

Pilot and Main Experimental Design

Two experimental studies are described: a pilot study and main study. Both studies investigate the three experimental questions presented earlier in this chapter, repeated here:

Question One: Does overlay provide a performance benefit for an overlap estimation task and an intersection-sketch task? How do participants perform with overlaid multi-layer DDS images compared to single-layer DDS images shown side-by-side?

Question Two: Where is the benefit crossover? If presenting targets overlaid in a single image allows for more accurate performance for an overlap estimation task and an intersection-sketch task, then 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?

Question Three: Are DDS alpha-blended layers more visually salient than DDS bump-mapped layers in a multi-layer visualization? How do the display parameters of individual target DDS layers influence their visual salience in the presence of distractors?

Both the pilot and main studies manipulated the same two independent variables during the experimental trials and collected the same performance data. The studies are presented in detail below. First, I outline the common experimental variables, both dependent and independent. Next, I present the experimental design, results, and analysis of the pilot study in full detail, following which I present the main study without repetitive detail, referring to the pilot study wherever possible. The section at the end of this chapter presents an overall comparison of the two studies and concludes with a summary of the joint findings.

Experimental Variables

Display Condition and *Target Display Type* are the two independent variables manipulated during experimental trials. The *Display Condition* variable encompasses two linked questions: (1) how the number of distractor layers affects performance on the overlap estimation and intersection sketch tasks and (2) how performance for two overlaid DDS data layers with distractors compares to performance for a side-by-side view of the data, without distractors. There are ten levels of the variable *Display Condition*, eight levels include overlaid DDS images of the two target layers with zero to seven distractor layers; these levels are referred to as *C0-C7*, respectively. The other two levels are referred to as *Side-by-side* and *Intersection* view. In the *Side-by-side* view, target shapes are displayed in two separate side-by-side DDS images, without distractors. In the *Intersection* view, target A and the intersection region of target A and target B are overlaid in a single image, without distractors. Example images for all levels of *Display Condition* are shown in Figures 3.19 through 3.34 for the pilot study and Figures 3.50 through 3.65 for the main study.

The *Side-by-side* and *Intersection* conditions are expected to produce worse and better performance, respectively, than the other conditions. There are two tasks required of the participants: to give a numeric estimate of the percentages of target A that is overlapped by target B and to sketch the intersection region of targets A and B. These tasks are most readily accomplished if the two targets are displayed overlaid in a single image. Thus, performance should be better for condition *C0* than for the *Side-by-side* view. If DDS layers are visually distinct, then additional layers should not detract from the participants' ability to see the target layers and the benefit of an overlay presentation should still outweigh the side-by-side presentation, thus performance should be better even for seven distractions, *C7*, than for the side-by-side view. The *Intersection* condition simplifies the task for the participants: because the intersection region is displayed directly, the number of cognitive steps required to complete the tasks are fewer. Thus *Intersection* should be comparable to *C0*, which has no distractors.

The second independent variable, *Target Display Type*, explores the question of whether DDS alpha-blended layers are more visually salient than the DDS bump-mapped layers, in the presence of both DDS alpha-blended and DDS bump-mapped distractor layers. To test this, *Target Display Type*, has three values, one where both targets are displayed with DDS alpha-blended layers, a second where both are displayed with DDS bump-mapped layers, and a third where target A is displayed with DDS alpha-blending and target B is displayed with DDS bump-mapping.

Within each study the same display parameters are assigned to all layers, targets and distractors, for the three different *Target Display Type* levels, only the instructions to the participant, indicating which layers are targets, changes. For example, in the main study, the target layers in the *Color-Color* session are always the red and green spots. Distractors are added in the following order: blue spots, large bumps, purple spots, medium bumps, yellow spots, small bumps, and cyan spots. In the *Color-Bump* session, the same nine colors and spot sizes are used, however, in this case, the target layers are the blue spots and the medium bumps, and distractors are added in the following order: red spots, green spots, large bumps, purple spots, small bumps, yellow spots, and cyan spots. In both sessions the display parameters for the layers are the same; it is the instructions to the participants that determine which layers are targets and which are distractors.

The goal of this study is to determine, for a set of nine DDS layers, if all nine are visually distinct from one another. Therefore the display layers in the experiment remain constant, and five different layers are specified and perceived as targets during the three different sessions. If participants are able to visually separate each of the five layers – even in images with nine layers, then this result provides strong evidence that each layer is visually distinct.

Dependent Measures

Two tasks are the same for the pilot and main studies. The first task is to estimate the percentage of target A that is overlapped by target B. The second task is to sketch the intersection region between targets A and B. The pilot study has an additional, third, task: to sketch target A – referred to in the pilot study section as *single-target sketch*. The participants type their responses to the first question on the keyboard. Labeled, non-ruled, paper is provided for the sketches. A *trial* refers to the overlap estimate and sketches completed by the participants for one image.

Two performance measures are the same for the pilot and main studies. The first metric represents the error in overlap estimate, compared to the correct percentage of overlap. The error is calculated as the absolute value of the difference. Error is defined similarly in the experiments by Healey et al. [1996], discussed at the beginning of this Chapter.

The second (and third) metrics are sketch scores for the intersection sketch and the single-target sketch. The sketches are compared to the correct computer-derived shapes for grading. For the pilot study, sketches are graded on a binary scale of 0 for incorrect sketches and 1 for correct sketches. For the main, study sketches are graded on a three-point scale of -1 for incorrect, 0 for mostly correct, and 1 for exact, where to achieve an exact score all vertices and edges must be included.

Pilot Study: Fall 1999

Methods

Design

The experimental design is within-subject for the *Display Condition* independent variable – each participant experienced five different trials for each of the ten categories of *Display Condition: C0-C7, Intersection, and Side-by-side view*. However, the design is between-subject for the *Target Display Type* variable, as each participant was randomly assigned to just one of the three levels of *Target Display Type*. There are 50 observations per participant for each of the overlap estimation and the intersection sketch tasks. A total of 30 participants completed the study, giving 1500 overlap estimates, intersection sketches, and single-target sketches.

Participants

Number, age, and gender

Thirty people, twenty-three men and seven women, are participants in the study. Each is between the ages of 18 and 40.

Criteria for exclusion of participants

Participants are excluded during the recruitment process and before the start of the experiment based on the following criteria:

- Participants who do not have normal or corrected vision;
- Participants who do not have normal color vision;
- Participants under the age of 18 or over the age of 40 (age 40 is typically when age-related vision deterioration begins).

Recruiting and inducements

Participants are recruited through electronic mail postings within the Computer Science department at University of North Carolina, Chapel Hill. All are either staff members or students in the Department of Computer Science at the University of North Carolina, Chapel Hill. Participation is voluntary, and participants are not paid.

Population Sampling and Generalization

The sample of participants is not meant to represent the population of students on the UNC campus in general. All participants are from the Computer Science Department, thus it is not reasonable to generalize the findings to the larger campus population.

Target and Distractor Layers

Target and distractor shapes are generated from a pool of seven elementary shapes: equilateral triangles, squares, circles, horizontally aligned ellipses, vertically aligned ellipses, horizontally aligned rectangles, and vertically aligned rectangles. The shapes are each displaced from the center of the image with the restriction that part of the shape overlap the center, thus guaranteeing that all targets and distractors overlap.

Targets A are displaced to the left, and targets B are displaced to the right. Distractor shapes are displaced in any of the eight principal directions from the center of the image. Displacement and scale factors for each shape are randomly generated from a uniform distribution so that no two shapes were exactly the same. Table 3.1 shows the fifty pairs of target shapes and Table 3.2 shows the five sets of seven distractors. Figure 3.14 provides a histogram of overlap percentages for all target pairs.













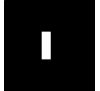









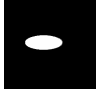









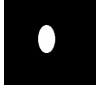



































































A											35.5%	25.6%	29.6%	30.8%	41.2%
B											50.0%	28.4%	85.2%	66.4%	63.3%
C											37.0%	42.5%	60.1%	36.2%	49.0%
D											49.6%	46.2%	32.0%	30.5%	30.4%
E											63.6%	6.1%	55.4%	32.5%	31.4%
F											50.7%	32.2%	28.5%	56.0%	25.5%
G											34.7%	47.5%	37.1%	72.2%	34.7%
H											65.3%	8.5%	14.4%	77.3%	30.7%
I											37.9%	55.4%	52.0%	58.9%	19.2
J											24.4%	46.4%	39.6%	63.0%	76.0%

Table 3.1: Target pair images and correct percentage of target A that is overlapped by target B. Target pairs are assigned to counterbalance group A-J (leftmost column). The five items in each counterbalance group is tested at every level of *Display Condition* and *Target Display Type*.

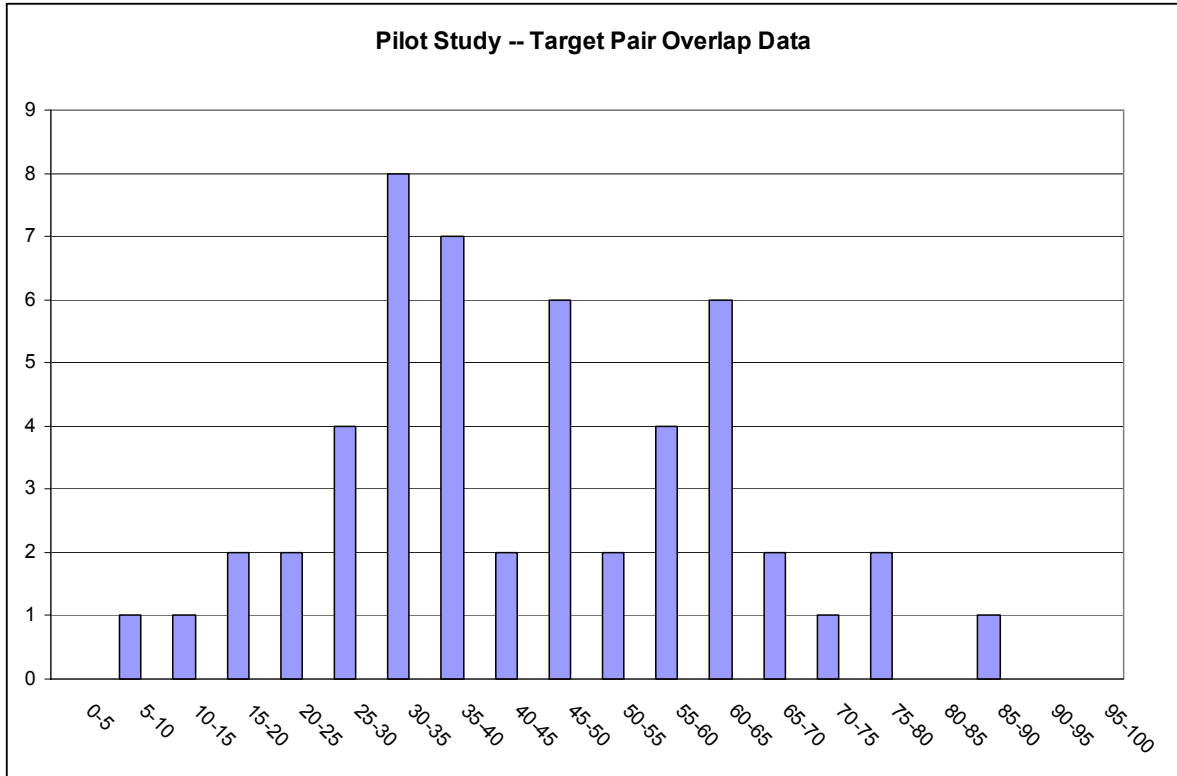


Figure 3.14: Histogram of overlap percentages.

Display Condition: Counterbalancing

The seven elementary shapes are each assigned to one of 50 target pairs. These target pairs are assigned to one of ten groups, which are then counterbalanced across participants and *Display Condition* levels (see Tables 3.1 and 3.3). This counterbalancing ensures that each target pair occurs once in each level of *Display Condition*, thus no one level of *Display Condition* is unfairly affected by a particularly easy or difficult target pair.

With seven distractors and five different trials per *Display Condition*, there are a total of 35 distractor shapes, five of each of the elementary shapes: equilateral triangles, circles, squares, horizontal ovals, vertical ovals, horizontal rectangles, and vertical rectangles. Distractors accumulated across levels, so that in *C7* the entire column of distractors in Table 3.2 is displayed in the trial image. Because distractors accumulate, participants view distractors in *C1* seven times as often as they view the distractors in the *C7* category. Participants may be better able to ignore familiar distractors, which may result in lower errors in trials with more distractors than if the distractors were always different.

	1	2	3	4	5
C1					
C2					
C3					
C4					
C5					
C6					
C7					

Table 3.2: Distractors are each assigned randomly to one of the conditions C1 through C7. The five items are assigned to one of the five target-pair items in Table 3.1.

Participant ID	C0	C1	C2	C3	C4	C5	C6	C7	Side-by-side-View	Intersection-View
1	A	B	C	D	E	F	G	H	I	J
2	J	A	B	C	D	E	F	G	H	I
3	I	J	A	B	C	D	E	F	G	H
4	H	I	J	A	B	C	D	E	F	G
5	G	H	I	J	A	B	C	D	E	F
6	F	G	H	I	J	A	B	C	D	E
7	E	F	G	H	I	J	A	B	C	D
8	D	E	F	G	H	I	J	A	B	C
9	C	D	E	F	G	H	I	J	A	B
10	B	C	D	E	F	G	H	I	J	A

Table 3.3: Target-pair groups A-J from Table 3.1 are counterbalanced across participants and *Display Condition* so that each target pair is presented in each experimental condition. There are five target pairs per group. Each row of this table represents 50 different trial images presented to a participant during one session. The entire table represents the 500 trial images used for one level of *Target Display Type*.

Target Display Type: Assignment of Color and Gaussian Standard Deviation

The spot color and standard deviations remain constant for the DDS layers for all three levels of *Target Display Type*. Only the instructions to the participants change: in the *Color-Color* session, target A is specified as the DDS alpha-blended layer distinguished by greenish-yellow hue and target B is specified as the DDS alpha-blended layer distinguished by a mustard hue. In the *Color-Bump* condition target A is specified as the DDS alpha-blended layer with lavender hue and target B is specified as the smallest DDS bump-mapped layer. In the condition *Bump-Bump* target A is specified as the smallest bump layer and target B is specified as the largest bump layer. The exact parameters for all nine layers and all three levels of *Target Display Type* are given in Tables 3.4 through 3.7. Figures 3.15 through 3.17 present color swatches for each layer. In the table the layer order column represents not only the order in which the layers were applied by the software, but also the perceived ordering (upper layers appear to lie on top of lower layers).

Target layers have value 1 inside the target shape area and 0.25 outside the shape area whereas distractor shapes have value 1 inside the shape area and value 0 outside. This is done to provide an additional visual cue to the participants as to which layers are the current target layers, however doing so may reduce the salience of the target boundary, as the transition is not from full-on to full-off.

Materials

All 50 target pairs are tested in each of the ten levels of *Display Condition* and in each of the three levels of *Target Display Type*. For each trial the target pairs are displayed with a different number of distractors and/or different display parameters – making each trial image unique. Fifty target pairs, each at ten different levels of *Display Condition*, means the generation of 500 different images for each level of *Target Display Type*, and a total of 1500 different trial images for the study overall.

Layer Order (Lowest-highest)	Layer Type	RGB Color Coordinates	HL S Color Coordinates	Relative Spot Size	Apparent height	Color-Color	Color-Bump	Bump-Bump
1	Large			1.0	2.0	C1	C2	target B
2	Medium			0.4	2.0	C5	C4	C5
3	Small			0.25	2.0	C7	C6	C3
4	Tiny			0.14	2.0	C3	target B	target A
5	Pink	236,102,166	331,169,199	0.83		C2	C1	C1
6	Lavender	184,137,254	264,196,251	0.5		C4	target A	C2
7	Turquoise	102,218,187	164,160,156	0.5		C6	C3	C4
8	Green/yellow	137,205,58	88,132,152	0.33		target A	C5	C6
9	mustard	246,187,33	43,140,235	0.33		target b	C7	C7

Table 3.4: The apparent ordering (lowest to highest), display parameters, and function of the nine DDS layers in the Pilot experiment. The last three columns on the right show the function of the layers in the different *Target Display Type Groups*. Five layers out of the nine were tested as targets in the experiment.

Function	Layer type	Layer Order
target A	green-yellow	8
target B	mustard	9
C1	large bump	1
C2	pink	5
C3	tiny bump	4
C4	lavender	6
C5	medium bump	2
C6	turquoise	7
C7	small bump	3

Table 3.5: Display parameter assignments for the targets and distractors in the *Color-Color* group.

Function	Layer type	Layer Order
target A	lavender	6
target B	tiny bump	4
C1	pink	5
C2	large bump	1
C3	turquoise	7
C4	medium bump	2
C5	green-yellow	8
C6	small bump	3
C7	mustard	9

Table 3.6: Display parameter assignments for the targets and distractors in the *Color-Bump* group.

Function	Layer type	Layer Order
target A	tiny bump	4
target B	large bump	1
C1	pink	5
C2	lavender	6
C3	small bump	3
C4	turquoise	7
C5	medium bump	2
C6	green-yellow	8
C7	mustard	9

Table 3.7: Display parameter assignments for the targets and distractors in the *Bump-Bump* group.

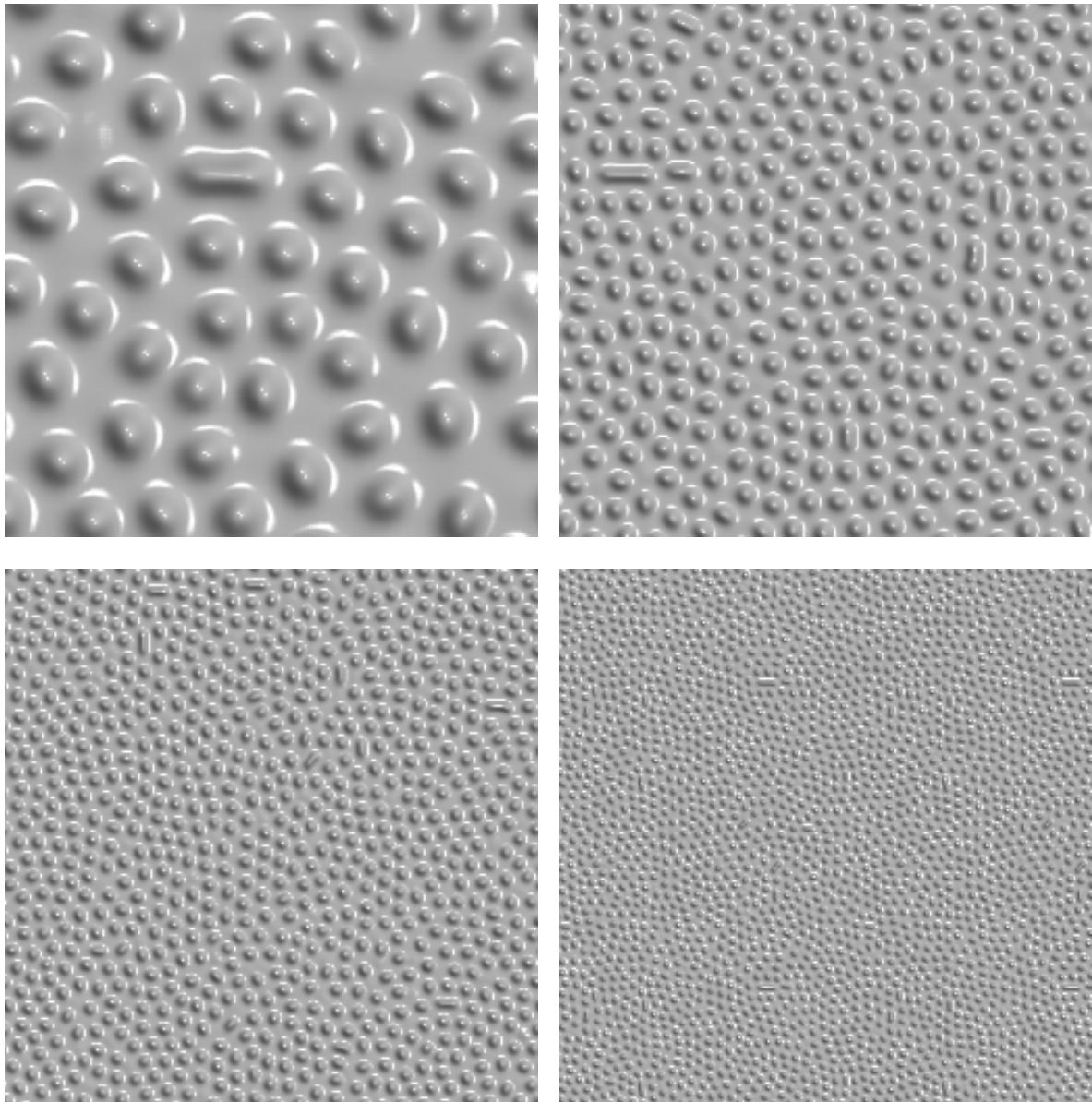


Figure 3.15: Swatches for the DDS bump-mapped layers, shown from largest to smallest. The bumps are at the scale seen by the participants. Both Pilot and Main studies use reaction-diffusion spots and bumps, shown here. The reaction-diffusion technique generates non-uniform spot shapes; this is readily seen in the image on the upper left where an oblong spot appears. All spots produced by the Gaussian generation technique described in Chapter Two are uniform.



Figure 3.16: Swatches for the DDS alpha-blended layers: pink, lavender, turquoise, and green-yellow. The spots are at the scale seen by the participants.



Figure 3.17: Swatch for the mustard DDS alpha-blended layer. The image is at the scale seen by the participants.

Dependent Measures

Participants complete three tasks. The first task is to estimate the percentage of target A that is overlapped by target B. The second task is to sketch the intersection region between targets A and B. The third task is to sketch target A – referred to below as *single target* sketch. The participants type their responses to the first question on the keyboard. Labeled, non-ruled, paper is provided for the sketches.

Three performance measures are collected. The first metric represents the error in the estimated percentage overlap, compared to the correct percentage of overlap, calculated as the absolute value of the difference. Error is defined similarly in the experiments by Healey et al. [1996], discussed at the beginning of this Chapter.

Grader0	Grader1	Count
0	0	190
0	1	121
1	0	71
1	1	1118
		1500

Table 3.8: Counts of intersection sketch scores by grader ID. The Pearson’s correlation between scores was 0.872 indicating high inter-rater reliability. Note that 192 sketches were judged wrong by one grader and exact by the other, out of 1500 total sketches – 12.8%. Also note that sketches were judged to be exact by both graders 1118 times – 74.5%.

Grader0	Grader1	Count
0	0	57
0	1	14
1	0	48
1	1	1381
		1500

Table 3.9: Counts of single target sketch scores by grader ID. The Pearson’s correlation between scores was 0.959, indicating high inter-rater reliability. Note that only 62 sketches were judged wrong by one grader and exact by the other, out of 1500 total sketches – 4%. Also note that sketches were judged to be exact by both graders 1381 times – 92.1%.

The second and third metrics are sketch scores for the intersection sketch and the single target sketch. The sketches are compared to the correct computer-derived shapes, and are graded on a binary scale of wrong: 0, correct: 1. All sketches are graded independently by two experimenters with Pearson’s correlation of 0.872 for the intersection sketches and a Pearson’s correlation 0.959 for the single target sketches. Both scores indicate high inter-rater reliability. Sketches are judged to be correct if they matched the desired shape in form, including the correct number of vertices and edges. Exact positioning and scale of the shape relative to the image boundaries is not considered. Tables 3.8 and 3.9 show scores by grader ID. Figure 3.18 shows example sketches. Overall the sketch scores are very high, with 74.5% of the intersection sketches and 92.1% of the single-target sketches receiving a grade of 1.

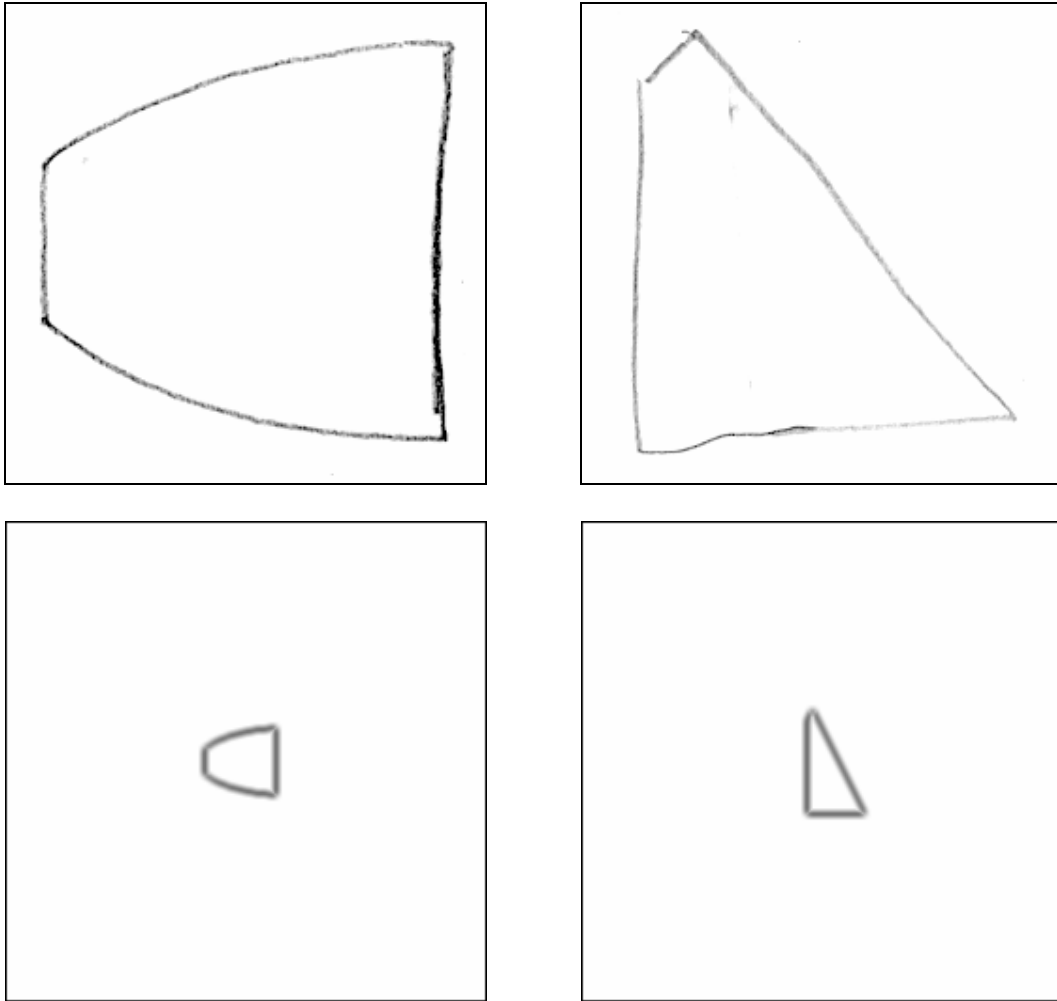


Figure 3.18: The top row shows example intersection sketches. The correct intersections are shown in the bottom row.

Procedures

Experiment schedule

Each participant completes the one session during a single visit to the lab. Each session, including instructions, lasts approximately 45-60 minutes.

The test program presents all instructions and trial images to the participant, including the instructions to read and sign a written consent form. The instructions explain the nature of the task and how to respond to the questions. Once the participants indicate they understand the instructions, the trials begin. Participants complete ten trials, followed by a break, trials resume when participants indicate (by pressing a key on the computer) that they are ready to continue.

Apparatus

Hardware Display System

Images are presented on the same monitor for all participants – a 19-inch color CRT monitor with a pixel resolution of 1600x1200, and 75-hertz non-interlaced refresh rate. Each image is 1024 by 1024 pixels, the rest of the screen contained instructions and a black background. Participants view the images while seated at a desk; no physical constraints are used to equalize participant distance from the desk.

Display Software

The display software was custom-designed for the experiment, and was implemented using OpenGL [Kempf and Frazier, editors, 1997].

Image Generation

Images were generated with the pfinan shading language [Olano and Lastra, 1998] running on the Pixel Flow graphics machine [Eyles et al., 1997].