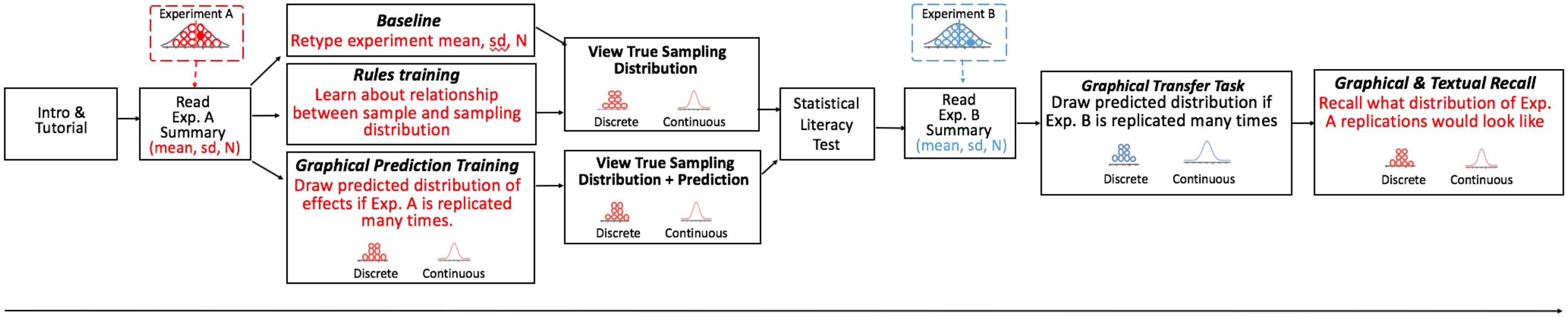
Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty

Study Interface



Intro & Tutorial

Experiment A

Read Exp. A Summary (mean, sd, N)

Baseline
Retype experiment mean, sd, N

Rules training
Learn about relationship between sample and sampling distribution

Graphical Prediction Training
Draw predicted distribution of effects if Exp. A is replicated many times.
Discrete Continuous

View True Sampling Distribution
Discrete Continuous

View True Sampling Distribution + Prediction
Discrete Continuous

Statistical Literacy Test

Experiment B

Read Exp. B Summary (mean, sd, N)

Graphical Transfer Task
Draw predicted distribution if Exp. B is replicated many times
Discrete Continuous

Graphical & Textual Recall
Recall what distribution of Exp. A replications would look like
Discrete Continuous

Intro & Tutorial
All participants are trained on how to use a graphical interface to sketch a distribution (for the graphical recall and transfer tasks).

Read About Exp. A
We assume an experiment has been replicated a large (e.g., infinite) number of times. Exp. A represents one sampled replication.

Training
All participants are assigned to one of three forms of training on how to interpret a visualized sampling distribution.

View Sampling Distribution
All participants are assigned to either a discrete or continuous visualization; those in the graphical prediction condition view the same format they used to make their prediction.

Numeracy Test
All participants take the Berlin Numeracy Test to measure statistical literacy.

Read About Exp. B
We assume a second experiment in a different domain has been replicated a large number of times. Exp. B represents one sampled replication.

Graphical Transfer
We test participants understanding of the relationship between sample data and replication uncertainty by asking them to predict the distribution of effects if Exp. B is replicated many times.

Graphical & Textual Recall
We test participants ability to recall the replication uncertainty in Exp. A by asking them to graphically and textually describe the distribution of effects if Exp. A is replicated many times.

Part 1. Read Experiment A Summary:

Read Experiment Description

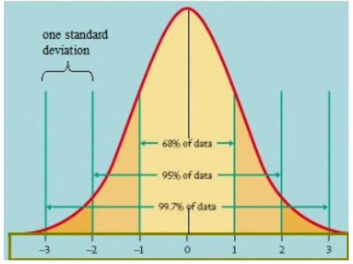
Read the experiment description below. Pay close attention the **average increase in activity and standard deviation of the increase** that the scientists observed. When you are finished reading, press Continue.

Experiment Description:

Scientists are studying how a new stimulant affects the level of physical activity of rats. They selected 40 rats for a recent experiment. The scientists monitored the activity of the rats two times over the course of two days: once after the rats had been given the stimulant, and once when the rats had been given a placebo solution (a solution that is known not to cause increased activity). For randomization, the order in which rats were given the stimulant versus the placebo was varied.

The scientists are interested in the rats' activity level. So, they used an infrared beam to measure the activity of the rats for 20 minutes. They started to measure activity 60 seconds after the rats had been given either the stimulant or the placebo. They then examined how much higher, on average, the activity score was when a rat was given the stimulant compared to when it was given the placebo. Below is their data.

The data includes the average increase in activity score that the scientists observed with the stimulant versus with the placebo. The table also presents the standard deviation, which is a measure of how much the score difference for each rat varied from one another. The standard deviation is the amount of increase such that approximately 70% percent of the rats will be within +/-1 standard deviation of the average and approximately 95% will be within +/-2 standard deviations (see the image below for an illustration).



Average Increase in Activity Score from Stimulant	268
Standard Deviation of Increase	151

Continue

Part 2a: Rules Training

Learn About Sampling Distributions

Description of a sampling distribution:

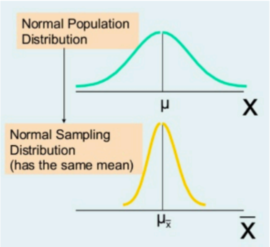
On the next page, you will view the *sampling distribution* for the experiment you read about. Imagine that scientists were to take all possible samples of some size (N) from a given population. For each sample, they compute a statistic, like the mean (average) of some variable. The sampling distribution is the probability distribution of this statistic. The sampling distribution describes the distribution of frequencies of a range of different outcomes that could possibly occur for that statistic and that population.

There is a predictable relationship between the standard deviation of the sampling distribution and the standard deviation of the observed data. (Recall from the previous page that the standard deviation is a number that quantifies the average deviation of any individual observation from the group of observations as a whole, where approximately 70% of the observations will be within +- 1 standard deviation of the mean.) The two are related based on the square root of the number of observations, or sample size, which is often abbreviated as N.

The relationship is:

Standard deviation of the sampling distribution = Data standard deviation / square root of N

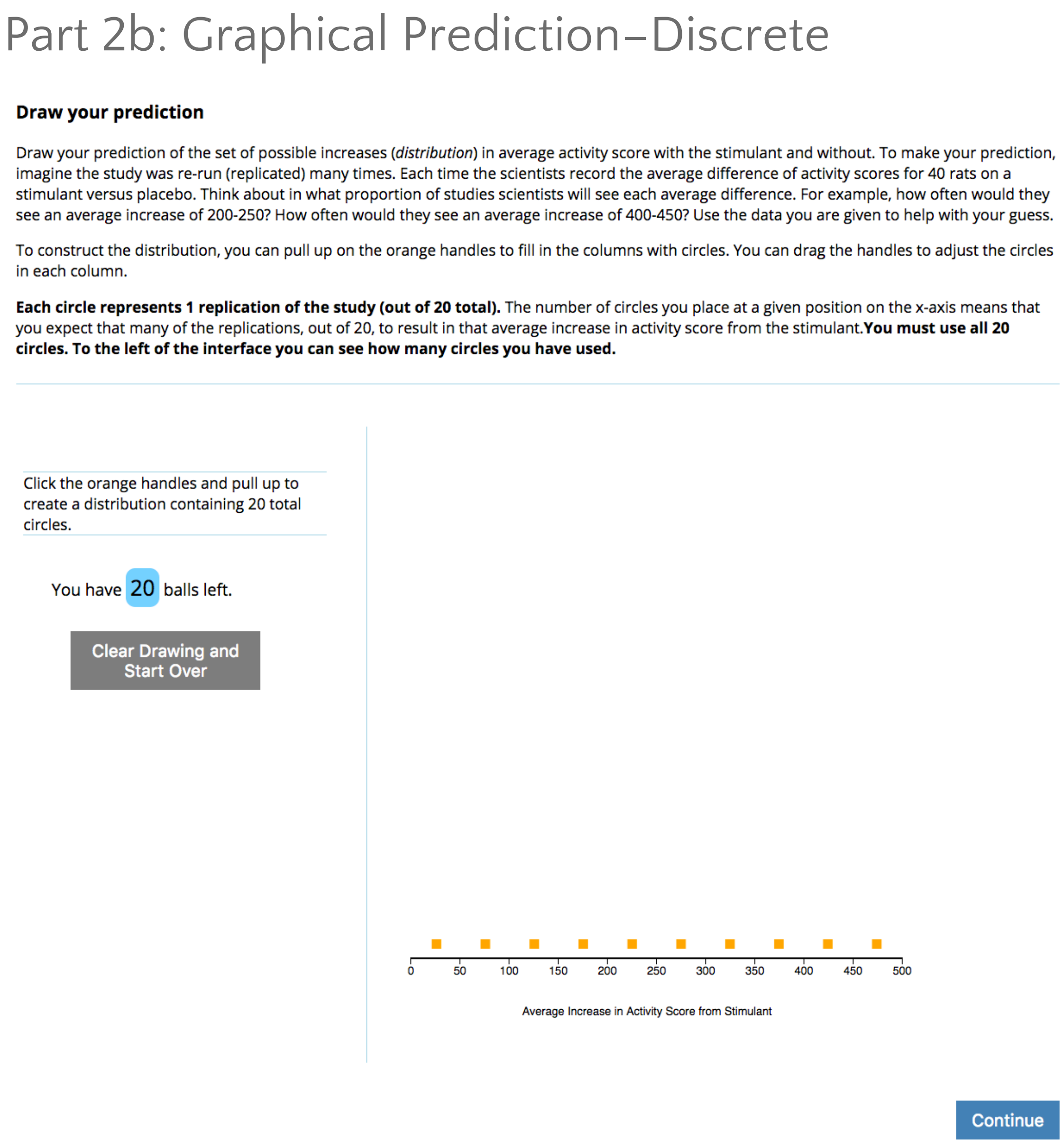
The below image helps illustrate the relationship.


Normal Population Distribution
μ
X
Normal Sampling Distribution (has the same mean)
$\mu_{\bar{x}}$
$\bar{X}$

Before continuing, calculate the standard deviation of the sampling distribution for an average of 30, observed for a sample of size 101, with a standard deviation of 12. Fill in the bottom two boxes and the form will help you with the calculation.

Standard deviation of the sampling distribution	SD/sqrt(N)
SD	Data standard deviatic
N	Number of observatio

Continue



Part 2b: Graphical Prediction–Discrete

Draw your prediction

Draw your prediction of the set of possible increases (*distribution*) in average activity score with the stimulant and without. To make your prediction, imagine the study was re-run (replicated) many times. Each time the scientists record the average difference of activity scores for 40 rats on a stimulant versus placebo. Think about in what proportion of studies scientists will see each average difference. For example, how often would they see an average increase of 200-250? How often would they see an average increase of 400-450? Use the data you are given to help with your guess.

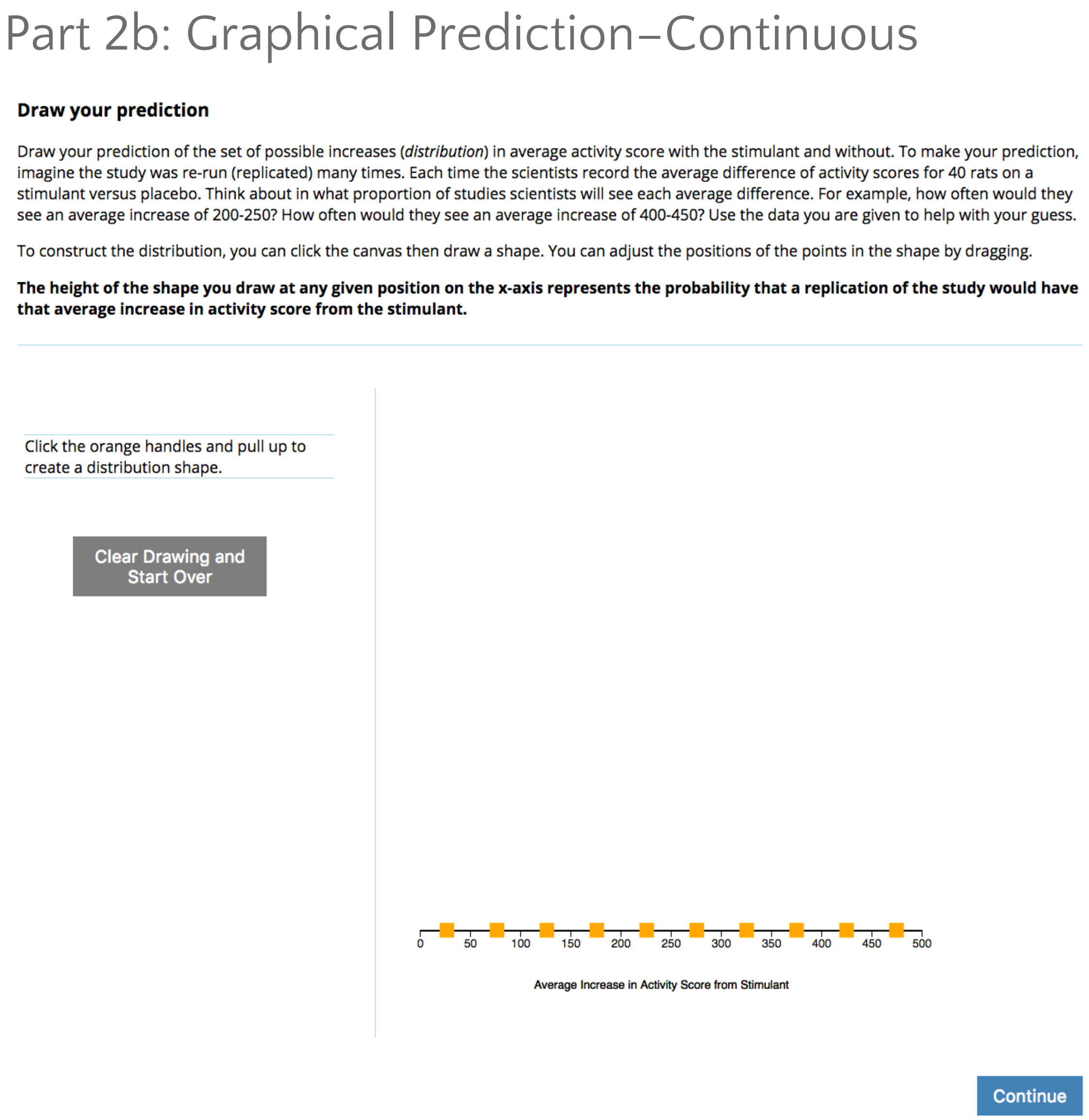
To construct the distribution, you can pull up on the orange handles to fill in the columns with circles. You can drag the handles to adjust the circles in each column.

Each circle represents 1 replication of the study (out of 20 total). The number of circles you place at a given position on the x-axis means that you expect that many of the replications, out of 20, to result in that average increase in activity score from the stimulant.**You must use all 20 circles. To the left of the interface you can see how many circles you have used.**

Click the orange handles and pull up to create a distribution containing 20 total circles.

You have 20 balls left.

Clear Drawing and Start Over
0
50
100
150
200
250
300
350
400
450
500
Average Increase in Activity Score from Stimulant
Continue



Part 2b: Graphical Prediction–Continuous

Draw your prediction

Draw your prediction of the set of possible increases (*distribution*) in average activity score with the stimulant and without. To make your prediction, imagine the study was re-run (replicated) many times. Each time the scientists record the average difference of activity scores for 40 rats on a stimulant versus placebo. Think about in what proportion of studies scientists will see each average difference. For example, how often would they see an average increase of 200-250? How often would they see an average increase of 400-450? Use the data you are given to help with your guess.

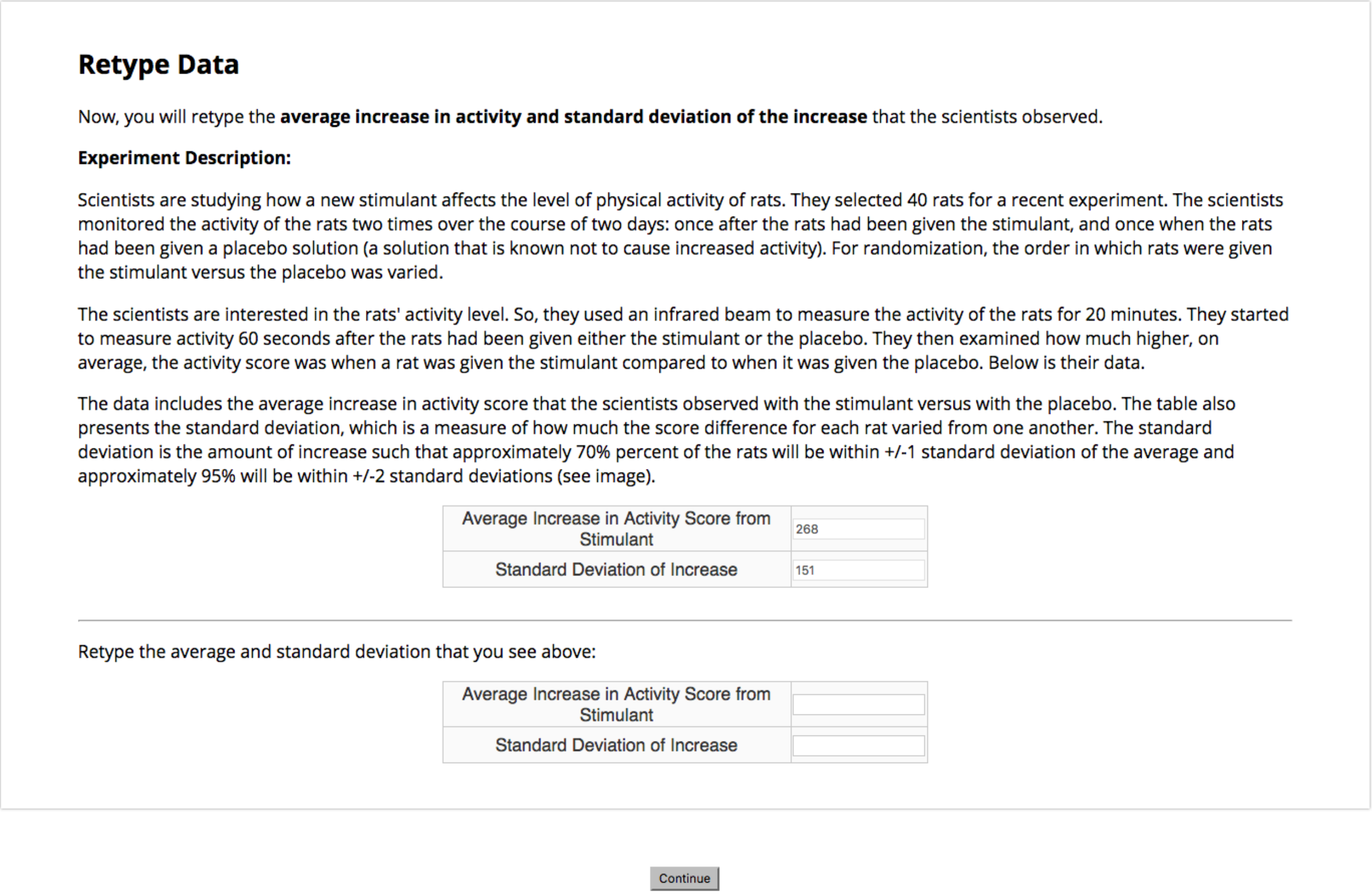
To construct the distribution, you can click the canvas then draw a shape. You can adjust the positions of the points in the shape by dragging.

The height of the shape you draw at any given position on the x-axis represents the probability that a replication of the study would have that average increase in activity score from the stimulant.

Click the orange handles and pull up to create a distribution shape.

Clear Drawing and Start Over
0
50
100
150
200
250
300
350
400
450
500
Average Increase in Activity Score from Stimulant
Continue

Part 2c: None



Retype Data

Now, you will retype the **average increase in activity and standard deviation of the increase** that the scientists observed.

Experiment Description:

Scientists are studying how a new stimulant affects the level of physical activity of rats. They selected 40 rats for a recent experiment. The scientists monitored the activity of the rats two times over the course of two days: once after the rats had been given the stimulant, and once when the rats had been given a placebo solution (a solution that is known not to cause increased activity). For randomization, the order in which rats were given the stimulant versus the placebo was varied.

The scientists are interested in the rats' activity level. So, they used an infrared beam to measure the activity of the rats for 20 minutes. They started to measure activity 60 seconds after the rats had been given either the stimulant or the placebo. They then examined how much higher, on average, the activity score was when a rat was given the stimulant compared to when it was given the placebo. Below is their data.

The data includes the average increase in activity score that the scientists observed with the stimulant versus with the placebo. The table also presents the standard deviation, which is a measure of how much the score difference for each rat varied from one another. The standard deviation is the amount of increase such that approximately 70% percent of the rats will be within +/-1 standard deviation of the average and approximately 95% will be within +/-2 standard deviations (see image).

Average Increase in Activity Score from Stimulant	268
Standard Deviation of Increase	151

Retype the average and standard deviation that you see above:

Average Increase in Activity Score from Stimulant	
Standard Deviation of Increase	

Part 3: View True Sampling Distribution–Discrete–Graphical Prediction

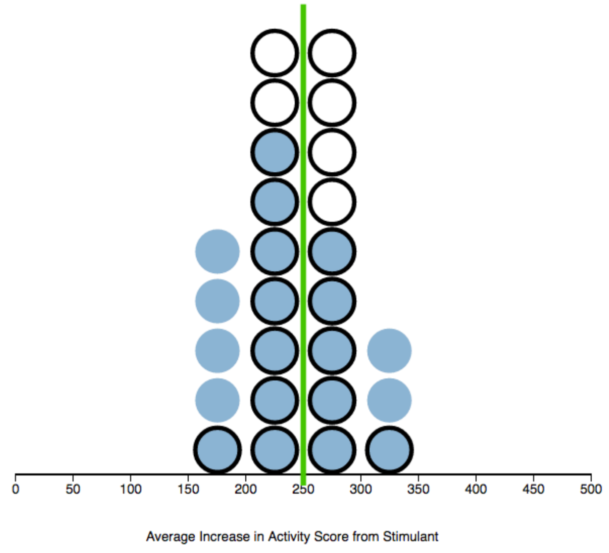
Each circle represents 1 replication of the study (out of 20 total). The number of circles stacked at a given position on the x-axis means that that many replications, out of 20 total, are likely to result in that average activity score difference.

Click the orange handles and pull up to create a distribution containing 20 total circles.

You have 0 balls left.

You're out of circles! Undo some selected circles by dragging down or click "Try Again" to start over!

Clear Drawing and Start Over



Continue

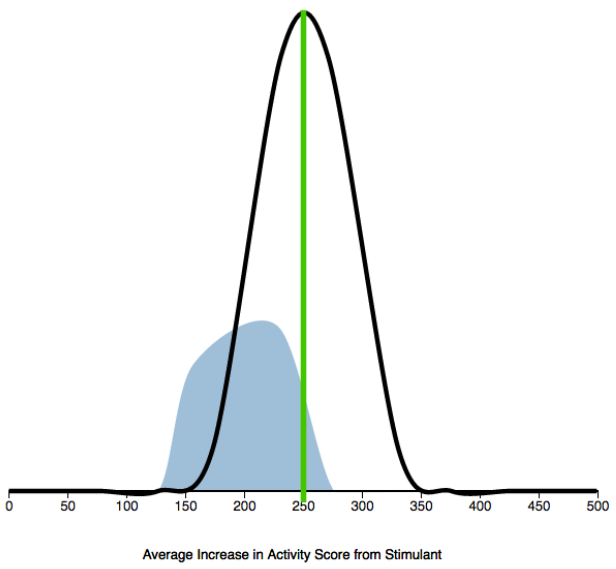
Now, examine the real distribution (the set of possible average increases that could be observed from the study), shown in black circles. Compare this true *sampling distribution for the increase* to your prediction. How did the center of your distribution compare to the center point (the average increase, shown in green) of the true distribution? How does the shape of your prediction compare to the true distribution's shape?

Part 3: View True Sampling Distribution–Continuous–Graphical Prediction

The height of the area at a given position on the x-axis depicts the probability of that average activity score difference.

Click the orange handles and pull up to create a distribution shape.

Clear Drawing and Start Over

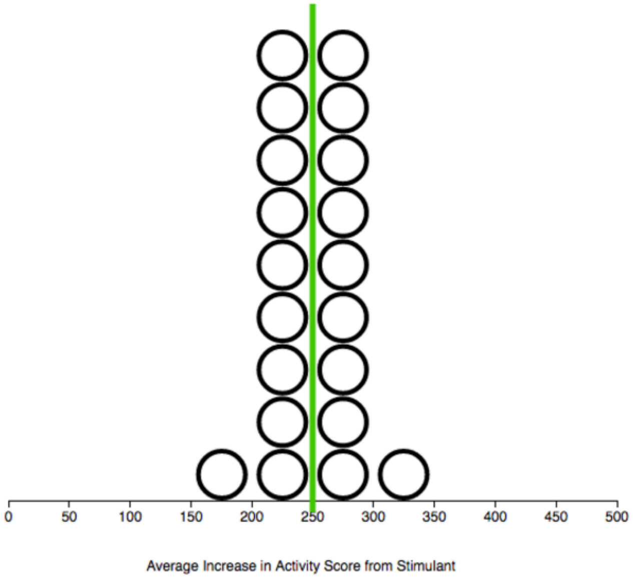


Continue

Now, examine the real distribution (the set of possible average increases that could be observed from the study), shown in black. Compare this true *sampling distribution for the increase* to your prediction. How did the center of your distribution compare to the center point (the average increase, shown in green) of the true distribution? How does the shape of your prediction compare to the true distribution's shape?

Part 3: View True Sampling Distribution–Discrete-Rules,None

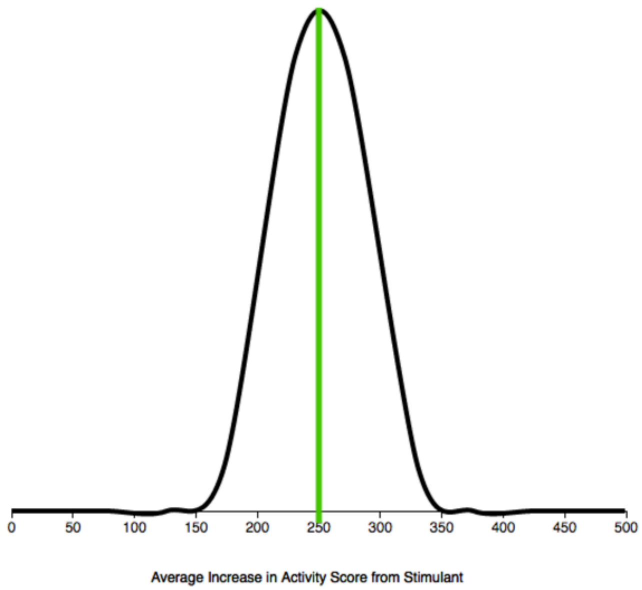
Average Increase in Activity Score from Stimulant	268
Standard Deviation of Increase	151



Now, retype the labels you see on the above graph. Type each number that appears on the graph, seperated by a comma, starting with the first number along the x-axis.

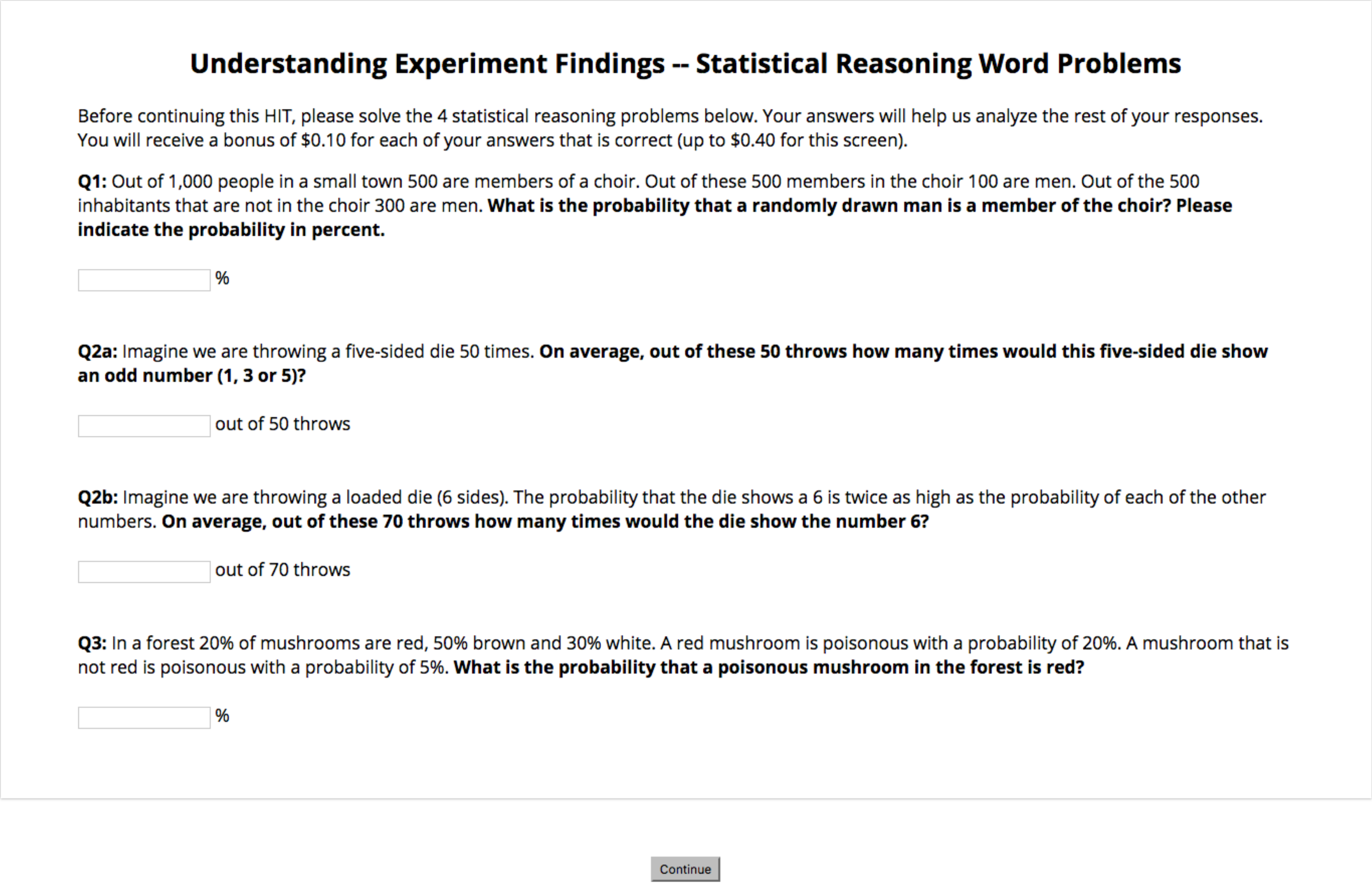
Part 3: View True Sampling Distribution–Continuous-Rules,None

Average Increase in Activity Score from Stimulant	268
Standard Deviation of Increase	151



Now, retype the labels you see on the above graph. Type each number that appears on the graph, seperated by a comma, starting with the first number along the x-axis.

Part 4: Statistical Literacy Test



Understanding Experiment Findings -- Statistical Reasoning Word Problems

Before continuing this HIT, please solve the 4 statistical reasoning problems below. Your answers will help us analyze the rest of your responses. You will receive a bonus of $0.10 for each of your answers that is correct (up to $0.40 for this screen).

Q1: Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. **What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.**

%

Q2a: Imagine we are throwing a five-sided die 50 times. **On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?**

out of 50 throws

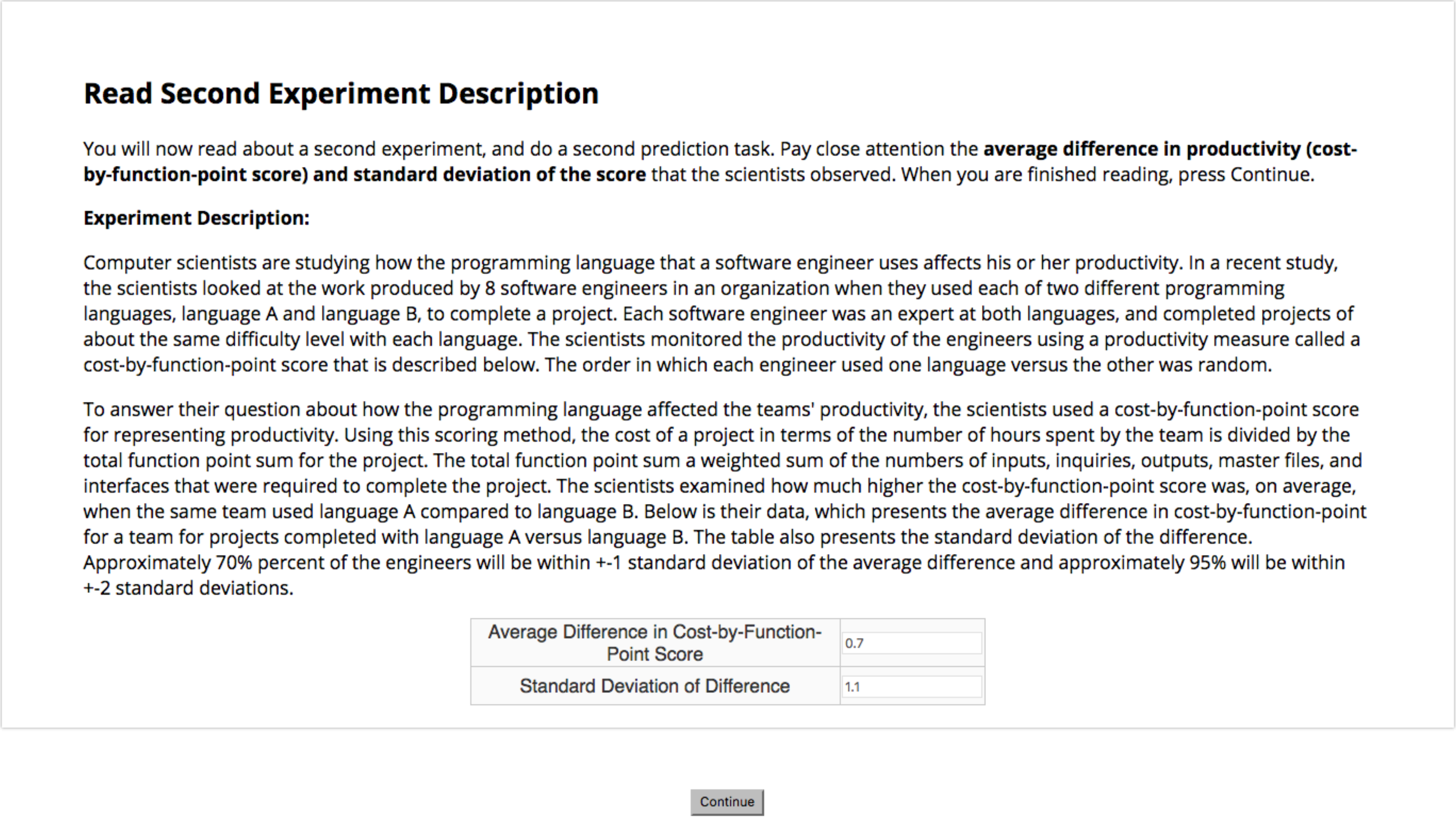
Q2b: Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. **On average, out of these 70 throws how many times would the die show the number 6?**

out of 70 throws

Q3: In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. **What is the probability that a poisonous mushroom in the forest is red?**

%

Part 5: Read Experiment B Summary



Read Second Experiment Description

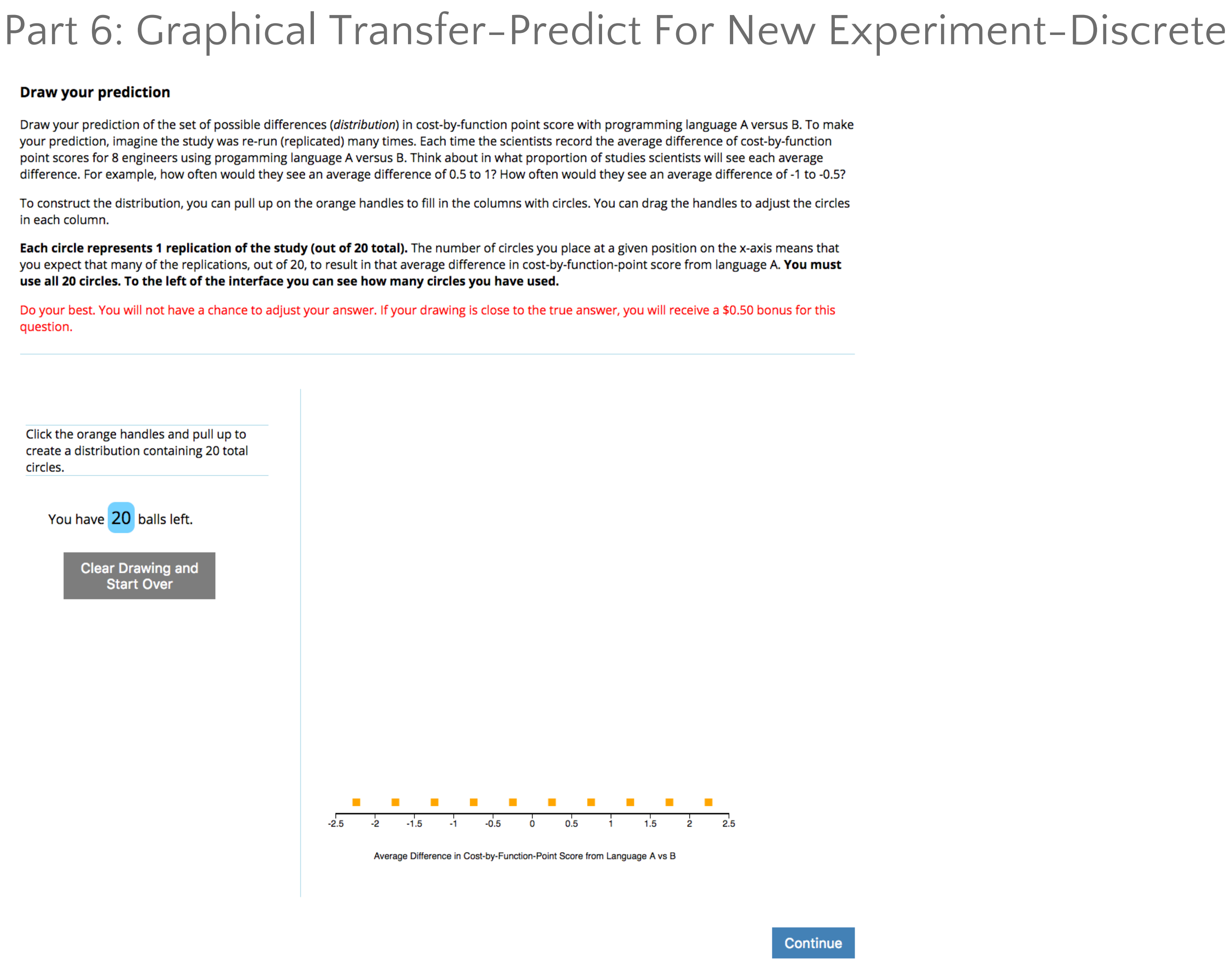
You will now read about a second experiment, and do a second prediction task. Pay close attention the **average difference in productivity (cost-by-function-point score) and standard deviation of the score** that the scientists observed. When you are finished reading, press Continue.

Experiment Description:

Computer scientists are studying how the programming language that a software engineer uses affects his or her productivity. In a recent study, the scientists looked at the work produced by 8 software engineers in an organization when they used each of two different programming languages, language A and language B, to complete a project. Each software engineer was an expert at both languages, and completed projects of about the same difficulty level with each language. The scientists monitored the productivity of the engineers using a productivity measure called a cost-by-function-point score that is described below. The order in which each engineer used one language versus the other was random.

To answer their question about how the programming language affected the teams' productivity, the scientists used a cost-by-function-point score for representing productivity. Using this scoring method, the cost of a project in terms of the number of hours spent by the team is divided by the total function point sum for the project. The total function point sum a weighted sum of the numbers of inputs, inquiries, outputs, master files, and interfaces that were required to complete the project. The scientists examined how much higher the cost-by-function-point score was, on average, when the same team used language A compared to language B. Below is their data, which presents the average difference in cost-by-function-point for a team for projects completed with language A versus language B. The table also presents the standard deviation of the difference. Approximately 70% percent of the engineers will be within +-1 standard deviation of the average difference and approximately 95% will be within +-2 standard deviations.

Average Difference in Cost-by-Function-Point Score	0.7
Standard Deviation of Difference	1.1



Part 6: Graphical Transfer-Predict For New Experiment-Discrete

Draw your prediction

Draw your prediction of the set of possible differences (*distribution*) in cost-by-function point score with programming language A versus B. To make your prediction, imagine the study was re-run (replicated) many times. Each time the scientists record the average difference of cost-by-function point scores for 8 engineers using progamming language A versus B. Think about in what proportion of studies scientists will see each average difference. For example, how often would they see an average difference of 0.5 to 1? How often would they see an average difference of -1 to -0.5?

To construct the distribution, you can pull up on the orange handles to fill in the columns with circles. You can drag the handles to adjust the circles in each column.

Each circle represents 1 replication of the study (out of 20 total). The number of circles you place at a given position on the x-axis means that you expect that many of the replications, out of 20, to result in that average difference in cost-by-function-point score from language A. **You must use all 20 circles. To the left of the interface you can see how many circles you have used.**

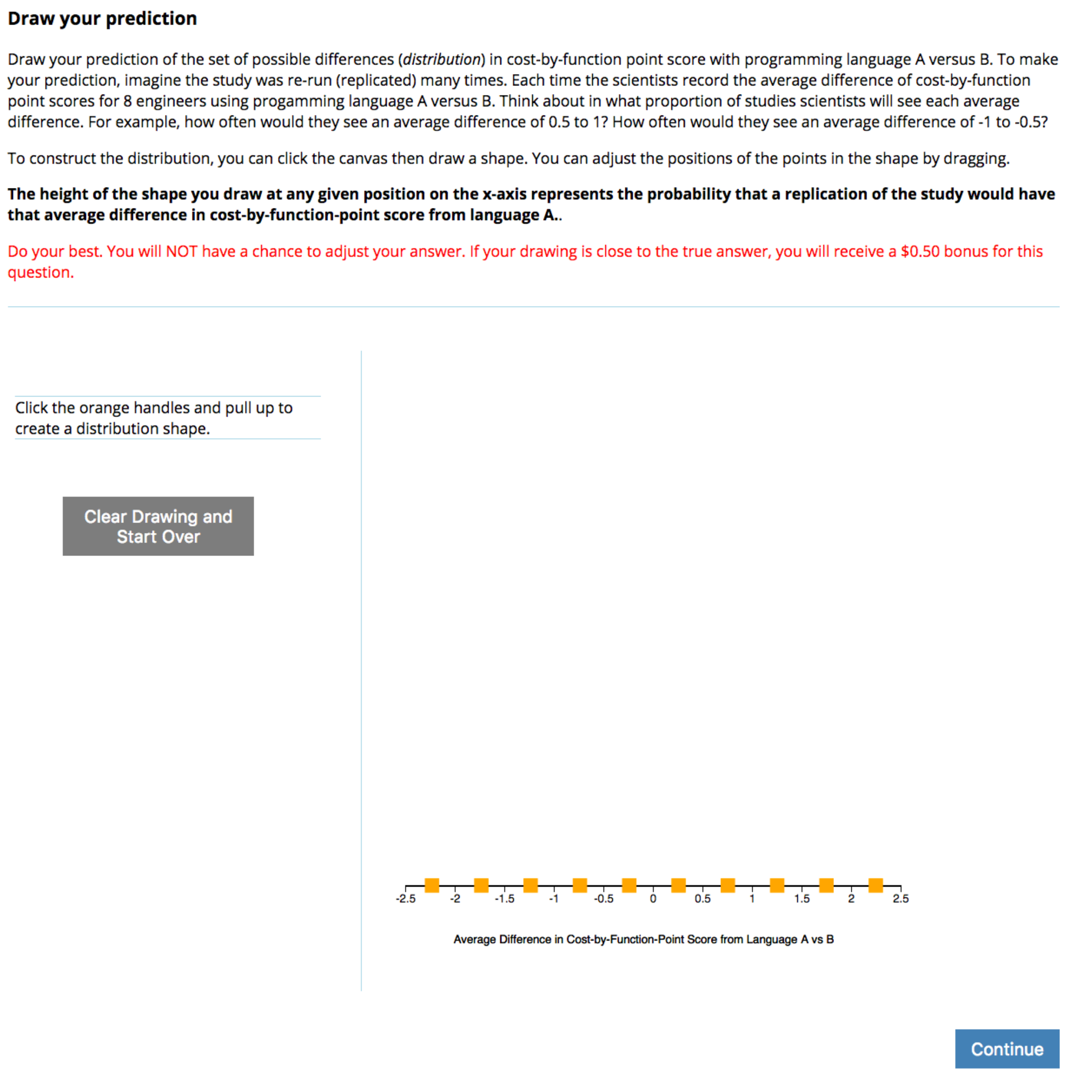
Do your best. You will not have a chance to adjust your answer. If your drawing is close to the true answer, you will receive a $0.50 bonus for this question.

Click the orange handles and pull up to create a distribution containing 20 total circles.

You have 20 balls left.

Clear Drawing and Start Over
■ ■ ■ ■ ■ ■ ■ ■ ■ ■
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
Average Difference in Cost-by-Function-Point Score from Language A vs B
Continue

Part 6: Graphical Transfer-Predict For New Experiment-Continuous



Draw your prediction

Draw your prediction of the set of possible differences (*distribution*) in cost-by-function point score with programming language A versus B. To make your prediction, imagine the study was re-run (replicated) many times. Each time the scientists record the average difference of cost-by-function point scores for 8 engineers using progamming language A versus B. Think about in what proportion of studies scientists will see each average difference. For example, how often would they see an average difference of 0.5 to 1? How often would they see an average difference of -1 to -0.5?

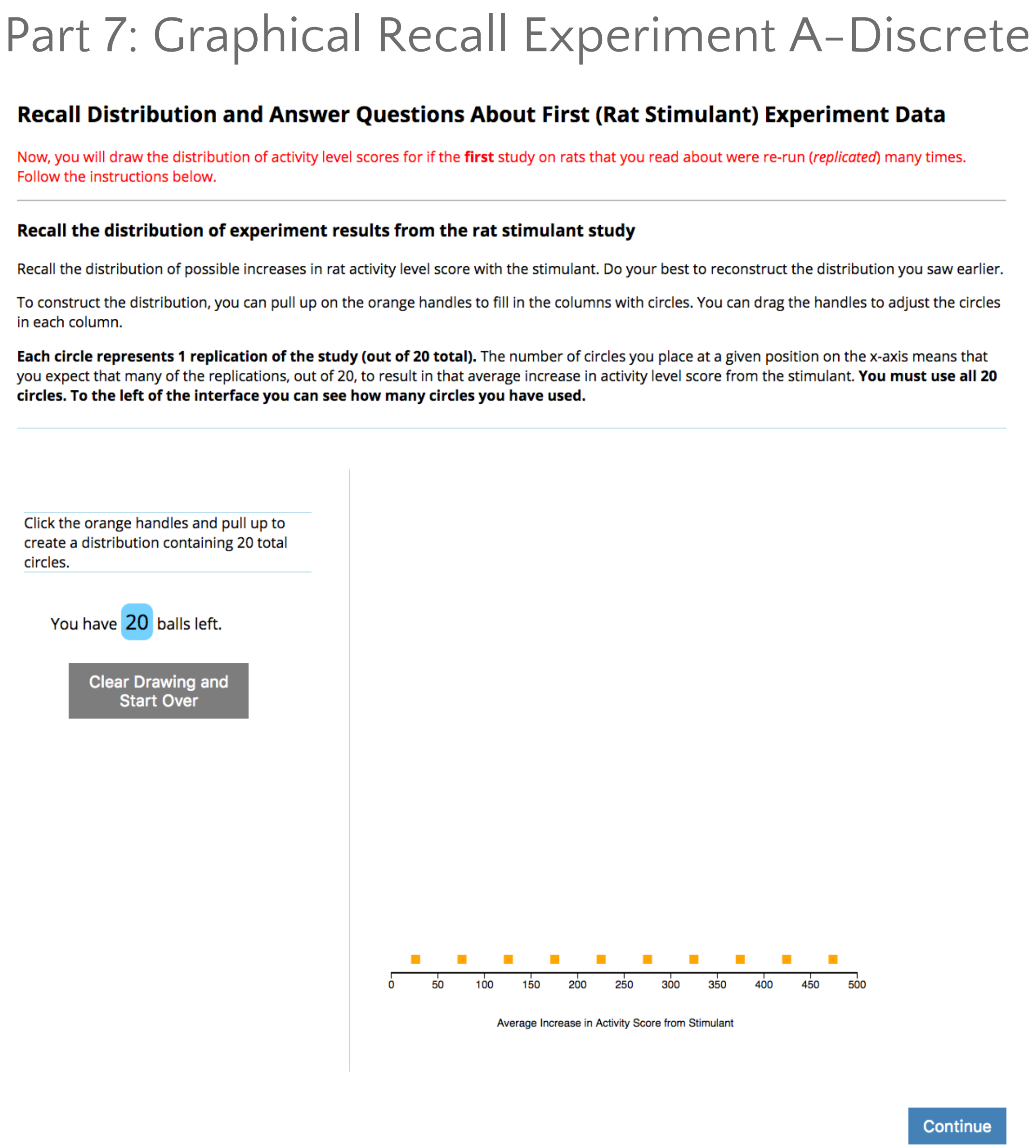
To construct the distribution, you can click the canvas then draw a shape. You can adjust the positions of the points in the shape by dragging.

The height of the shape you draw at any given position on the x-axis represents the probability that a replication of the study would have that average difference in cost-by-function-point score from language A..

Do your best. You will NOT have a chance to adjust your answer. If your drawing is close to the true answer, you will receive a $0.50 bonus for this question.

Click the orange handles and pull up to create a distribution shape.

Clear Drawing and Start Over
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
Average Difference in Cost-by-Function-Point Score from Language A vs B
Continue


Clear Drawing and Start Over
0
50
100
150
200
250
300
350
400
450
500
Average Increase in Activity Score from Stimulant
Continue

Part 7: Graphical Recall Experiment A-Discrete

Recall Distribution and Answer Questions About First (Rat Stimulant) Experiment Data

Now, you will draw the distribution of activity level scores for if the **first** study on rats that you read about were re-run (*replicated*) many times. Follow the instructions below.

Recall the distribution of experiment results from the rat stimulant study

Recall the distribution of possible increases in rat activity level score with the stimulant. Do your best to reconstruct the distribution you saw earlier.

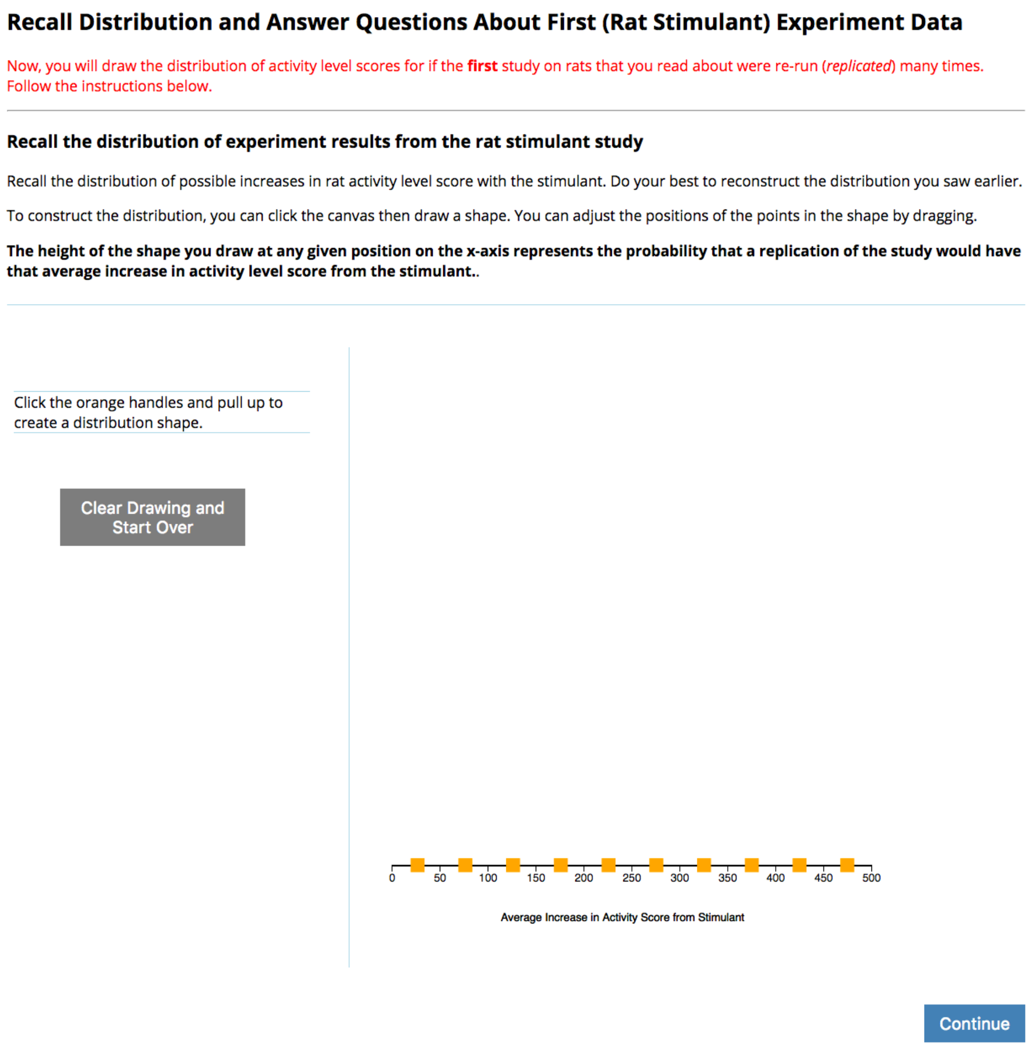
To construct the distribution, you can pull up on the orange handles to fill in the columns with circles. You can drag the handles to adjust the circles in each column.

Each circle represents 1 replication of the study (out of 20 total). The number of circles you place at a given position on the x-axis means that you expect that many of the replications, out of 20, to result in that average increase in activity level score from the stimulant. **You must use all 20 circles. To the left of the interface you can see how many circles you have used.**

Click the orange handles and pull up to create a distribution containing 20 total circles.

You have 20 balls left.

Part 7: Graphical Recall Experiment A–Continuous



Recall Distribution and Answer Questions About First (Rat Stimulant) Experiment Data

Now, you will draw the distribution of activity level scores for if the **first** study on rats that you read about were re-run (*replicated*) many times. Follow the instructions below.

Recall the distribution of experiment results from the rat stimulant study

Recall the distribution of possible increases in rat activity level score with the stimulant. Do your best to reconstruct the distribution you saw earlier.

To construct the distribution, you can click the canvas then draw a shape. You can adjust the positions of the points in the shape by dragging.

The height of the shape you draw at any given position on the x-axis represents the probability that a replication of the study would have that average increase in activity level score from the stimulant..

Click the orange handles and pull up to create a distribution shape.

Clear Drawing and Start Over
0
50
100
150
200
250
300
350
400
450
500
Average Increase in Activity Score from Stimulant
Continue

Part 7: Textual Recall Experiment A

Use your drawing to answer the probability questions. You can adjust your drawing if you want to. You will receive a bonus of $0.10 for every question that you get correct or close to correct (up to $0.80 for the set of questions).

Q: If the study were repeated many, many times, what percentage of replications would observe an average activity score difference of **between 50 and 400?**

□ out of 100

Q: If the study were repeated many, many times, what percentage of replications would observe an average activity score difference of **less than 400?**

□ out of 100

Q: If the study were repeated many, many times, what percentage of replications would observe an average activity score difference of **less than 200?**

□ out of 100

Q: If the study were repeated many, many times, what percentage of replications would observe an average activity score difference of **of approximately 400?**

□ out of 100

Q: If the study were repeated many, many times, what percentage of replications would observe an average activity score difference of **of between 300 and 350?**

□ out of 100

Q: If the study were repeated many, many times, what percentage of replications would observe an average activity score difference of **of between 200 and 250?**

□ out of 100

Q: If the study were repeated many, many times, what percentage of replications would observe an average difference of **between 200 and 300?**

□ out of 100

Q: If the study were repeated many, many times, what percentage of replications would observe an average activity score difference of **less than 100?**

□ out of 100