Supplemental Data

Distinct Mechanisms Mediate Visual Detection and Identification

James M. Hillis and David H. Brainard

Supplemental Experimental Procedures

Eight paid volunteers who were unaware of the experimental hypotheses and the two authors participated in the experiments. Two participants did not complete observations for the paint condition, so there were only eight participants for this condition. All participants had normal or corrected to normal acuity and normal color vision as assessed by an Ishihara color-blindness test.

Stimuli were presented on a computer-controlled calibrated RGB monitor with 14-bit resolution per channel and a refresh rate of 75 Hz. Participants viewed the stimuli monocularly through a small square aperture. Their head position was stabilized with a chin rest placed 86 cm from the monitor.

Background patterns were simulated perspective projection images of checkerboards (see Figure 1). We generated the shadowed checkerboard by simulating a 60% decrease in the illumination of the shadowed region. The penumbra was created with a Gaussian filter. For the painted checkerboard, the luminance of the checks along the diagonal was decreased by 60%. Table S1 provides the chromatic and luminance properties of the checkerboard images.

Test spots were $2.1^\circ \times 0.8^\circ$ ellipses (perspective projections of circles) presented for 0.5 s. They were blurred at the edges and were ramped on and off over .147 s (11 frames) with a Gaussian profile (σ = .04 s). Thus, test intensity was static for .21 s.

Discrimination Experiment

Discrimination performance was measured in a two-interval forced-choice task. The intensity in one interval was set at I_p. We call this the pedestal intensity. The intensity in the other interval was the pedestal plus an increment, $I_p + \Delta I$. Participants indicated the interval they believed contained the increment. Trials alternated between the two possible locations in each (shadow/paint) condition. After initiating the experiment with a key press, a fixation mark appeared for 0.4 s in one of the two locations. The participant shifted gaze to this location, and two spots were presented sequentially: One test spot had the intensity of the pedestal, I_p, and the other had the intensity of the pedestal plus test, $I_p + \Delta I$. The order was random, and the participant's task was to indicate which interval has the test. After the participants indicated their response by pressing a key, a fixation mark appeared for 0.4 s at the other test location, and the test sequence was repeated at this location. The participant shifted gaze between the two test locations in this trial-by-trial fashion until the session was complete.

One pedestal intensity, I_p, and one context (either painted or shadowed checkerboard) were used in an experimental session. Test intensity, ΔI, was controlled by self-terminating, adaptable staircases [S1]. Four independent staircases were run: two for the left and two for the right target locations. For each target location, one three-down, one-up, and one two-down, one-up staircase ensured efficient sampling of test intensities near threshold performance. Each staircase terminated after 12 reversals. Sessions for each pedestal/context pair were repeated twice: For one of the sessions, the painted or shadowed diagonal was on the negative oblique; for the other, it was along the positive. The orientation of the shadow or paint did not affect performance, and data were pooled across sessions. Each session lasted between 10 and 15 min.

Matching Experiment

The effect of context on appearance was measured with an asymmetric matching task. Participants matched the appearance of sequentially presented test spots located as shown in Figure 1. Spatial and temporal parameters of the targets were essentially identical to those used in the discrimination experiment. The only differences were that (1) the intensity of the spots in the two target intervals at one location was the same and (2) the participant responded only after the targets were presented in both target locations (as opposed once after each target location in the discrimination experiment). The response was a button press that changed the intensity of the spot at the second location. The selected intensity change was then used in the subsequent spot presentation. The participant indicated when a perceptual match had been obtained.

Specific instructions given to the participant in the matching experiment were: "Make the brightness of the adjustable spot the same as the fixed spot, disregarding, as much as possible, other areas of the display. That is, make it look like the amount of light coming from the adjustable spot is the same as that coming from the fixed spot." We used these instructions to bias participants toward relying on a low-level percept.

After initiation of the experiment with a key press, a fixation mark appeared for 0.4 s in one of the two locations. The participant shifted gaze to this location, and a spot of fixed intensity was presented twice at this location. The spot was shown twice so that the stimulus temporal properties could be equated with the discrimination experiment. After the second presentation, a fixation mark appeared for 0.4 s at the other test location. The participant shifted gaze to this location, and an adjustable match stimulus was presented twice. The participant then pressed a key to increase or decrease the intensity of this match stimulus for the next presentation. Observers controlled the size as well as the direction of the intensity change: There were four available step sizes. The smallest step size was set to a value near detection threshold determined from pilot experiments, and the other three step sizes were 3, 20, and 50 times greater.

Presentation of the fixed test and adjustable match continued until the participant pressed a button indicating that a perceptual match had been obtained. The participant was then asked to rate, on a scale of zero (match was unsatisfactory) to three (match was perfect), the quality of the match. After rating the match, a fixation marker appeared at the fixed-test location and was followed by presentation of a test of different intensity. There were eight tests of different intensity presented in each session. Each session was repeated at least four times. There were four different types of sessions: fixed test inside the shadow/paint (i.e., in the right test location shown in the images in Figure 1) and fixed test outside the shadow/paint. Session order was randomized and counter balanced across participants.

Data Analysis

Data Aggregation

Data were pooled across participants. To pool the discrimination data, we normalized the psychometric data for each participant. The normalization consisted of division of the test intensities for each pedestal and each of the four test locations (two for the paint condition and two for the shadow condition) for each participant by an estimate of their individual detection threshold. Detection threshold was taken as the value obtained for $I_p = 0$ at the test location outside the shadow in the shadow condition (i.e., the test location on the left in the top image of Figure 1). This normalization eliminated differences in absolute sensitivity across participants. After normalizing, we rescaled the test values for all participants by the average normalizing factor to keep the units commensurate with physical values.

To estimate just-noticeable intensity differences (JNDs) for each pedestal and test location, we pooled the trial-by-trial normalized data across subjects and fit these with cumulative normal functions by using a maximum-likelihood criterion. The advantage of pooling the data on a trial-by-trial basis is that we can directly estimate maximum-likelihood parameters for the pooled data. Arriving at such an estimate from individual threshold estimates is more complicated: There is a varying degree of confidence associated with each threshold, and taking a simple average does not take this into account. Along with the two parameters of the cumulative normal, we included a parameter for key-press errors [S2]. This "lapse rate" was constrained to be less that 5%. Across all participants and pedestals, the average lapse rate, estimated from these fits, was 1.9%. JNDs shown in Figure 2 correspond to 75% correct points estimated

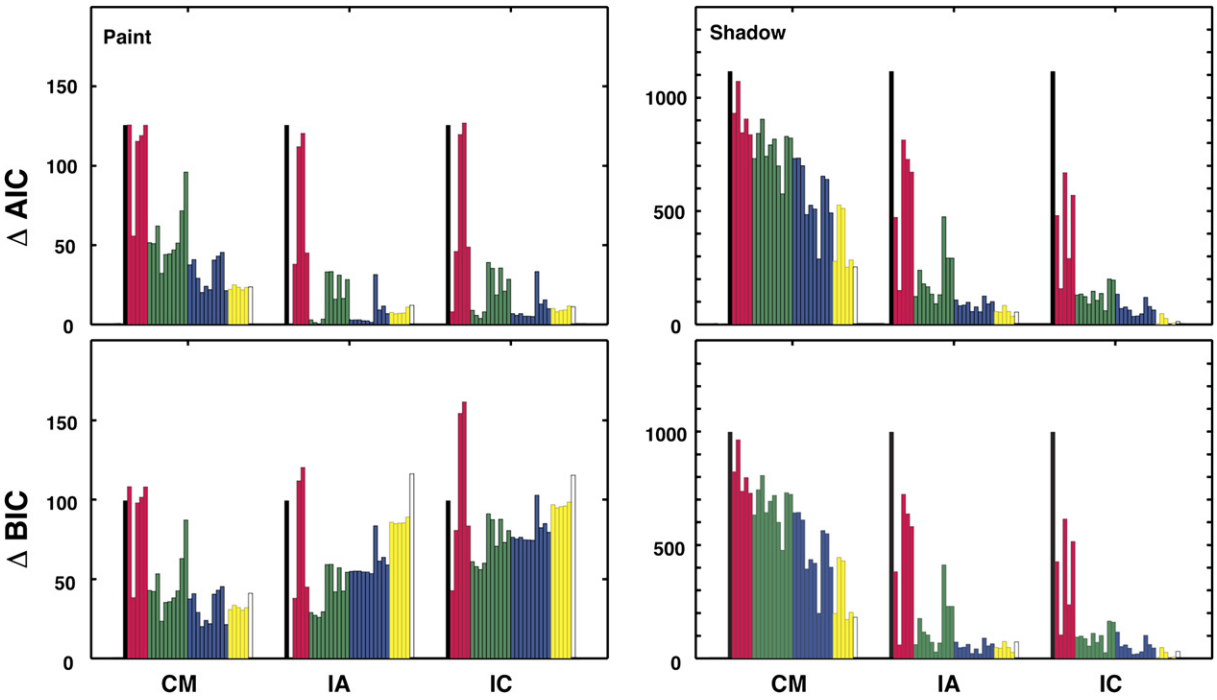

Figure S1. Δ*AIC* and Δ*BIC* Results for the Full Complement of Models

The two left panels are results from the paint condition, and the two right panels are results from the shadow condition. The three groups of bars within each plot correspond to the three variants (labeled on the x axis). The six colors within each group correspond to the no-adaptation model (black), one (red)-, two- (green), three- (blue), four- (yellow), and five- (white) parameter adaptation models. Note the large difference in y axis scale for the shadow and paint conditions.

from the fits. The error bars were 95% confidence intervals determined by boot-strapping data from the maximum-likelihood parameters [S2].

The matching data were averaged across all participants. Trials in which the participant could not get a satisfactory match (i.e., where their rating for the setting was zero) were excluded from the analysis. A total of 22 trials (~2%) were excluded for this reason.

Model-Fitting Methods

We used numerical search to find model parameters, which we refer to generically as θ, that maximized the likelihood function: $L(\theta|data_{match}, data_{discrim})$. The parameters are the five parameters in the main text's Equation 1, which is restated here for convenience:

$$R = M\frac{(gI+s)^p}{(gI+s)^q+1} \quad (1)$$

To estimate the match-setting variance for each test intensity, we used the variance of the residuals between the individual matches and their corresponding means. We allowed for the fact that participants' settings were more variable at some test intensities than others. Matches to the low and high test intensities tended to be more variable than those for the midrange test intensities. By using the Gaussian noise assumption and variance estimated from the residuals, we could calculate the likelihood of observing the individual match settings.

We fit the three variants (common mechanism, independent adaptation, and independent channels) of all 31 parametric adaptation models described in the main text, plus a no-adaptation model, to data from the paint and shadow conditions separately. Thus, we fit a total of 93 models to the data from each condition (there are 93 models including the zero parameter model rather than 94 because the five parameter adaptation model is the same for the IA and IC variants). Each model was fit with numerical search over the models' parameters so that the likelihood of the data given the parameters could be maximized. The 93 models differed in terms of the number of free parameters. For clarification of the differences, it is helpful to consider a concrete example. Suppose we choose the two-parameter model of adaptation in which parameters *g* and *p* of Equation 1 are allowed to vary across contexts. This adaptation model limits the change in the shape of the intensity-response function across the two test locations to that which can be produced by changes in the values of *g* and *p*. Values of the other three parameters, *s*, *q*, and *M*, are yoked to be the same at two test locations. For the CM variant of this model, *g* and *q* are allowed to differ as a function of test location but are the same for appearance and discrimination judgments (two adaptation parameters × two test locations + three fixed parameters = seven parameters in all). For the IA variant, *g* and *q* are allowed to differ as a function of test location and form of judgment (appearance or discrimination) but *s*, *q*, and *M* are the same at each test location and each form of judgment (two adaptation parameters × two test locations × two types of judgment + three fixed parameters = 11 parameters). For the IC variant, *g* and *q* are allowed to differ as a function of test location and form of judgment (appearance or discrimination), whereas *s*, *q*, and *M* can vary with form of judgment but not with test location (two adaptation parameters × two test locations × two types of judgment + three fixed parameters × two types of judgment = 14 parameters). Hillis and Brainard [S3] provide further details on the modeling approach.

To facilitate the process of finding global likelihood maxima, we fit the simplest models first (i.e., the no-adaptation model, which has five free parameters, was fit first) and used parameter values from these fits as starting points for higher-dimensional searches. For higher-dimensional searches, we used starting points determined by fits from nested variants (e.g., we used the best fit from CM variant as starting points for the IA variant) and nested adaptation models (e.g., we used the fit from single-parameter adaptation models as starting points for two-parameter adaptation models, etc.). With few exceptions, both of these approaches led to the same numerical solution. Thus, there is some assurance that our fits were determined by the global maxima of the likelihood functions. For the few exceptions that did occur, we chose the parameters that provided higher likelihoods.

To select among the 93 possible models for the paint and shadow conditions, we used the likelihoods obtained from the maximum

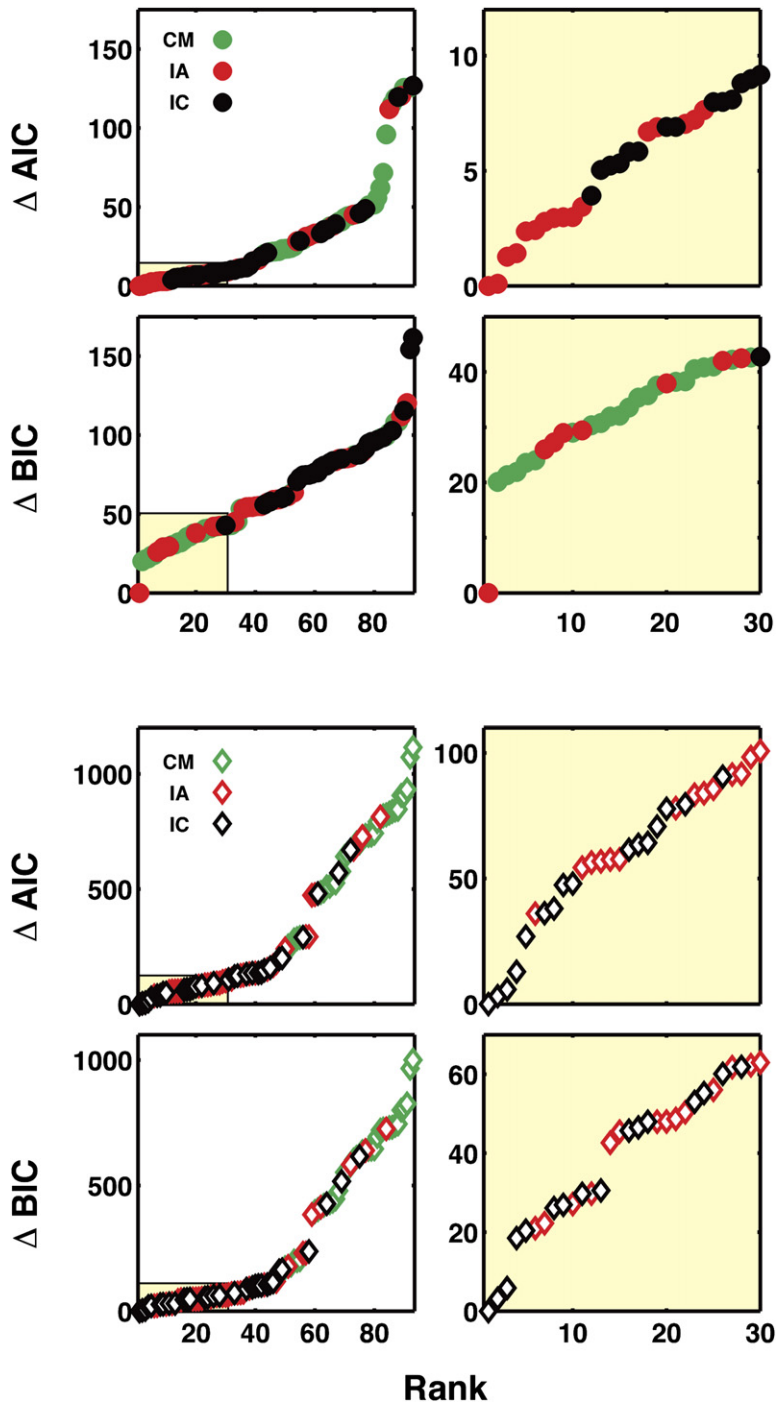

Figure S2. Δ*AIC* or Δ*BIC* as a Function of Model Rank

The paint and shadow conditions are shown in the top four and bottom four panels, respectively. For providing greater detail for the result of the top-ranked models, the right panels show results for the top 30 models (the yellow rectangles in the left panels indicate the regions that are shown in the right panels). Different symbol colors represent the three model variants. Note the large difference in the y axis scale for the paint and shadow conditions.

likelihood fits to calculate information theoretic criteria [S4–S6] and employed cross-validation techniques.

Model-Fitting Results

We computed an information criterion, AIC [S4], and the Bayesian information criterion, BIC [S6], by using the following equations:

$$AIC = -2\ln\left(L\left(\hat{\theta}|data_{match}, data_{discrim}\right)\right) + 2K$$

$$BIC = -2\ln\left(L\left(\hat{\theta}|data_{match}, data_{discrim}\right)\right) + K \cdot \ln(n)$$

in which L is the likelihood, $\hat{\theta}$ are estimated maximum-likelihood parameters, K is the number of free parameters, and n are the number of data points (which we counted as the number of matches plus the number of forced-choice judgments). Figure S1 shows the results. The bar plots in the top four panels show Δ*AIC* and Δ*BIC* (Δ*AIC* = *AIC* − min(*AIC*)and Δ*BIC* = *BIC* − min(*BIC*)) for the full complement of models considered. The *AIC* and *BIC* values are estimates of the expected relative distance between the model and the true model. Thus, smaller values of Δ*AIC* and Δ*BIC* indicate better models. Within the set of models considered, the most preferred model by the *AIC* criterion has a Δ*AIC* value of zero, and the most preferred model by the *BIC* criterion has a Δ*BIC* value of zero. The three groups of bars within each plot correspond to the three variants (labeled on the x axis). The six colors within each group correspond to the no-adaptation model and the one- to five-parameter adaptation models (note that the no-adaptation model is the same for all three model variants and that the five parameter model is

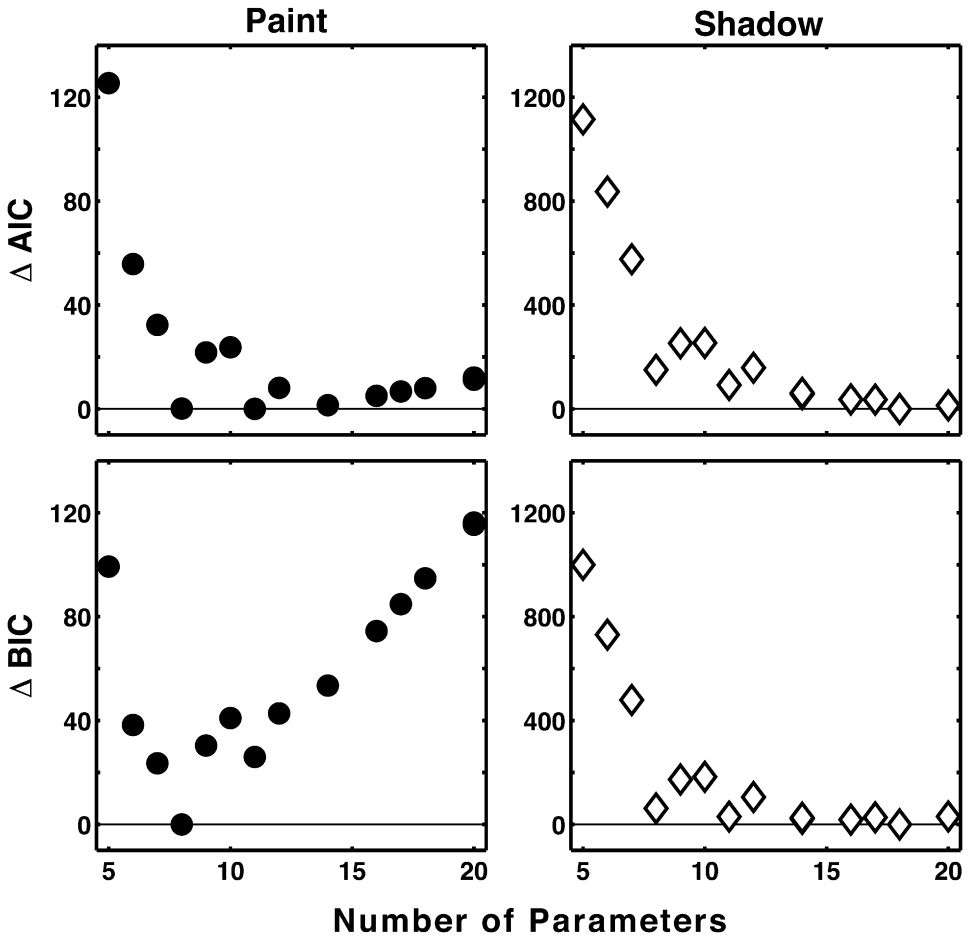

Figure S3. Δ*AIC* and Δ*BIC* of the "Best" Models as a Function of Number of Parameters

"Best" here refers to the model with the smallest Δ*AIC* or Δ*BIC* value within the set of models that had n parameters. As shown in Table S2, there is some overlap of the number of model parameters across the three variants considered (e.g., a one-parameter IA model and three three-parameter CM models each have the same number of parameters: eight). These plots are intended to emphasize the fact that the Δ*AIC* and Δ*BIC* minima for the paint condition occur at a fewer number of model parameters (eight) than the shadow condition (eighteen). The two left and two rights panels show results for the paint and shadow conditions, respectively. The fits associated with each of these models can be seen at http://color.psych.upenn.edu/supplements/shadowpaint/mainwebshdpnt.html.

the same for the IA and IC variants). The two left panels are results from the paint condition, and two right panels are results from the shadow condition. The two top and two bottom panels show *AIC* and *BIC* results, respectively.

The first result obtained from this model-selection analysis is the rejection of the CM variants in favor of the IA and IC variants. The *AIC* results clearly favor the IA and IC variants over the CM variant, for both paint and shadow conditions. This is indicated in the bar plots by the fact that the bars in the IA and IC groups generally have smaller values than the bars for the CM group. The *BIC* also favors the IA and IC variants for the shadow condition. For the paint condition, the *BIC* analysis is less clear cut by visual inspection of this bar plot. Figure S2 replots the data from Figure S1 in a manner that emphasizes the comparison across variants. The x axis of these plots is the overall model rank. The y axis is the Δ*AIC* or Δ*BIC* value. The top four panels are results from the paint condition, and bottom four panels are results from the shadow condition. The two right panels within each block are "zoomed-in" versions of the left panels so that greater detail on the results for the models ranked in the top 30 could be provided. Different symbols represent the three variants. We see that the best models tend to be the IA and IC variants: Red and black symbols are clustered toward the bottom left of each plot. The only exception is the Δ*BIC* values for the paint condition. In this case, there are several three-parameter CM models that are ranked in the top ten. However, there is a clear break between the top-ranked IA model (which was the model that allowed only *g* to vary across context) and the top-ranked CM models. Overall, CM variants are soundly rejected in favor of the more complex IA and IC variants for both the paint or shadow conditions.

A second feature of the analysis is that the CM variants are rejected more decisively for the shadow condition than for the paint condition. In this regard, note that the range of Δ*AIC* and Δ*BIC* values in Figures S1 and S2 is much greater for the shadow than for the paint condition. Related to this is the fact that both criteria indicate that more complex models of adaptation are required for the shadow than for the paint condition. To highlight this, we grouped the 93 models by their overall number of free parameters and then identified the best model within each group. Figure S3 plots the Δ*AIC* or Δ*BIC* values of the best model within each group as a function of number of parameters. For the paint condition, the minima occur at eight parameters for both Δ*AIC* and Δ*BIC*. The minima for the shadow condition occur at 18 parameters. The fits associated with each of these models can be seen at http://color.psych.upenn.edu/supplements/shadowpaint/mainwebshdpnt.html.

Table S2 provides Δ*AIC* and Δ*BIC* values for the no adaptation model and the best one- to five-parameter models for each of the variants. For both *AIC* and *BIC*, the top model for the shadow condition was a four-parameter, IC model. There was essentially no support for any simpler model. Burham and Anderson [S5] indicate that when there are fewer than 100 models under consideration, Δ*AIC* > 10 provides essentially no support for the model. By this criterion, only the top three models received nonnegligible support, and these

Table S1. Color Coordinates of Checkerboard Stimuli

	Black in Shadow	Black in Light/White in Shadow	White in Light	Background
L	0.013734	0.023089	0.038135	0.030512
M	0.011268	0.018972	0.031297	0.025038
S	0.0083938	0.014183	0.023343	0.018666
X	.33	.33	.33	.33
Y	.33	.33	.33	.33
Y	9.2	15.1	24.9	20
Isomerization Rates				
R^*_L	2212	3912	6645	5260
R^*_M	1566	2790	4748	3754
R^*_S	247	446	763	601
Isomerization Totals for Test Duration and Size				
R^*_L	138.09×10^5	244.2×10^5	$414.9 \times *10^5$	328.4×10^5
R^*_M	48.89×10^5	87.1×10^5	148.2×10^5	117.2×10^5
R^*_S	0.71×10^5	1.3×10^5	2.2×10^5	1.74×10^5

are all four-parameter IC models. By *AIC*, the most preferred models in the paint condition were one-, two-, and three-parameter models of the IA variant. By the *BIC*, the IA variant of the one-parameter *g* model was most preferred by a considerable margin.

Supplemental References

S1. Treutwein, B. (1995). Adaptive psychophysical procedures. Vision Res. *35*, 2503–2522.

S2. Wichmann, F.W., and Hill, N.J. (2001). The psychometric function: II.Bootstrap-based confidence intervals and sampling. Percept. Psychophys. *63*, 1314–1329.

S3. Hillis, J.M., and Brainard, D.H. (2007). Do common mechanisms of adaptation mediate color discrimination and appearance? Contrast adaptation. J. Opt. Soc. Am. A *24*, 2122–2133.

S4. Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Automat. Contr. *19*, 716–723.

S5. Burnham, K.P., and Anderson, D.R. (2002). Model Selection and Multi-model Inference (New York: Springer-Verlag).

S6. Schwartz, G. (1978). Estimating the dimension of a model. Annals of Statistics *6*, 461–464.

Table S2. Information-Criteria Results

			Paint		Shadow	
Variant	Adaptation Model	Number of Parameters	Δ*AIC*	Δ*BIC*	Δ*AIC*	Δ*BIC*
ALL	*None*	5	125.44	99.27	1114.86	997.52
CM	*s*	6	55.78	38.30	836.72	728.41
CM	*g*, *M*	7	32.33	23.55	576.06	476.78
CM	*g*, *p*, *q*	8	20.19	20.09	289.12	198.87
CM	*g*, *p*, *q*, *M*	9	21.75	30.34	252.47	171.24
CM	*g*, *s*, *p*, *q*, *M*	10	23.69	40.97	253.98	181.78
IA	*g*	8	0.10	0.00	150.16	59.90
IA	*g*, *s*	11	0.00	25.98	91.63	28.45
IA	*g*, *q*, *M*	14	1.42	53.46	56.34	20.24
IA	*g*, *s*, *p*, *M*	17	6.70	84.82	35.97	26.95
IC	*g*	12	8.10	42.76	158.16	104.01
IC	*g*, *q*	14	3.92	55.97	61.47	25.37
IC	*g*, *q*, *M*	16	5.04	74.46	36.18	18.13
IC	*g*, *s*, *p*, *M*	18	7.98	94.79	0.00	0.00
IA/IC	*g*, *s*, *p*, *q*, *M*	20	11.19	115.38	12.87	30.92