Preprint version. Final published version is:
Roshan, E. and Funt, B. "Computational color prediction versus least-dissimilar matching," J. Opt. Soc. Am. A 35, B292-B298 (March 2018)

Computational Color Prediction versus Least-Dissimilar Matching

EMITIS ROSHAN[1] AND BRIAN FUNT[1,*]

[1]*School of Computing Science, Simon Fraser University, Vancouver, BC, V5A 1S6, Canada*
**Corresponding author: funt@sfu.ca*

Received XX Month XXXX; revised XX Month, XXXX; accepted XX Month XXXX; posted XX Month XXXX (Doc. ID XXXXX); published XX Month XXXX

The performance of color prediction methods CIECAM02, KSM², Waypoint, Best Linear, MMV center, and relit color signal are compared in terms of how well they explain Logvinenko & Tokunaga's asymmetric color matching results ("Colour Constancy as Measured by Least Dissimilar Matching," *Seeing and Perceiving*, vol. 24, no. 5, pp. 407-452, 2011). In their experiment, 4 observers were asked to determine (3 repeats) for a given Munsell paper under a test illuminant which of 22 other Munsell papers was the least-dissimilar under a match illuminant. Their use of "least-dissimilar" as opposed to "matching" is an important aspect of their experiment. Their results raise several questions. Question 1: Are observers choosing the original Munsell paper under the match illuminant? If they are, then the average (over 12 matches) color signal (i.e., cone LMS or CIE XYZ) made under a given illuminant condition should correspond to that of the test paper's color signal under the match illuminant. Computation shows that the mean color signal of the matched papers is close to the color signal of the physically identical paper under the match illuminant. Question 2: Which color prediction method most closely predicts the observers' average least-dissimilar match? Question 3: Given the variability between observers, how do individual observers compare to the computational methods in predicting the average observer matches? A leave-one-observer-out comparison shows that individual observers, somewhat surprisingly, predict the average matches of the remaining observers better than any of the above color prediction methods.

OCSI codes: 330.0330, 330.1720

http://dx.doi.org/10.1364/AO.99.099999

1. INTRODUCTION

Logvinenko & Tokunaga [1] conducted an asymmetric color matching experiment in which observers view a Munsell paper under one light (the test illuminant) and then choose the least dissimilar matching paper from a set of 22 papers under a second light (the match illuminant). There were 4 observers and 3 repetitions each. The papers under both lights are all visible simultaneously. See Fig. 1 for a photograph of the setup. The papers are rearranged between trials. Note that these are real papers under real illuminants, not colored patches on a digital display nor colors obtained using hidden illuminants to simulate reflectance changes [2, 3]. The experiment involved 6 illuminants of approximately equal illuminance, green (G), blue (B), neutral (N), yellow (Y), red1 (R1) and red2 (R2), and all 30 possible pairs were used as test/match illuminant conditions. However, since the two red illuminants are very similar, in this paper we exclude one of them (R2). Considering only the non-identical pairs of 5 of the illuminants, there are respectively 5 and 4 possible illuminants as the test and match lights and so 20 illumination conditions. The illumination condition is specified by G2N or Y2B and so on throughout the paper. For instance, G2N means the test and match field are, respectively, illuminated by green and neutral.

The Logvinenko & Tokunaga (L&T henceforth) experiment differs from many other asymmetric color matching experiments in that subjects are not asked to make exact asymmetric matches, but rather to identify the colored paper that appears least-dissimilar. They argue that the classic asymmetric matching has a major shortcoming in that the observers who set a match report that color matches are not always perceptually identical. They point out that the light-color dimension of object color means that an exact asymmetric color match is impossible in principle. Hence, they ask their observers not to find an exact match but rather a least-dissimilar match [1, 4].

There are other types of color matching but each has its own shortcomings. In memory matching, the samples under different lights to be compared are shown successively, not at the same time. When there is a delay between successive views this necessarily involves memory [5]. Allowing time for the eyes to adapt to each illuminant, the observers need to keep the color information in mind but it is hard to remember it perfectly after a long delay. In Haploscopic matching, a sample under the first light is shown to the right eye. A copy of the same sample under a different light is shown simultaneously (or successively) to the left eye so that each eye becomes adapted to a different light. Haploscopic matching experiments assume that the two eyes are independent with respect to sensitivities and chromatic adaptation mechanisms, which may well be valid for the sensory mechanism but may not hold for cognitive mechanisms. One half of the field of view corresponding to each eye will be projected into the left brain and the other half will be projected into the right region and then the signals will be mixed in a way that is not yet fully understood [6].

During each trial of the L&T experiment, a laser pointer is used to indicate a test colored paper (a Munsell paper from the matte collection) from the left-hand panel and observers are asked to identify the least-dissimilar paper from the right-hand panel. As L&T point out, a perfect asymmetric match will usually be impossible due to metamer mismatching (i.e., the fact that two different reflectances may reflect metameric lights under one illuminant, but non-metameric lights under a second illuminant). Further analysis of the effect of metamer mismatching in the context of this experiment is provided by Logvinenko et al. [7].

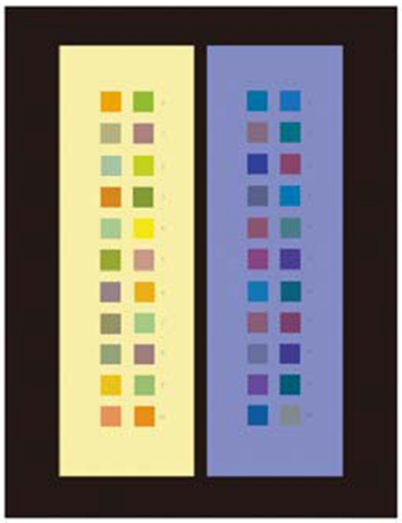

Fig. 1. The asymmetric matching setup used by L&T [1] showing the example of the left-hand panel in yellowish light and the right-hand panel in bluish light. "Each stimulus array contained 20 papers from every other page in the Munsell book of maximal Chroma … completed with grey (N5/) and black (N1/) papers (i.e., 22 papers in all). The stimulus array dimensions were 39 x 55 cm." [1] The papers are rearranged between trials.

Based on the L&T asymmetric matching results, we compare several color prediction methods to determine which best models observer performance. In particular, we compare von-Kries-rule-based CIECAM02 [8], KSM[2] [9], Wpt [10], Best Linear 3x3 transform [11], MMV (metamer mismatch volume) center [12] and Relit color signal (LMS cone response or XYZ) of the test paper under the match illuminant. Details of these methods are given below. In all cases, we assume that the methods have accurate information about the test and match illuminants.

In analyzing the methods relative to the L&T data, we address three questions: (i) Are observers generally choosing the physically identical Munsell paper under the match illuminant? (ii) Which computational method most closely corresponds to the observer average? and (iii) How does the performance of individual observers compare to the computational methods in predicting the least-dissimilar matches of the average observer?

2. BACKGROUND

Numerous methods for predicting 'color' under a change of illumination have been proposed. Derhak and Berns [10] make the distinction between chromatic adaptation transforms (CATs) and material adjustment transforms (MATs). A CAT is intended to predict what color signal under the match condition will appear the same as under the test condition. Of course, there is the issue of what 'the same' means. Derhak and Berns define the goal of a MAT as "…to predict material constancy or how sensor excitations for an object color change with changes in observing conditions" [10]. The problem with this definition is, as established by Logvinenko et al. [7], that as a result of metamer mismatching intrinsic object colors that are independent of the illuminant simply do not exist—hence material constancy in the Derhak and Berns sense does not exist either since from such material constancy intrinsic object colour would immediately follow.

Logvinenko [13] distinguishes between the intrinsic color of an object independent of the illumination—which he proves does not exist—and the material color that can be associated with object/light pairs. He bases the definition on the asymmetric-match relation determined by least-dissimilar matching. "Let us designate the least-dissimilar-colour-matching relation as ~m. Therefore, $(x_1, p_1) \sim m\ (x_2, p_2)$ means that the object x_1 under the light p_1 looks least dissimilar (amongst all the other objects under this light) to the object x_2 under the light p_2. … The existence of the least-dissimilar-colour-matching relation allows us to distinguish a particular type of object colour resemblance…When two object/light pairs belong to the same ~m equivalence class we will say that they have the same material colour." (p. 153-154)

As long as we bear in mind that we will not obtain constancy or "material color equivalency" [10] we can still investigate methods of predicting—given a color signal from a given surface reflectance under a first light—what its color signal is likely to be under a second light. Wpt [10] is one such color signal predictor. However, the issue we address here is not whether one CAT or color signal predictor is better than another, but rather whether or not any of them successfully predicts the least-dissimilar matches made by the observers in L&T's experiment.

Color signal predictors can be divided into two categories: those that require full knowledge of the spectral power distributions of both the test and match illuminants; and those that require only the color signals of the perfect reflector under each illuminant. In the first category are Relit, Best Linear [11], Wpt [10] and MMV center [12]. In the second category are von-Kries-based CIECAM02 [8] and KSM[2] [9].

The Relit color signal is simply the color signal of the given test paper under the match (second) illuminant. Computing it requires the full spectral reflectance function of the surface as well as the SPDs of the second illuminant. Since L&T used matte Munsell papers, we assume that the color signal $(\varphi_1, \varphi_2, \varphi_3)$ resulting from light impinging on sensors $S_i(\lambda)$ (i = 1…3) from a surface of spectral reflectance $x(\lambda)$ illuminated by light with spectral power distribution $I(\lambda)$ is:

$$\varphi_i(x) = \int_{\lambda_{max}}^{\lambda_{min}} x(\lambda) I(\lambda) S_i(\lambda) d\lambda \left(i = 1, 2, 3\right) \qquad \textbf{(1)}$$

The Relit 'prediction' of the color signal is, of course, not really a prediction at all but rather, under the assumption of matte reflectance, a straightforward calculation of what the actual color signal will be.

Wpt involves a 3x3 linear matrix transformation of the test color signal to the match color signal. The 3x3 transformation is determined based on the SPDs of the illuminants and a training set consisting of the reflectances of all the papers in the Munsell collection. In order to satisfy other design requirements, Wpt does not, in fact, determine the optimal 3x3 matrix. In comparison, the Best Linear method [11] is based on using the optimal 3x3 matrix mapping the color signals from the training set (1600 Munsell papers) under the test illuminant to the match illuminant.

MMV center prediction is based on computing metamer mismatch volumes. For a given color signal under the test illuminant, the set of color signals it could theoretically become under the match illuminant defines a convex volume in color signal space called the metamer mismatch volume (MMV). Computing the MMV requires full knowledge of the SPDs of both illuminants. Logvinenko et al. [12] propose using the color signal at the geometric center of the MMV as a candidate for what the color signal under the test illuminant is likely to become under the match illuminant, and we label that prediction method "MMV centre".

In the second category of color signal prediction methods—those that require only the color signals of the illuminants, not their full SPDs—we consider von-Kries-based CIECAM02 [8] and KSM^2 [9]. At the heart of CIECAM02 is the chromatic adaptation transform CAT02, which applies the standard von Kries (diagonal) transformation after a sharpening transformation [14, 15]. The degree of adaptation can vary from zero, for no adaptation, to 1, for complete adaptation. We tested CIECAM02 with 10 different values specified for D (0.1, 0.2, …, 0.9, 1) instead of computing it as a function of the adapting field factors. We found that CIECAM02 performed the best with D equal to 1. Therefore, we set D to 1 when computing the CIECAM02 prediction results reported below.

Also in the second category is KSM^2, developed by Mirzaei et al. [9]. KSM^2 uses Gaussian-like functions (called wraparound Gaussians) to represent both the illuminations and the reflectance. Given the color signal of a light (its full SPD is not required), a metameric Gaussian SPD can be found that is fully specified by 3 parameters: K the scaling, S the sigma, M the peak wavelength. As illustrated in Fig. 2, to make a color signal prediction, KSM^2 finds three Gaussian functions, one representing an SPD metameric to the test illuminant, a second metameric to the match illuminant, and a third representing a reflectance metameric to the given test color signal under the Gaussian SPD metameric to the test illuminant. It then computes the match color signal of that Gaussian reflectance under the match Gaussian illuminant and uses that color signal as its prediction.

L&T suggest that the least-dissimilar match may be based on the central wavelength component of Logvinenko's ADL coordinates [16]. In terms of ADL coordinates, Logvinenko proved for any arbitrary strictly positive illuminant that for each spectral reflectance function there exists a unique rectangular spectral reflectance function specified by three numbers, purity (α), spectral bandwidth (δ), and central wavelength (λ) that is a metamer under that illuminant. An example of an αδλ (ADL) metamer is shown in Figure 2(c). L&T suggest "It seems plausible to expect the same rectangular spectral reflectance function to be assigned the same material colour under different illuminations. If also the least dissimilar match is based on the equality of material colours then we can make a prediction for our stimulus papers evaluating the colour stimulus shift produced by the illuminants used in our experiment. The prediction is rather simple: the least dissimilarity between differently illuminated papers is to be achieved by the pair with the same rectangular metamers. As purity and spectral band did not vary systematically over the stimulus sample, this suggestion amounts, at first approximation, to the prediction that in our experiment the least dissimilar match should be determined by the central wavelength" (p. 429 [1]). In other words, the L component.

L&T test their hypothesis and conclude, "… the observers' matches drastically violate the central wavelength equality prediction" (p. 431 [1]). In any case, we test this central-wavelength hypothesis again here but using the M of KSM^2 [9] rather than the L of ADL.

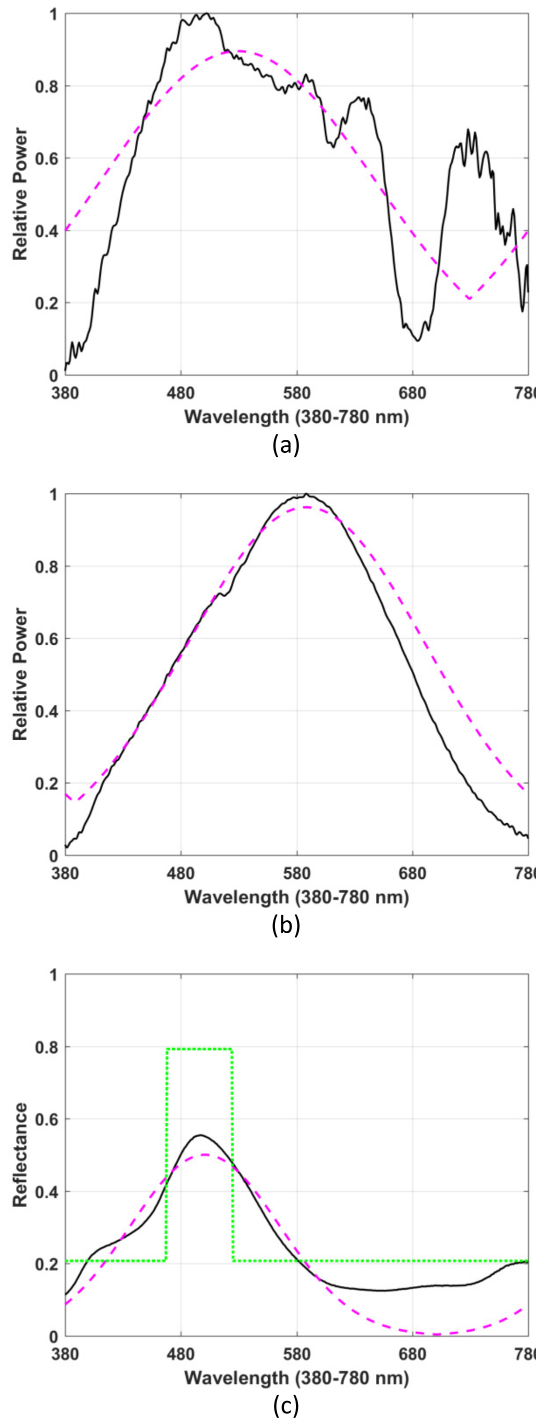

Figure 2. Solid black curves are the given spectral power distributions and reflectance. Dashed magenta curves are their Gaussian metamers and the dotted green curve is the ADL metamer with α=0.58, δ=54, λ=497. (a) Wraparound Gaussian SPD metameric to the test light. (b) Wraparound Gaussian SPD metameric to the match light. (c) Wraparound Gaussian reflectance producing the same color signal when lit by the wraparound Gaussian SPD from (a) as the color signal of the original reflectance (solid black curve in (c)) under the test light. The rectangular spectral reflectance is also metameric to the original reflectance under the test light.

3. OBSERVERS CHOOSE ORIGINAL MUNSELL PAPER?

Before addressing the issue of how well the various computational methods model the asymmetric matches made by the L&T observers, we consider the issue of whether or not observers are generally choosing the physically identical Munsell paper under the match illuminant as least-dissimilar to the test paper? To answer this question, for each test paper under the test illuminant, we compute the average XYZ under the match illuminant of the Munsell papers chosen as least-dissimilar and calculate how far in terms of Euclidean distance that average is from the actual XYZ of the test paper under the match illuminant, and finally average the results over all 20 test papers.

For each illumination condition, 4 observers with 3 repeats made least-dissimilar matches. All 20 chromatic papers were used as test papers. For each of the 20 test papers, therefore, there are 12 least-dissimilar matches reported, resulting in 240 matches for each illumination condition. Considering the 20 non-identical pairs of lights used in the asymmetric matching experiments, we have 20 x 240, or 4800 matches in total. The average Euclidean distance between the matched paper and the XYZ of the physically identical Munsell paper under the match illuminant is 6.0. For comparison, the average XYZ difference between a given Munsell paper and the nearest of the other 19 papers under the Neutral illumination is 6.6. In other words, the observers are on average choosing as least dissimilar a paper that is either the physically identical paper or one that is close to it in color.

Our analysis is in agreement with L&T's analysis: "… when the test illuminant was neutral or yellow the average mismatch was roughly one hue step. The mismatch for the other four test illuminants was approximately two hue steps. Therefore, while the exact match rate for these illuminations … is quite low (less than 30%) the average mismatch does not exceed two hue steps" ([1] p. 415). An 'exact match' is defined as the observer choosing the physically identical paper.

These results suggest, perhaps not surprisingly, that observers generally find the match paper that is physically identical to the test paper to be the least dissimilar one.

4. PREDICTING OBSERVER AVERAGE MATCHES

To determine which method most closely predicts observer least-dissimilar matching behavior, we consider the 12 (4 observers, 3 repeats) matches made for each test paper under a given illumination condition and compute the *average-observer-match* as the average of the color signals of the 12 matched papers under the match illuminant. Each computational method is used predict the color signal of the test paper under the match illuminant. A method's *prediction error* is calculated as the Euclidean distance between the average-observer-match color signal and the color signal the method predicts.

We compare the performance of the computational color prediction methods to one another using the Wilcoxon signed-rank one-sided and two-sided tests [17]. The Wilcoxon test is a non-parametric statistical hypothesis test based on the sum of the signed ranks of a set of paired samples. In the present case, the paired samples are the prediction errors for the 20 papers under a given illumination condition of the two methods being compared. All the tests are performed at the 5% significance level.

More specifically, the 20 test papers result in 20 average-observer-match values for a given pair of test and match illuminants, along with a corresponding set of 20 predictions made by each algorithm. Three tests are performed to compare each pair (Method 1 and Method 2) of methods—one two-sided test and two one-sided tests. The null hypotheses for these tests are as follows.

- Two-sided test: the null hypothesis is that the median prediction errors of the two methods are equal.
- Right-tailed test: the null hypothesis is that the median prediction error of Method 1 is greater than the median prediction error of Method 2.
- Left-tailed test: the null hypothesis is that the median prediction error of Method 2 is greater than the median prediction error of Method 1.

The results of the three Wilcoxon tests will lead to one of the following cases.

Case I: The null hypothesis of the two-side test cannot be rejected at the 5% significance level. In this case the performance of Method 1 and Method 2 can be considered to be equivalent.

Case II: The null hypothesis of the two-side test can be rejected and the right-tailed test cannot be rejected, but the null hypothesis of the left-tailed test can be rejected. In this case, Method 2 can be considered to be better (lower median prediction error) than Method 1.

Case III: The null hypothesis of the two-side test can be rejected and the left-tailed test cannot be rejected, but the null hypothesis of the right-tailed test can be rejected. In this case, Method 1 can be considered to be better (lower median prediction error) than Method 2.

Table 1. Comparison of algorithms in predicting the average-observer-match in each of the 20 different illumination conditions. The numbers in columns 3-5 indicate how many times across the 20 different illumination conditions that each Case (see text for definition of the Cases) occurs. Informally, Case III indicates Method 1 is 'better' than Method 2, Case II that Method 2 is better than Method 1, and Case I that they perform similarly.

Method 1	Method 2	Case III	Case II	Case I
Relit	KSM^2	10	1	9
Relit	MMV Center	20	0	0
Relit	CIECAM02	10	0	10
Relit	Wpt	8	0	12
Relit	Best Linear	5	0	15
KSM^2	MMV Center	17	0	3
KSM^2	CIECAM02	5	2	13
KSM^2	Wpt	5	5	10
KSM^2	Best Linear	1	6	13
MMV Center	CIECAM02	0	18	2
MMV Center	Wpt	0	19	1
MMV Center	Best Linear	0	20	0
CIECAM02	Wpt	0	6	14
CIECAM02	Best Linear	1	8	11
Best linear	Wpt	1	7	12

Note that the results in Table 1 show the relative performance of the methods, not their absolute performance. In other words, the methods might be doing equally poorly rather than equally well. In terms of absolute performance, Table 2 lists the accuracy of each method's predictions averaged over the 400 cases. The accuracy is measured in terms of the Euclidean distance between the prediction and the average XYZ of the 12 least-dissimilar matches, and similarly for CIE1976 u′v′ coordinates. Although most of the results reported in this study are in terms of XYZ, almost identical ranking results were obtained using Euclidean distances in Hunter-Pointer-Estevez LMS space and the CIEDE2000 metric.

The results in Table1 and Table 2 are aggregated over all 20 Munsell papers and all 20 illumination conditions. L&T [1] provide a detailed analysis of how the average 'exact match' rate varies both with the illumination condition and with the test paper.

Table 2. Accuracy in Predicting Average Observer Matches. Mean and median of the Euclidean distance in XYZ and CIE1976 u′v′ between each method's predictions and the average observer match across 400 cases.

Method	Mean XYZ	Median XYZ	Mean u′v′	Median u′v′
Relit	5.21	3.45	0.024	0.015
Best Linear	5.56	4.17	0.040	0.023
Wpt	6.20	4.44	0.096	0.025
KSM²	8.08	4.50	0.043	0.030
CIECAM02	7.61	5.99	0.040	0.030
MMV Center	39.85	23.44	0.072	0.040

5. OBSERVERS PREDICTING OTHER OBSERVERS

In the previous section the performance comparison is between computational methods. All those methods might be equally good or bad but how does their performance compare to that of the observers relative to one another? Clearly there will be variability in the least-dissimilar matches made by the different observers. To what extent do the observers agree with one another and is a match made by an individual observer any better or worse a predictor of the average observer match than those made by the various computational methods?

To answer this question, we used a leave-one-observer-out comparison in which one observer is excluded and the 9 remaining trials (3 observers, 3 repeats per paper) are combined to create a 3-observer average for each illumination condition. The mean of the excluded observer's 3 trials is then used as a predictor of this 3-observer average. This process is repeated for each of the 4 observers resulting in predictors Obs1,…,Obs4 of the 4 different, 3-observer averages.

Table 3 compares the individual observers to the computational methods in predicting the 3-observer average. Table 3 also includes results based on picking the paper that has the closest 'hue' using M from KSM² as the hue measure, which interestingly does slightly better than using all 3 components of KSM².

Table 3. Observers versus Computational Methods. Similar to the Table 1 but in this case comparing via the Wilcoxon test how well each method/observer predicts the 3-observer average of least-dissimilar matches. The numbers in columns 3-5 indicate how many times across the 20 different illumination conditions that each Case (see text for definition of the Cases) occurs. Informally, Case III indicates the given method is 'better' than the particular observer, Case II that the observer is better than the method, and Case I that they perform similarly.

Method	Observer	Case III	Case II	Case I
KSM²	Obs1	0	13	7
	Obs2	0	15	5
	Obs3	0	11	9
	Obs4	0	14	6
Relit	Obs1	0	9	11
	Obs2	0	11	9
	Obs3	2	6	12
	Obs4	0	9	11
Wpt	Obs1	0	13	7
	Obs2	0	14	6
	Obs3	0	9	11
	Obs4	0	12	8
CIECAM02	Obs1	0	14	6
	Obs2	0	16	4
	Obs3	0	13	7
	Obs4	0	18	2
Best Linear	Obs1	0	12	8
	Obs2	0	10	10
	Obs3	0	7	13
	Obs4	0	12	8
MMV Center	Obs1	0	20	0
	Obs2	0	20	0
	Obs3	0	19	1
	Obs4	0	20	0
M of KSM²	Obs1	0	11	9
	Obs2	0	14	6
	Obs3	0	9	11
	Obs4	0	13	7

From Table 3, it is clear that human observers predict the 3-observer average better than the computational methods do, as indicated by the fact that the numbers in the Case II column are substantially larger than those in the Case III column.

6. RESULTS USING THE PROCESS OF ELIMINATION

In a discussion concerning the results described in Section 5 above, John McCann [18] suggested that perhaps the observers were exploiting the fact that there were only 20 chromatic papers from which to choose and this might in some way be affecting the L&T matching results. In order to address that concern, in this section we provide the computational methods with this additional information to see if they are then able to predict the observers' least-dissimilar matches correctly.

Although the L&T observers were instructed simply to identify the least-dissimilar looking paper, the observers were aware that the same 20 papers were present under both the test and match illuminants so it is conceivable that they used that extra information to do an overall best fit of the least-dissimilar matches for of the 20 papers under the match illuminant to those under the test illuminant. Although we cannot know what observers were doing when they made their least-dissimilar matches, we can have the computational methods exploit that extra information.

Table 4 shows the results corresponding to those in Table 3 but when the algorithms minimize the overall dissimilarity across all 20 papers before deciding on the match for the given test paper.

It is clear from Table 4 that the extra information does improve the computational methods' predictions of the 3-observer average (Case I numbers are larger than those in Table 3); nonetheless, the individual observers still are statistically better roughly half the time (Case II) . In other words, even when the computational methods are modified to exploit a process-of-elimination type strategy they are still are not as good as the human observers in predicting the other observers' least-dissimilar matches.

7. DISCUSSION

The Logvinenko & Tokunaga [1] asymmetric matching experiment is interesting because it is based on least-dissimilar matching of real papers under real lights. The question the L&T experiment addresses differs from that of many corresponding color experiments, which tend to abstract color away from what its purpose might be. Given this different set of experimental data, we have evaluated several color signal prediction methods in terms of how well they correspond to observers' least-dissimilar matching. Note that, as mentioned above, Best Linear, Wpt, and MMV centers require the full spectra of the test and match illuminants, while KSM² and CIECAM02 require only their color signals. In other words, the former ones may or may not predict human performance, but they cannot possibly provide a computational model of any aspect of trichromatic color perception.

Our analysis shows that observers tend to find the physically identical test paper to be the least-dissimilar match paper. Since there is a forced choice of 1 paper out of 20, this does not mean, however, that observers would always consider that paper to be the least-dissimilar if there were an effectively infinite choice of papers. Note also that because of the possibility of metamer mismatching it is a mistake to interpret the physically identical paper under the match illuminant as the 'correct' answer. An observer is not wrong to find some other paper to be least dissimilar. If the test/match paper were to be replaced by one of different (but metameric under the test light) reflectance then the color signal under the match illuminant will be different from the original situation even though nothing in the test condition visibly changed.

Table 4. Results corresponding to those in Table 3 but allowing the algorithms to include minimizing the total dissimilarity across all 20 papers simultaneously.

Method	Observer	Case III	Case II	Case I
KSM²	obs1	0	9	11
	obs2	0	11	9
	obs3	2	7	11
	obs4	0	12	8
Relit	obs1	0	9	11
	obs2	0	11	9
	obs3	2	6	12
	obs4	0	9	11
Wpt	obs1	0	9	11
	obs2	0	11	9
	obs3	2	6	12
	obs4	0	10	10
CIECAM02	obs1	0	11	9
	obs2	0	12	8
	obs3	0	10	10
	obs4	0	15	5
Best Linear	obs1	0	10	10
	obs2	0	10	10
	obs3	2	6	12
	obs4	0	10	10
MMV Center	obs1	0	9	11
	obs2	0	11	9
	obs3	1	10	9
	obs4	0	14	6
M of KSM²	obs1	0	10	10
	obs2	0	13	7
	obs3	0	9	11
	obs4	0	9	11

Interestingly, none of the methods is as effective as each individual observer in predicting the 3-observer average of the other observers' matches. This implies that all the computational methods studied are not capturing some important aspect of the observers' least-dissimilar matching strategy. L&T [1] argue for the existence of both lighting and material dimensions of object color and propose the concept of an across-illuminant color map. Perhaps once their across-illuminant color map is fully specified it will provide a full model of the L&T asymmetric matching results. All we can say in the meantime, however, is that the computational models we tested do not explain those results adequately.

Funding Information.
Natural Sciences and Engineering Research Council of Canada.

Acknowledgment.
Portions of this work were presented at the Human Vision and Electronic Imaging Conference in 2017 as "Evaluation of Color Prediction Methods in Terms of Least Dissimilar Asymmetric Matching." We would like to thank the reviewers for their very constructive and important suggestions and A. Logvinenko and R. Tokunaga for providing their experimental data.

References

1. A. Logvinenko and R. Tokunaga, "Colour Constancy as Measured by Least Dissimilar Matching," Seeing and Perceiving **24**, no. 5, 407-452, (2011).
2. D. Brainard, W. Brunt and J. Speigle, "Color constancy in the nearly natural image. 1. Asymmetric matches," J. Opt. Soc. Am. A **14**, no. 9, 2091-2110, (1997).
3. V. de Almeida, P. Fiadeiro and S. Nascimento, "Color constancy by asymmetric color matching with real objects in three-dimensional scenes," Visual Neuroscience **21**, no. 03, 341-345, (2004).
4. A. Logvinenko and L.T. Maloney. "The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching," Perception & Psychophysics **68**, no. 1, 76-83, (2006).
5. J. Pérez-Carpinell, V.J. Camps and J.A. Diaz, "Color memory in protanomals and deuteranomals: matching time effect," Color Research & Application **28**, no. 2, 151-154, (2003).
6. A.K.R. Choudhury, *Principles of Colour and Appearance Measurement: Object Appearance, Colour Perception and Instrumental Measurement* (Elsevier, 2014).
7. A. Logvinenko, B. Funt, H. Mirzaei and R. Tokunaga, "Rethinking Colour Constancy," PLOS ONE **10**, no. 9, e0135029 (2015).
8. N. Moroney, M. D. Fairchild, R. W. Hunt, C. Li, M. Luo and T. Newman, "The CIECAM02 colour appearance model," in *Proceedings of Tenth IS&T Color Imaging Conference* (2002), no. 1, pp. 23-27.
9. H. Mirzaei and B. Funt, "Gaussian Illuminants and Reflectances for Colour Signal Prediction," in *Color and Imaging Conference, Society for Imaging Science and Technology* (2014), pp. 212-216.
10. M. Derhak and R. Berns, "Introducing Wpt (Waypoint): A color equivalency representation for defining a material adjustment transform," Color Research & Application **40**, no. 6, 535-549, (2014).
11. B. Funt, and J. Hao, "Non-diagonal Color Correction," in *Proceedings of IEEE Conference on Image Processing* (IEEE, 2003) vol. 1, pp. I-481.
12. A. Logvinenko, B. Funt and C. Godau, "Metamer Mismatching," IEEE Trans. Image Process. **23**, 34-43, (2014).
13. A. Logvinenko, "Object-colour manifold," J. Computer Vision **101**, no. 1, 143-160, (2013).
14. G. Finlayson, M. Drew and B. Funt, "Spectral Sharpening: Sensor Transformations for Improved Color Constancy," J. Opt. Soc. Am. A **11**, 1553-1563, (1994).
15. J. A. von Kries, "Chromatic Adaptation," *Sources of Colour Vision* (1970).
16. A. Logvinenko, "An Object-color Space," J. Vision **9**, no. 11, (2009).
17. S. Siegel, N. J. Castellan, *Nonparametric statistics for the behavioral sciences,* Second ed., McGraw-Hill, New York, (1988).
18. J. McCann, personal communication, Feb. 1, (2017).