazayeri, M., & Shadlen, M. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, 13(8), 1020–1026.
- Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A*, 13(10), 1981–1991.
- Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America*, 96(1), 307–312.
- Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, 25(6), 1215–1226.
- Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, 62(2–3), 81–92.
- Nemes, V. A., Parry, N. R. A., & McKeefry, D. J. (2010). A behavioural investigation of human visual short term memory for colour. *Ophthalmic & Physiological Optics*, 30(5), 594–601.
- Nemes, V. A., Parry, N. R. A., Whitaker, D., & McKeefry, D. J. (2012). The retention and disruption of color information in human short-term visual memory. *Journal of Vision*, 12(1), 26.1–26.14.
- Newhall, S., Burnham, R., & Clark, J. (1957). Comparison of successive with simultaneous color matching. *Journal of the Optical Society of America*, 47(I), 216–227.
- Nilsson, T. H., & Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 7(1), 141–150.

- Olivers, C. N. L., & Schreij, D. (2014). Visual memory performance for color depends on spatiotemporal context. *Attention, Perception & Psychophysics*.
- Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PLoS ONE*, *9*(1), e86488.
- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*(11), 1–15.
- Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, *10*, 1–22.
- Shevell, S. K., & Kingdom, F. A. A. (2008). Color in complex scenes. *Annual Review of Psychology*, *59*, 143–166.
- Siple, P., & Springer, R. D. M. (1983). Memory and preference for the colors of objects. *Perception & psychophysics*, *34*(4), 363–370.
- Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1458), 1329–1346.
- Speigle, J., & Brainard, D. (1996). Is Color Constancy Task Independent? In *Proceedings of the 4th IS&T/SID Color Imaging Conference* (pp. 167–172). Scottsdale: AZ.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585.
- Uchikawa, K., & Ikeda, M. (1981). Temporal deterioration of wavelength discrimination with successive comparison method. *Vision Research*.
- Uchikawa, K., Kuriki, I., & ToNE, Y. (1998). Measurement of color constancy by color memory matching. *Optical Review*, *5*(1), 59–63.
- Uchikawa, K., & Shinoda, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research & Application*, *21*(6), 430–439.