[Note for Doug: It is quite easy for me to regenerate the figures with different choices of point sizes, line widths, scale, label font sizes, etc.]

**Psychophysics Methods**

In the data we are reporting in the poster, there were only two context images, what we have been calling ‘original paint’ and one what we have been calling ‘original shadow’. The paint image served as the reference context, while the shadow image served as the comparison context. There was also a control condition where both context images were the paint image. Which side of the display each image appeared no was randomized across trials.

The two context images are illustrated in Figure 1 (Figure1.pptx), with the disks of the same luminance embedded in each: disk luminance 0.25 (left pair), 0.5 (middle pair), and 0.75) right pair. The top row shows ‘original paint’ images and the bottom row shows ‘original shadow’ images. The image pairs are patterned on Adelson’s checker-shadow illusion, combined with Gilchrist’s modification where the illusion occurs for disks placed on the checks. We have adopted the stimuli such that the lightness effect occurs across reference disks viewed in the center of two separate images, rather than within single images. In our version of the illusion, the reference disks always have the same immediate surround (the center check is the same luminance in the paint and shadow versions and the same average global surround. Indeed, the average luminance of each corresponding check in the paint and shadow versions is the same. The only difference between shadow and paint images is the spatial distribution of light in the first and second off diagonals. In the paint image, each check is spatially uniform. In the shadow image, the luminance is governed by a cumulative normal, producing a penumbra-like gradient. The feature of studying lightness effects across the paint-shadow context images is that they cannot be mediated by local contrast effects, nor by adaptation to the overall luminance of the context images. These two factors are matched. The mean luminance of each image was 0.485, and the luminance of the immediate surround of the disks had a luminance of 0.170.

On each trial, there was a test disk and a match disk, one in each context On half the trials, the test disk was in the reference context, while in the other half it was in the comparison context. The two context images, with disks embedded, appeared simultaneously to the left and right of fixation for 1 second. The subject’s task was to indicate which image (left or right) contained the lighter disk. The subject responded by pressing a button on the left or right of a game controller. There was in inter-trial interval of at least one second between the offset of one pair of images and the onset of the next pair. This interval could be longer if the subject took more than 1 second to respond.

There were three test disk luminances, with luminances on a normalized [0-1] scale of 0.25, 0.5, and 0.75. The match disk’s luminance was adjusted over trials by a staircase.

A luminance of 1 corresponded to about 285 cd/m2 – the exact value varied across the experiment as the monitor aged and across a change in monitors part way through the experiments. The CIE x,y chromaticity of the stimuli was approximately [0.30, 0.32].

The size of the context images was 3.5° of visual angle, and they were presented centered vertically and with their centers located at +/- 3.5° horizontally. Each check in the checkerboards was a 0.7° square. The disks had a diameter of 0.35°, and were centered on the center square of the context images.

There were two separate staircases for each choice of reference context image and test disk luminance, so 12 staircases in each session (2 choices of reference context image x 3 test luminances x 2 staircases for each combination). For each combination of context image and test disk luminance, one staircase was 2 up 1 down and the other was 1 up 2 down. (The reason for this is that the two staircase together spread the trials out a bit around the point of subjective equality, and do a better job of determining not only the point of subjective equality, that is the 50% point of the fit psychometric function, but also of the slope of the psychometric function through that point - aka discrimination threshold. In the back of my mind, I thought it might be interesting some day to look at how thresholds differ across the various paint shadow conditions, but to date have not done so. )

At the start of each session there were 5 practice trials, chosen randomly from the set of possible trial types and with the comparison luminance drawn randomly from the luminance range [0,1].

Each staircase started at a randomly chosen value from the luminance range [0,1], quantized to 20 levels (0.05 steps). The initial staircase stepsize was 0.20. After two reversals, this decreased to 0.10. After the third reversal, the step size decreased to 0.05 and remained there for the remaining trials of the staircase.

[Note for DB. If I had this to do again, I’d shrink the smallest step size. I don’t think this is a matter of great concern, but we probably can measure a bit more precisely than we are, at least for the lowest probe luminance. Go back and look at the early method of constant stimuli data and check how steep the psychophysical functions are in the data, as opposed to the fits to aggregated data. Could add a supplemental figure showing staircase progression for each probe intensity, or include these in an online site that we set up.]

Each individual staircase was 20 trials – that is there were a fixed number of trials per staircase. Thus there were 285 trials per session.

For each combination of context image and test disk luminance, we combined the data from the two staircases and fit a cumulative normal to it, using a maximum likelihood fitting method. The fit was implemented using the Palamedes toolbox and allowed for guess and lapse limits that could range between 0 and 5%. The PSE for each condition was obtained from this fit.

**Psychophysics Results**

Four subjects completed the original paint/shadow conditions, along with the control paint/paint condition. All were naïve as to the purpose of the experiment and had visual acuity of 20/40 or better as tested with a Snellen eye chart. Subjects ran two sessions of each condition (paint/shadow and control paint/paint) to complete a single replication. Subjects AQR and CNJ replicated the experiment twice; subjects BAF and EJE replicated it once. The conditions reported here are a subset of a larger set of conditions run by each subject.

[Note for DB. ParametricConditions2, conditions 31 (control) and 32 (paint/shadow) for AQR, BAF, and CNJ. ParametricConditions2, conditions 51 (control) and 52 (paint/shadow) for AQR, CNJ, and EJE.]

A example psychometric function for one subject/condition is shown in Figure 2 (Figure2.eps). The reference context was the paint image, the comparison context was the shadow image. The test was embedded in the reference image and match was in the comparison image. The test disk luminance was 0.5, as indicated by the vertical green line above the x-axis. The fraction of times that the match disks were judged lighter is plotted as a function of match disk luminance. For purposes of the plot, match luminances generated by the staircase procedure were sorted and aggregated into groups of five trials. The plotted points represent average fraction match lighter versus average match luminance for the five-trial groups. The red line through the data is the maximum likelihood cumulative normal fit to the individual trial data. The PSE obtained from the fit is plotted above the x-axis as the vertical red line. The fact that the red line is shifted to the left of the green line shows the perceptual effect of context on lightness. Disks appear lighter in the shadow image than the paint image (see Figure 1), so the luminance of the PSE less than that of the test disk.

Supplemental Figure 1 shows an example psychometric function for the same subject for a control condition where both reference and comparison context images were the paint image. This figure shows the expected result that the PSE is essentially identical to the reference disk luminance.

[Note for DB: Source for Figure 2 in data analysis tree: cnj-c32\_pnt\_rot0\_shad4\_blk40\_cen40\_vs\_shd\_rot0\_shad4\_blk40\_cen40\_t1-2\_example.eps. Source for Sup Figure 1: cnj-c31\_pnt\_rot0\_shad4\_blk40\_cen40\_vs\_pnt\_rot0\_shad4\_blk40\_cen40\_t1-2\_example.eps.]

[Presentation decision: I picked fairly clean examples of psychometric functions.]

Figure 3 (Figure3.eps) summarizes the paint/shadow data for one subject for one two-session experiment. Each point represents a pair of disk luminances that match in appearance. The red points plot the data for the first session, the blue point for the second session. For or trials where the test was in the reference context, the PSE is plotted against test intensity. For these trials, the PSE is generally less than that of the test, as in the case shown in Figure 2. For trials where the test disk was in the comparison context, the test disk luminance is plotted against the PSE. This reversal keeps the sign of the effect shown in the figure consistent across the two types of trials. For comparison, Supplemental Figure 2 shows for the same subject/experiment the data for the control condition. Supplemental Figure 3 shows plots in the same format as Figure 3 for each experiment/replication for the paint shadow condition.

[Note for DB: Source for Figure 3: cnj condition 52 data in ParametricConditions3. Source for Sup Figure 2: cnj condition 51 data in ParametricConditions3.]

[Presentation decision: I picked a fairly examples of single experiment data. You can look at all of the paint/shadow data in Supplemental Figure 3.]

We summarized the size of the effect of context on lightness by the downward shift of the data shown in Figure 3, relative to the positive diagonal. For these data, that shift is -0.06 normalized luminance units. The shift was determined by the best least squares fit of a line of unity slope to the data, including only the data in the range between 0.25 and 0.75 on the x-axis. This restriction was implemented to facilitate comparison with the physiological data. Points whose values on the y-axis fell outside the range [0-1] were also excluded. Such points can occur if the measured psychophysical function did not span the value 0.5, so that the PSE is determined by extrapolation of a poorly constrained PSE. The fit is shown by the solid black line. The corresponding shift for the control data shown in Supplemental Figure 2 is 0.01.

[Data analysis decision: the range restriction is to match the range we use in the analysis of the physiological data, where the separation between paint and shadow is most evident in the middle test range.]

[Data analysis decision: to fit with slope only, intercept only, or both? The summary plots with both suggest to me that we can’t really identify both a slope and intercept, which makes me want to fit with just one. My recollection is that some of the data sets from the decoding really want an intercept, so currently am fitting the psychophysics that way. The intercept is equivalent to the mean shift between test and PSE. You can look at Supplemental Figure 3 to see the fits. I don’t think the data massively contradicts the choice to fit with an intercept, although I suspect that if you looked really hard at the pattern of the residuals you’d reject the intercept only fit in favor of slope and intercept.]

[Note for Doug: There are cryptic titles on Figure 3 and Sup Figure 2, that you would want to delete from the version that goes on the poster.]

Figure 4 shows the intercepts obtained for each subject and replication for the paint shadow condition (red points) and for the control condition (black points). The mean intercept for the paint shadow condition (obtained by first averaging the data within subject across replications and then averaging across subjects) was -0.06. The corresponding mean for the control condition was more than a factor of 10 smaller,  
-0.004. The paint shadow intercepts for AQR, BAF, CNJ, and EJE respectively were  
-0.08, -0.04, -0.06, and -0.07, with the corresponding control values being -0.01, -0.009, -0.008, and 0.01. The paint shadow effect can clearly be reliably measured with our procedures.