Main Study Findings

Tables 3.31 through 3.38 show the average, median, and standard deviations for the overlap estimate and intersection measures.

The same analyses that are performed for the pilot study are performed for the main study data. The first analysis answers two questions 1) how performance for overlaid DDS images with distractors compares to that for the *Side-by-side* view images without distractors and 2) how additional distractor layers influence task performance. The second analysis answers the question of whether DDS alpha-blended layers are more visually distinct in the presence of distractors than DDS bump-mapped layers.

Questions One and Two: How do participants perform with multi-layer DDS images compared to single-layer DDS images shown side-by-side? Is there a point where the multitude of additional layers causes enough visual interference that the task is more accurately performed looking at the targets side-by-side?

This analysis sought to answer the question of how task performance for overlaid DDS images compared to performance for the side-by-side DDS images. In the analysis, performance for the *Side-by-Side* condition is compared to performance with *C7*, the seven distractor condition. Because the values of the linear model predict that *C7* will have the highest error percentages for the overlap estimation task and the lowest sketch scores for the sketch task, *C7* represents worst-case performance for overlaid DDS images. Because performance for the *Side-by-side* condition was always worse on average than for *C7* and because *C7* represents worst-case performance, it is possible to statistically compare the *Side-by-side* with just *C7* and draw conclusions for all the overlaid DDS conditions. All analyses were performed using SPSS for Windows statistical software [SPSS, 2001].

Both the overlap estimation task and the sketch task produced the same results in the analysis. The main effect of *Display Condition* was found to be significant ($p < 0.0001$), indicating that performance for the *Side-by-Side* and *C7* was significantly different. This is true for all *Target Display Type* groups, whereas in the pilot study both *Color-Bump* and *Bump-Bump* groups saw performance decreases with additional layers. The main effect of *Target Display Type* was not found to be significant, nor was the interaction effect of *Display Condition * Target Display Type*. Figures 3.66 through 3.71 show plots of average performance by group.

	Mean	Median	Std Deviation
Overall	.079	.053	.088

Table 3.31: Overall mean, median, and standard deviation of participant error in the overlap estimation task. Error is defined as the absolute value of the difference between the estimated area and the actual area. – *Display Condition* levels *C0*-*C7*.

	Mean	Median	Std Deviation
Color-Color	.070	.050	.075
Color-Bump	.079	.048	.098
Bump-Bump	.088	.062	.089

Table 3.32: Overall mean, median, and standard deviation of participant error by *Target Display Type* averaged across all eight levels of *Display Condition* for the overlap estimation task– *Display Condition* levels *C0*-*C7*.

	Mean	Median	Std Deviation
C0	.067	.042	.089
C1	.064	.043	.069
C2	.070	.050	.077
C3	.069	.046	.078
C4	.080	.055	.083
C5	.096	.067	.104
C6	.098	.067	.106
C7	.088	.062	.087

Table 3.33: Overall mean, median, and standard deviation of participant error by *Display Condition* averaged across all three *Target Display Type* levels for the overlap estimation task– *Display Condition* levels *C0*-*C7*.

	Color-Color			*Color-Bump*			*Bump-Bump*		
Display Condition	Mean	Median	Std Deviation	Mean	Median	Std Deviation	Mean	Median	Std Deviation
C0	.067	.040	.092	.063	.039	.089	.072	.049	.087
C1	.054	.040	.051	.055	.041	.058	.082	.054	.090
C2	.068	.050	.073	.075	.049	.090	.067	.054	.067
C3	.069	.046	.072	.060	.038	.069	.080	.058	.090
C4	.071	.056	.070	.065	.040	.084	.103	.082	.088
C5	.071	.050	.073	.101	.063	.114	.116	.089	.113
C6	.079	.057	.092	.125	.088	.131	.089	.065	.083
C7	.083	.062	.070	.091	.054	.107	.092	.081	.081
S-S	.179	.125	.159	.186	.160	.168	.176	.114	.169
I	.049	.037	.040	.049	.036	.047	.060	.050	.058

Table 3.34: Error statistics for the overlap estimation task by *Target Display Type* and *Display Condition*. The bottom two rows are the *Side-by-side* and *Intersection* views. Note that the values for the *Side-by-side* view are over twice those for *C7* in all three groups.

	Mean	Median	Std Deviation
Overall	.648	1.000	.577

Table 3.35: Overall mean, median, and standard deviation of the sketch scores for the intersection sketch task – *Display Condition* levels *C0-C7*. Lower sketch scores indicate poorer performance, and a score of 1 indicates a correct sketch. In all cases the median score is equal to one, showing that over half of the sketches were judged to be correct.

	Mean	Median	Std Deviation
Color-Color	.719	1.000	.503
Color-Bump	.662	1.000	.574
Bump-Bump	.563	1.000	.637

Table 3.36: Overall mean, median, and standard deviation by *Display Condition* averaged across all eight levels of *Display Condition* for the intersection sketch task– *Display Condition* levels *C0-C7*.

	Mean	Median	Std Deviation
C0	.784	1.000	.431
C1	.780	1.000	.437
C2	.713	1.000	.508
C3	.677	1.000	.549
C4	.601	1.000	.594
C5	.541	1.000	.655
C6	.523	1.000	.672
C7	.563	1.000	.652

Table 3.37: Overall mean, median, and standard deviation by *Display Condition* averaged across all three *Target Display Type* levels for the intersection sketch task– *Display Condition* levels *C0-C7*.

	Color-Color			*Color-Bump*			*Bump-Bump*		
Display Condition	Mean	Median	Std Deviation	Mean	Median	Std Deviation	Mean	Median	Std Deviation
C0	.793	1.000	.389	.840	1.000	.403	.720	1.000	.490
C1	.823	1.000	.389	.817	1.000	.390	.700	1.000	.512
C2	.773	1.000	.440	.783	1.000	.411	.583	1.000	.625
C3	.690	1.000	.533	.730	1.000	.497	.610	1.000	.607
C4	.683	1.000	.523	.663	1.000	.500	.457	1.000	.714
C5	.690	1.000	.497	.497	1.000	.713	.437	1.000	.708
C6	.683	1.000	.542	.433	1.000	.739	.453	1.000	.694
C7	.613	1.000	.635	.533	1.000	.670	.543	1.000	.653
S-S	.233	.500	.823	.240	.500	.774	.273	.500	.796
I	.533	1.000	.587	.590	1.000	.537	.363	.500	.709

Table 3.38: Statistics for the intersection sketch task by *Target Display Type* and *Display Condition*. The bottom two rows are the *Side-by-side* and *Intersection* views.

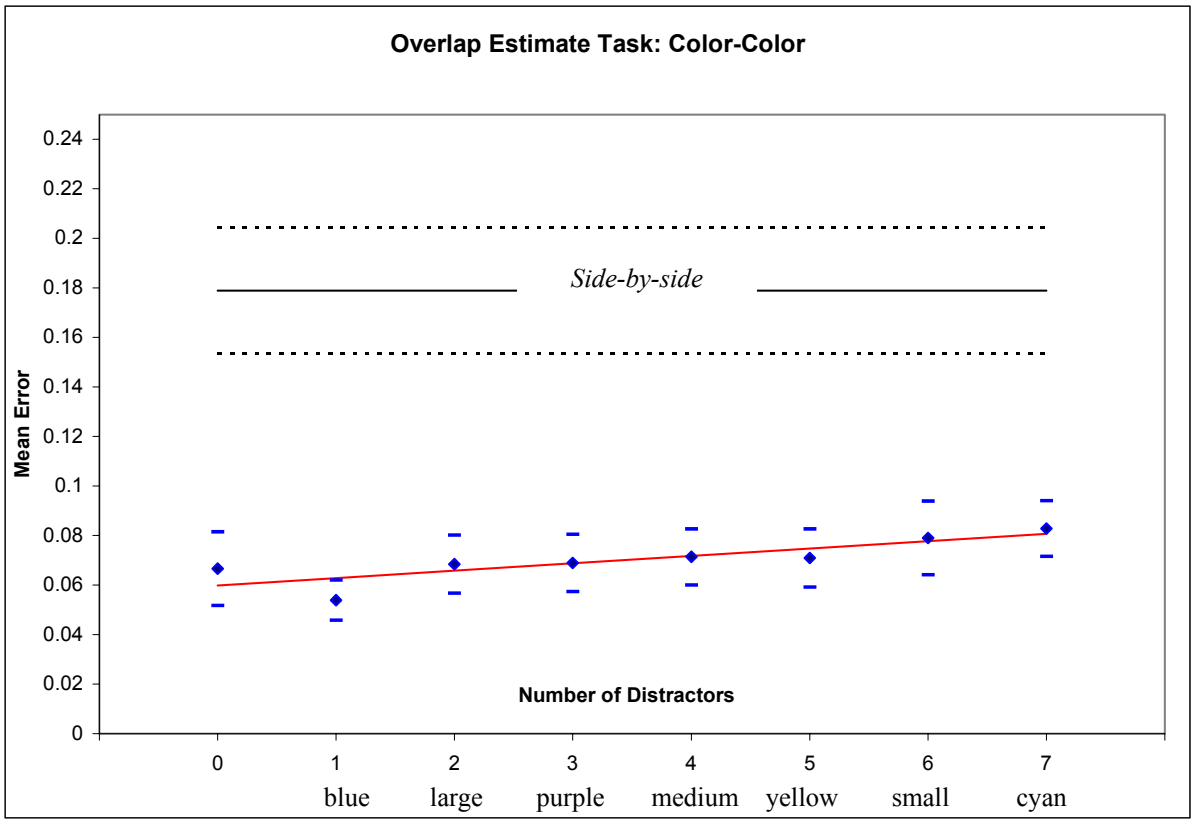

Figure 3.66: Performance for the Overlaid DDS conditions compared to the *Side-by-Side* condition for the *Color-Color* group. The blue diamonds mark the average error for each *Display Condition* level *C0-C7*. The blue horizontal bars above and below the diamonds mark the standard error, which is a measure of the amount of variation in the data. The red line is the linear fit. The solid black horizontal line at the top of the graph shows the average performance for the *Side-by-side view*. The black dashed lines above and below it show the standard error for the *Side-by-side view*. Distractor type is listed at the bottom of the graph.

People performed significantly worse when the target images were displayed side-by-side than when the target images were displayed in one image with seven other distractor shapes. This illustrates the power of overlaying images when performing spatial correlation tasks.

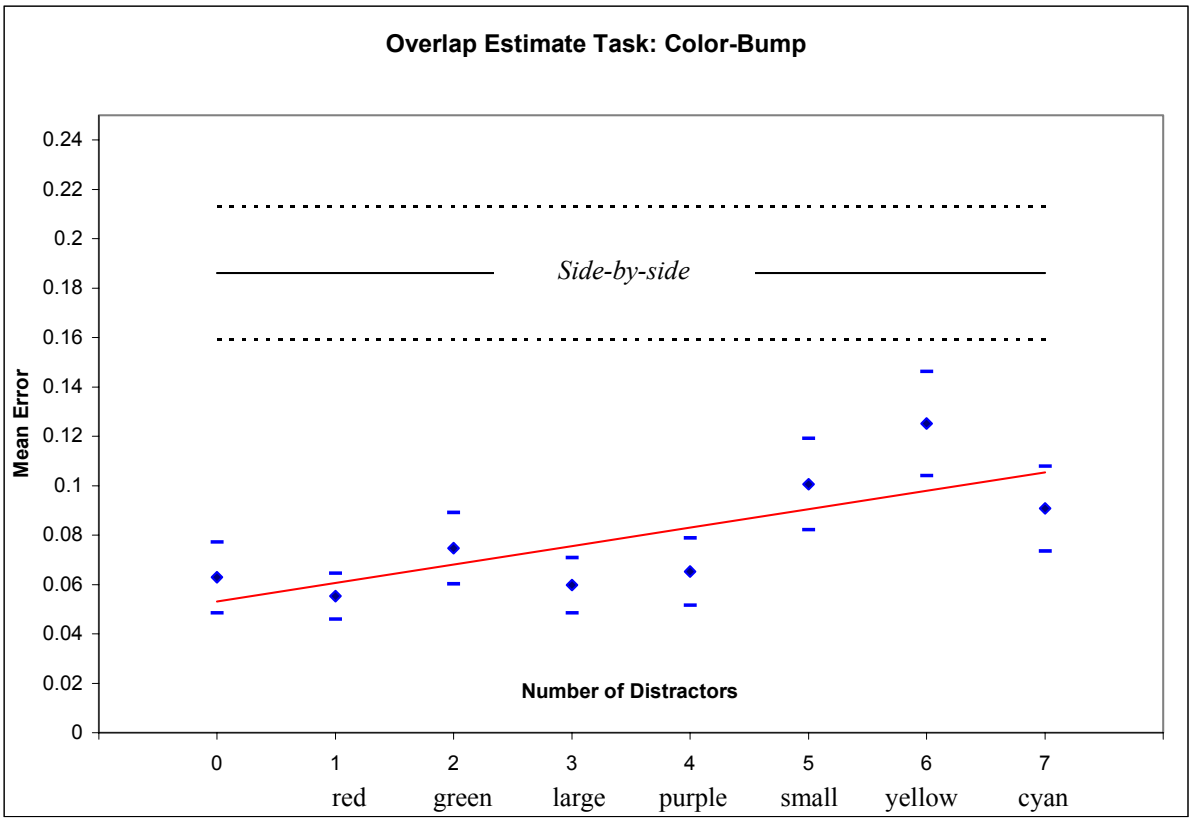

Figure 3.67: Performance for the Overlaid DDS conditions compared to the *Side-by-Side* condition for the *Color-Bump* group. The blue diamonds mark the average error for each *Display Condition* level *C0-C7*. The blue horizontal bars above and below the diamonds mark the standard error, which is a measure of the amount of variation in the data. The red line is the linear fit. The solid black horizontal line at the top of the graph shows the average performance for the *Side-by-side view*. The black dashed lines above and below it show the standard error for the *Side-by-side view*. Distractor type is listed at the bottom of the graph.

People performed significantly worse when the target images were displayed side-by-side than when the target images were displayed in one image with seven other distractor shapes. In the graph condition *C6* actually performed worse than *C7*, when *C6* is compared to *the Side-by-side* condition it is also significantly better ($p < 0.007$).

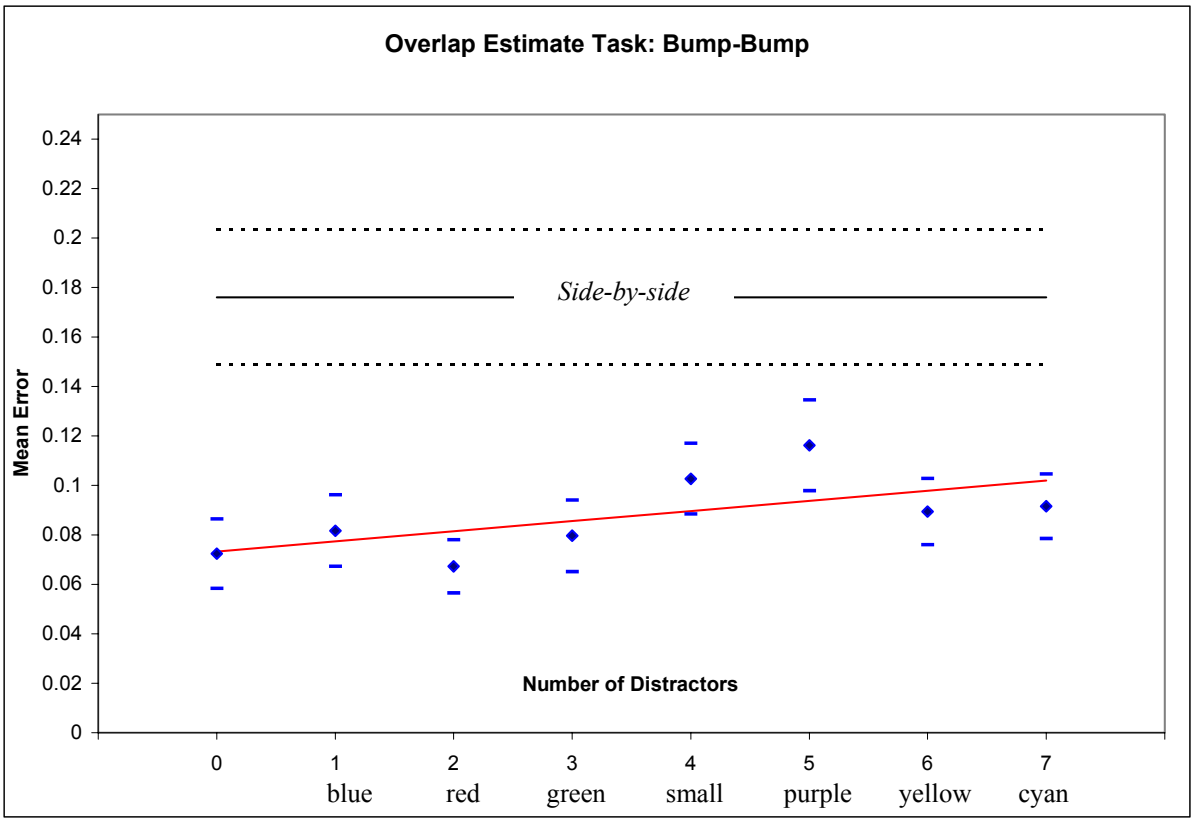

Figure 3.68: Performance for the Overlaid DDS conditions compared to the *Side-by-Side* condition for the *Bump-Bump* group. The blue diamonds mark the average error for each *Display Condition* level *C0-C7*. The blue horizontal bars above and below the diamonds mark the standard error, which is a measure of the amount of variation in the data. The red line is the linear fit. The solid black horizontal line at the top of the graph shows the average performance for the *Side-by-side view*. The black dashed lines above and below it show the standard error for the *Side-by-side view*. Distractor type is listed at the bottom of the graph.

People performed significantly worse when the target images were displayed side-by-side than when the target images were displayed in one image with seven other distractor shapes. In the graph condition *C5* actually performed worse than *C7*, when *C5* is compared to the *Side-by-side* condition it is also significantly better ($p = 0.001$).

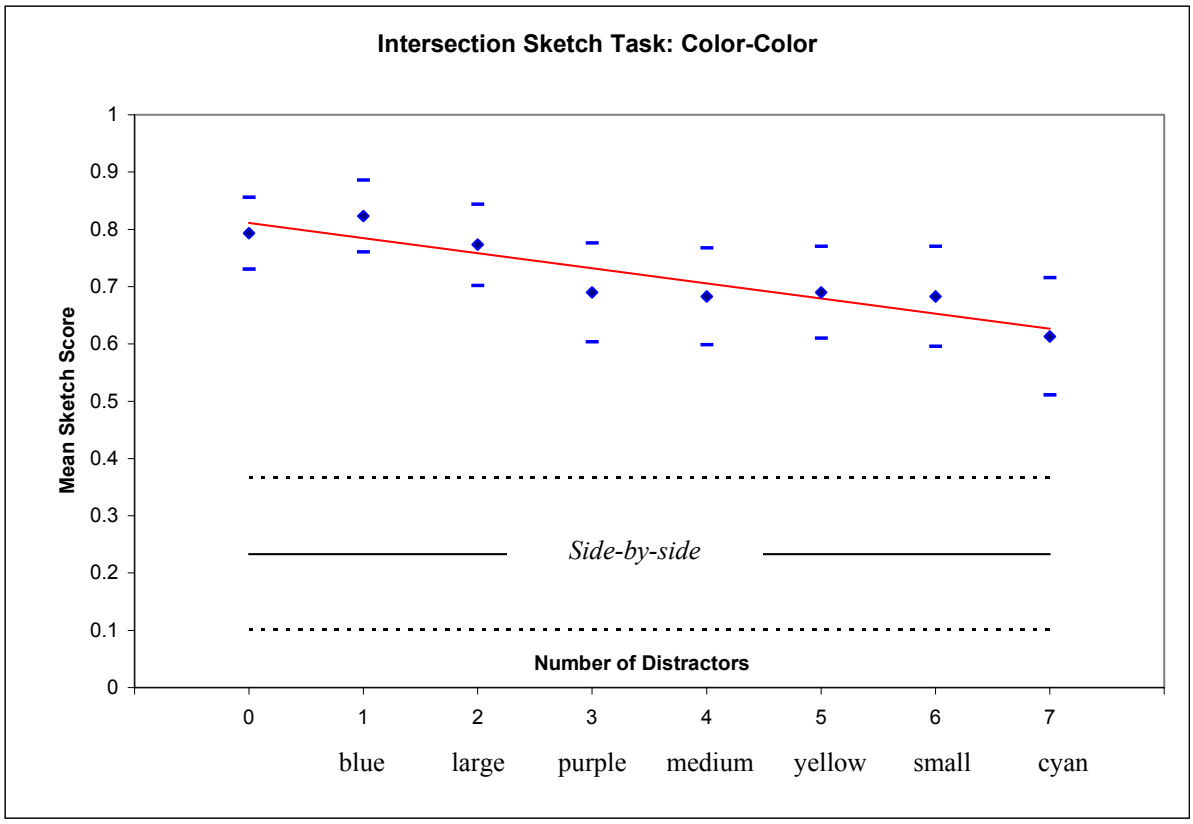

Figure 3.69: Performance for the Overlaid DDS conditions compared to the *Side-by-Side* condition for the *Color-Color* group. The blue diamonds mark the average error for each *Display Condition* level *C0-C7*. The blue horizontal bars above and below the diamonds mark the standard error, which is a measure of the amount of variation in the data. The red line is the linear fit. The solid black horizontal line at the top of the graph shows the average performance for the *Side-by-side view*. The black dashed lines above and below it show the standard error for the *Side-by-side view*. Distractor type is listed at the bottom of the graph.

People performed significantly worse when the target images were displayed side-by-side than when the target images were displayed in one image with seven other distractor shapes.

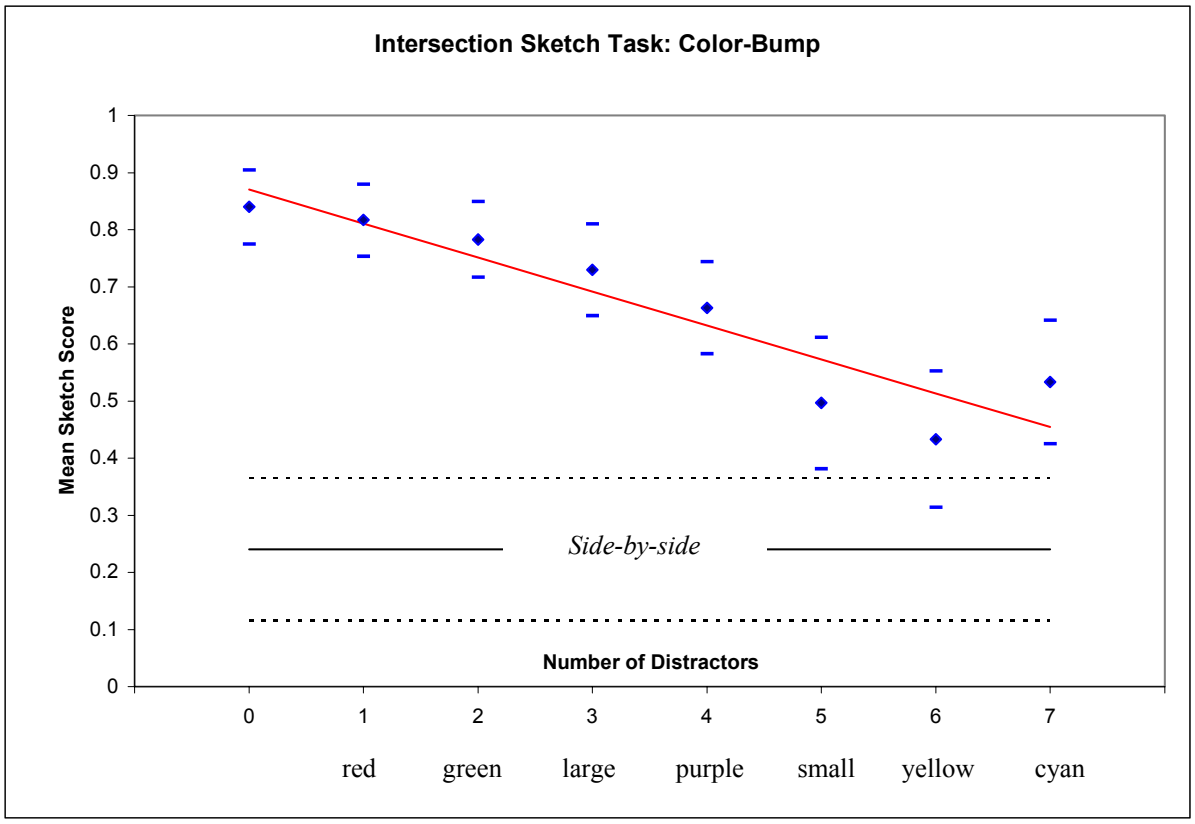

Figure 3.70: Performance for the Overlaid DDS conditions compared to the *Side-by-Side* condition for the *Color-Bump* group. The blue diamonds mark the average error for each *Display Condition* level *C0-C7*. The blue horizontal bars above and below the diamonds mark the standard error, which is a measure of the amount of variation in the data. The red line is the linear fit. The solid black horizontal line at the top of the graph shows the average performance for the *Side-by-side view*. The black dashed lines above and below it show the standard error for the *Side-by-side view*. Distractor type is listed at the bottom of the graph.

People performed significantly worse when the target images were displayed side-by-side than when the target images were displayed in one image with seven other distractor shapes. In the graph condition *C6* actually performed worse than *C7*, when *C6* is compared to the *Side-by-side* condition it is also significantly better ($p < 0.013$).

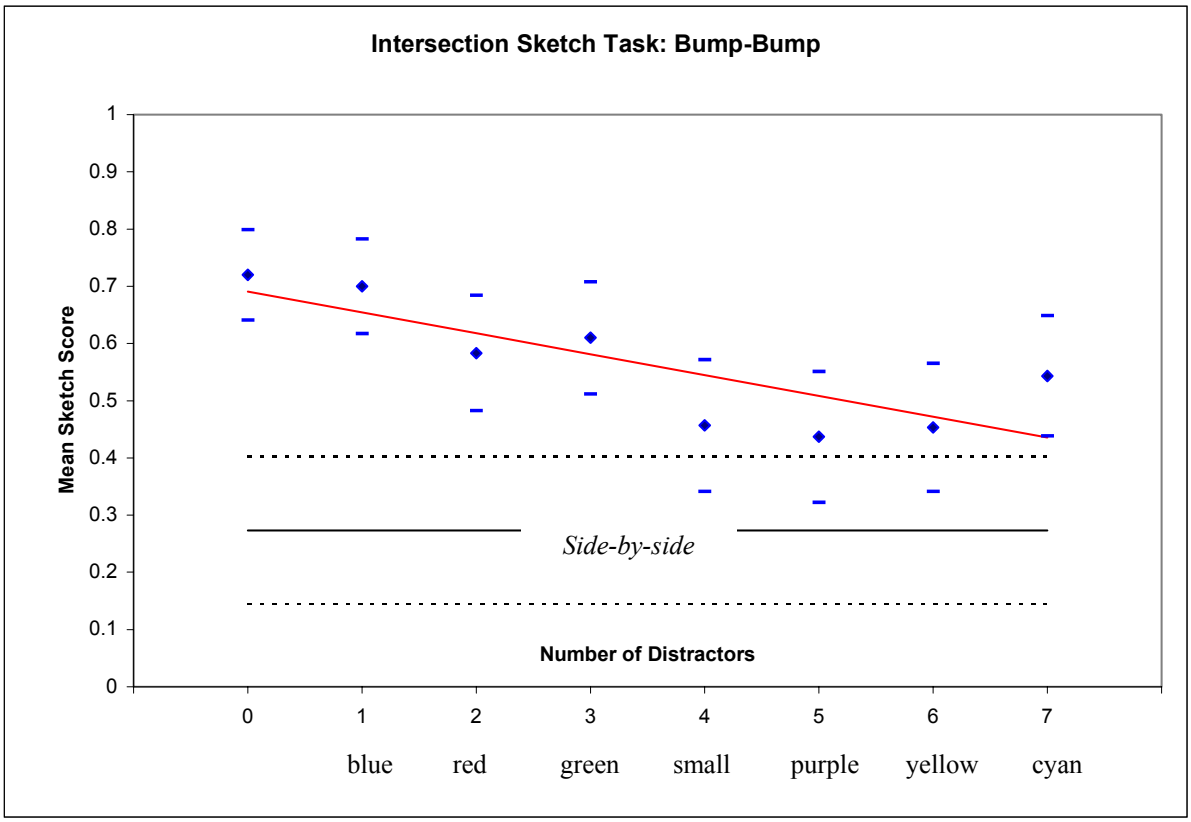

Figure 3.71: Performance for the Overlaid DDS conditions compared to the *Side-by-Side* condition for the Bump-Bump group. The blue diamonds mark the average error for each *Display Condition* level *C0*-*C7*. The blue horizontal bars above and below the diamonds mark the standard error, which is a measure of the amount of variation in the data. The red line is the linear fit. The solid black horizontal line at the top of the graph shows the average performance for the *Side-by-side view*. The black dashed lines above and below it show the standard error for the *Side-by-side view*. Distractor type is listed at the bottom of the graph.

People performed significantly worse when the target images were displayed side-by-side than when the target images were displayed in one image with seven other distractor shapes. In the graph condition *C5* actually performed worse than *C7*; when *C5* is compared to the *Side-by-side* condition it is also significantly better ($p < 0.044$).

Question Three: Are DDS alpha-blended layers more visually salient than DDS bump-mapped layers in a multi-layer visualization?

This analysis looks at participant performance on the two outcome measures, error in overlap estimation and sketch shape accuracy, and relates the outcome measures to the number of distractors in the test images. The analysis also considers differences in *Target Display Type* groups and whether performance on either task was different for the *Color-Color*, *Color-Bump*, or *Bump-Bump* groups.

A regression F-test is performed to test for significant slope due to the number of distractors in the trial images, this is the same analysis as in the pilot study, see page 170 for details.

Tables 3.39 and 3.40 show the SAS output for the F-test. Plots of the results are shown in Figures 3.66 and 3.67. For both measures the number of distractors significantly affects performance ($p < 0.001$). *Target Display Type* is also found to be significant – performance is different when both targets are displayed with DDS alpha-blending versus DDS bump-mapping ($p = 0.0098$) for the overlap estimation task and ($p = 0.0002$) for the sketch task. The interaction term, *Display Condition* * *Target Display Type* is also significant ($p = 0.0232$) for the overlap estimation task and ($p = 0.0148$) for the sketch task. A significant interaction effect indicates that the slopes for each *Target Display Type* group are significantly different in direction, as is shown in the graphs. That both measures produce the same statistical result strengthens the evidence, as the two measures were very different in terms of the task involved. The linear models based on the estimated intercepts and slopes from the F-test are given below:

Overlap Estimation Task:

When analyzed by *Target Display Type* the effect of the number of distractors is significant: *Color-Color* ($p = 0.0028$); *Color-Bump* ($p = 0.0003$).; *Bump-Bump* ($p = 0.0008$).

$$\text{Error}_{\text{Color-Color}} = 0.06 + 0.003 * \text{Number of Distractors} \quad (3.10)$$

$$\text{Error}_{\text{Color-Bump}} = 0.05 + 0.007 * \text{Number of Distractors} \quad (3.11)$$

$$\text{Error}_{\text{Bump-Bump}} = 0.07 + 0.004 * \text{Number of Distractors} \quad (3.12)$$

Type 3 Tests of Fixed Effects: Overlap Estimate

Effect	Num DF	Den DF	F Value	Pr > F
DC	1	29	23.06	<.0001
TDT	2	58	5.01	0.0098
DC*TDT	2	58	4.02	0.0232

Table 3.39: SAS output for the main effects of *Display Condition*, and *Target Display Type*, and the interaction effect of *Display Condition* * *Target Display Type* for the overlap estimation task. The number of observations, N, is 3600 as only conditions *C0-C7* are included in the analysis.

Type 3 Tests of Fixed Effects: Intersection Sketch

Effect	Num DF	Den DF	F Value	Pr > F
DC	1	29	33.44	<.0001
TDT	2	58	10.12	0.0002
DC*TDT	2	58	4.54	0.0148

Table 3.40: SAS output for the main effects of *Display Condition*, and *Target Display Type*, and the interaction effect of *Display Condition* * *Target Display Type* for the sketch task. The number of observations, N, is 3600 as only conditions *C0-C7* are included in the analysis.

Sketch Task:

When analyzed by *Target Display Type* the effect of the number of distractors is significant: *Color-Color* ($p = <.00010.\ 0028$); *Color-Bump* ($p < .0001$); *Bump-Bump* ($p = 0.0031$).

$$\text{Sketch}_{\text{Color-Color}} = 0.81 - 0.026 * \text{Number of Distractors} \quad (3.13)$$

$$\text{Sketch}_{\text{Color-Bump}} = 0.87 - 0.059 * \text{Number of Distractors} \quad (3.14)$$

$$\text{Sketch}_{\text{Bump-Bump}} = 0.69 - 0.036 * \text{Number of Distractors} \quad (3.15)$$

The slopes for the *Color-Color* and *Bump-Bump* groups are similarly shallow and the slope for the *Color-Bump* group is significantly steeper than for the other two groups. The line for the *Color-Bump* group crosses that of the *Color-Color* group between 1 and 2 distractors for both outcome measures.

In the graphs in Figures 3.66 and 3.67, the lines for the *Color-Color* and *Bump-Bump* groups appear to be near parallel to one another, with a significant offset. In the analysis, parallel lines offset from one another should result in a significant main effect for *Target Display Type*, but should not result in a significant interaction of *Display Condition* * *Target Display Type*. If the *Color-Bump* group is removed from the analysis, the output from the SAS software shows no significant interaction effect ($p = 0.4276$) for the overlap task and ($p = 0.3778$) for the sketch task. This indicates that the slopes for the *Color-Color* and *Bump-Bump* groups are indeed near-parallel.

The significant interaction effect comes from the change in performance characteristics for the *Color-Bump* group as distractors are added to the test images. At zero and one distractors performance for the *Color-Bump* group is near that of the *Color-Color* group, while at seven distractors performance for the *Color-Bump* group is near that of the *Bump-Bump* group. The line for the *Color-Bump* group is parallel to neither the *Color-Color* nor the *Bump-Bump* groups.

It is interesting to ask why the effect of the number of distractors at seven layers is the same for the *Color-Bump* and the *Bump-Bump* groups. If the decrease in performance was caused by interference with the visual discrimination of the DDS bump-mapped layers, then one might think that the effect would be twice as strong for two bump-mapped targets as for one. One possible explanation for the behavior revealed by the graphs is that the effect of distractors on the visual discrimination of the DDS bump-mapped layers is stronger than that for the DDS alpha-blended layers *up to a constant amount*. This will be discussed in more detail at the end of the Chapter.

Magnitude of Effect for Overlap Estimation Task

The difference in performance for zero distractors and seven distractors predicted by the analysis for the *Color-Color* group is only 2 percentage points, and a comparison of means for C0 and C7 found no significant difference at the 0.05 level *($p = 0.10$)*. The difference in performance for zero distractors and seven distractors predicted by the analysis for the *Bump-Bump* group is only slightly higher: 2.86 percentage points, and a comparison of means for C0 and C7 found no significant difference at the 0.5 level *($p = 0.077$)*. The difference in performance for zero distractors and seven distractors predicted by the analysis for the *Color-Bump* group is only slightly more: 5 percentage points, and a comparison of means for C0 and C7 found no significant difference at the .05 level *($p = 0.07$)*.

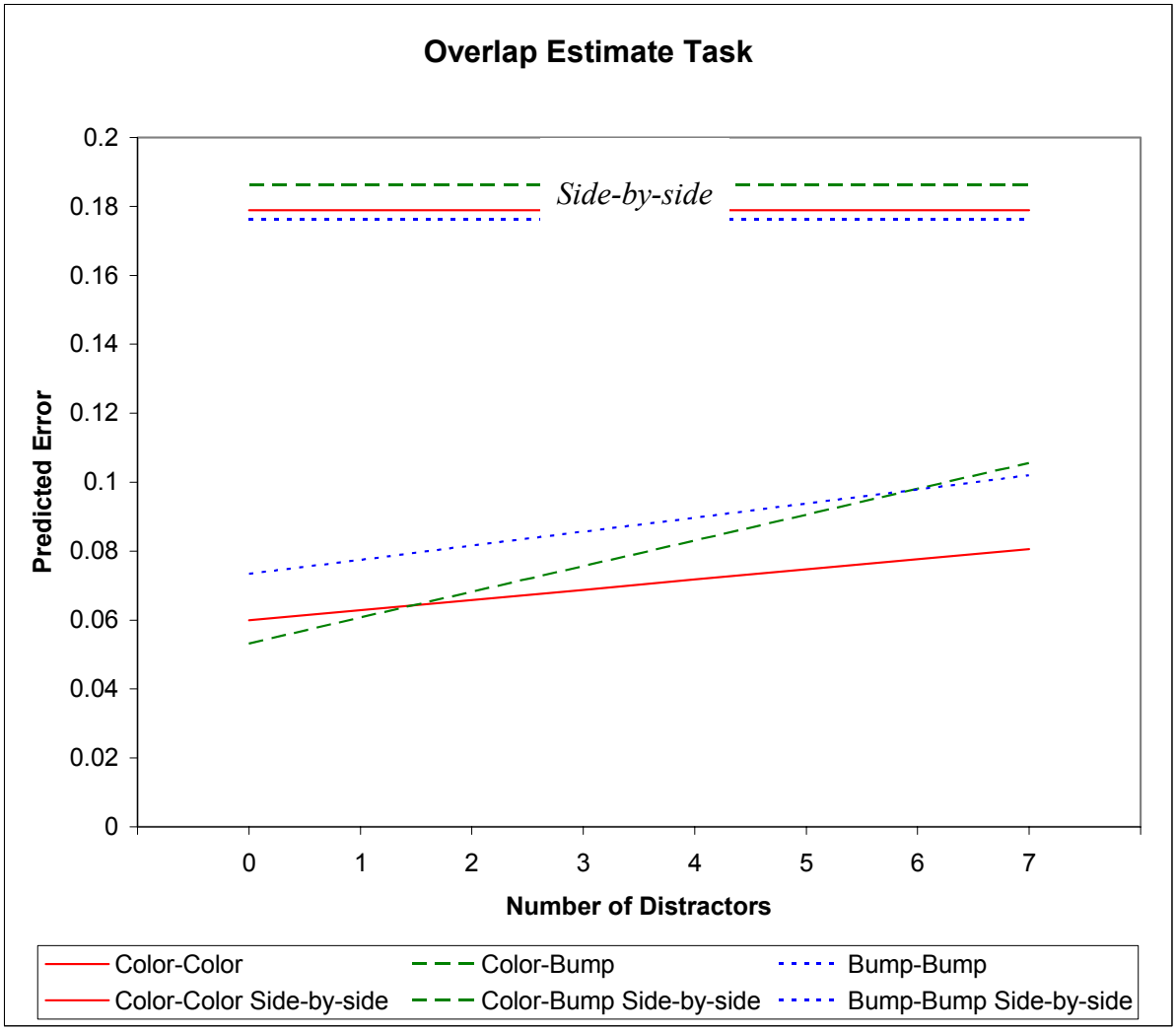

Figure 3.72: Plots of the linear fits for the overlap estimation task. The lines for the *Color-Color* and *Bump-Bump* groups are close to parallel, with an almost constant displacement. Interestingly, the *Color-Bump* group starts out lower than the *Color-Color* group and ends higher than the *Bump-Bump* group. Performance for the *Color-Bump* group is like the *Color-Color* group for few distractors, and becomes like the *Bump-Bump* group as the number of distractors increases.

How many distractors could be handled using each of the techniques? This experiment does not answer that. As a speculation, linear extrapolation would put the crossing point at 40 distractors for the *Color-Color* group, 25 distractors for the *Bump-Bump* group, and 18 distractors for the *Color-Bump* group.

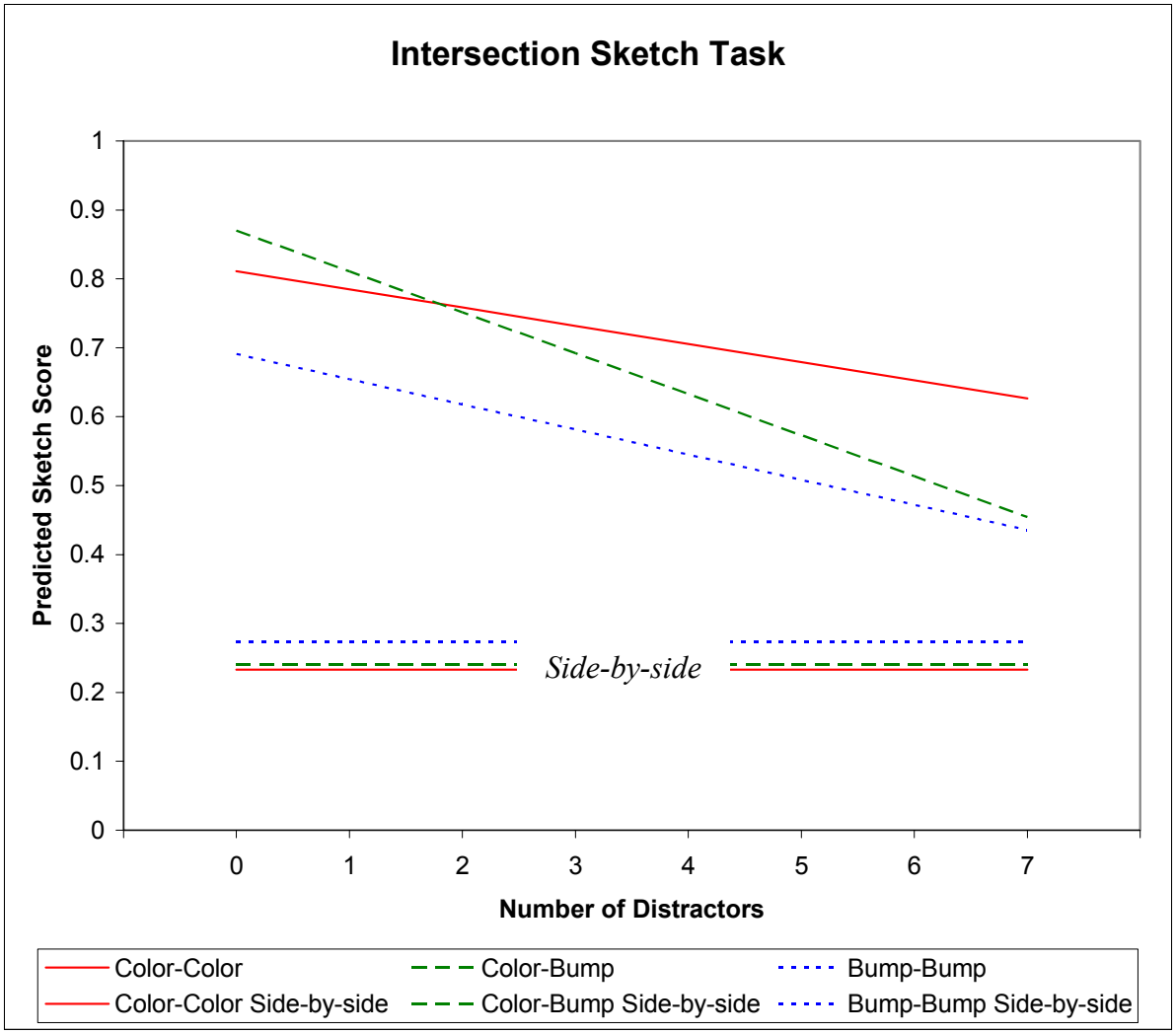

Figure 3.73: Plots of the linear fits for the sketch task. The results are the same as for the overlap estimation task, showing strong reliability. The lines for the *Color-Color* and *Bump-Bump* groups are close to parallel, with an almost constant displacement. Interestingly, the *Color-Bump* group starts out with better performance than the *Color-Color* group and ends with slightly better performance than the *Bump-Bump* group. Performance for the *Color-Bump* group is like the *Color-Color* group for few distractors, and becomes like the *Bump-Bump* group as the number of distractors increases.

In the comparison of means analysis, when performance is compared to the *Side-by-side view*, all three *Target Display Type* groups are significantly better than the *Side-by-side view* and the effect of *Target Display Type* is not found to be a significant factor. However, in the linear fit analysis, *Target Display Type* and the interaction term *Target Display Type* * *Display Condition* are found to be significant, indicating that there is a measurable difference between groups. Figures 3.66 through 3.71 show clear differences between performance for the *Color-Color* and *Color-Bump* and *Bump-Bump* groups. Although the effect is smaller than in the pilot study, there is a break in performance for the *Color-Bump* group when the third bump layer is added to the images, distractor layer six in Figure 3.67. There is also a break in performance for the *Bump-Bump* group when the third bump layer is added in distractor layer four in Figure 3.68. This effect is similar to the one seen in the pilot study when there were up to four bump layers in the trial images. The conclusion is that the bump-mapped layers are interfering more with each other than with the alpha-blended layers. The question of whether alpha-blended layers are more visually salient than bump-mapped layers in the presence of distractors is more readily answered when the results for the pilot and main study are compared; this is the topic of the next section.