The Logic of Illusion

Cara Collins

Helen Tyson Middle School

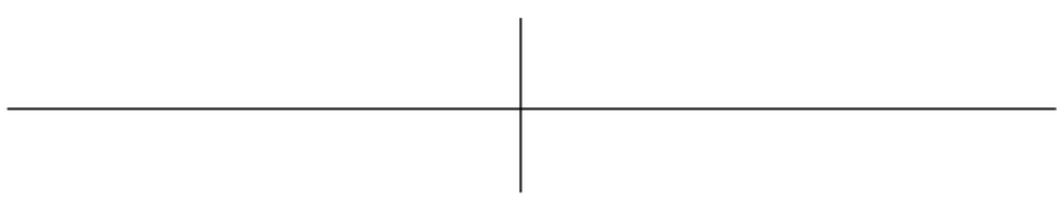
The logic of making statistical decisions is counter-intuitive for my students. It relies on a test of the hypothesis that observed variability is due to chance alone and not cause + chance. The eye illusion experiment is a good opportunity to help students understand this logic.

Foundations

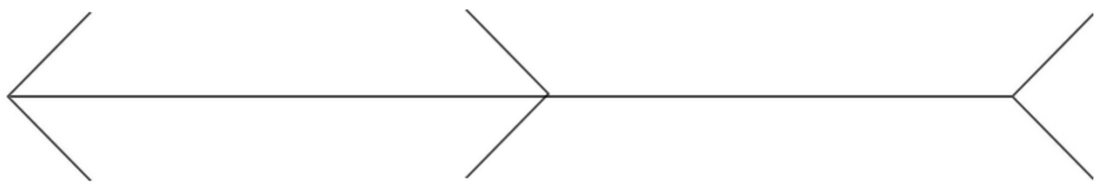
Before engaging in the logic of making statistical decisions, students should understand how to build models of a chance process, as developed in Unit 6 for the repeated measurement process. Students should also understand that good models generate simulated samples with characteristics that are consistent, but not identical, with the real-world measurements. The key here is that when a student sees a sample generated by a model, the student knows that this sample, even if it matches characteristics of the real-world sample, is only one sample. As my students like to say of any particular simulated sample, "maybe you just got lucky." To get a better sense of model fit, repeated runs of the model are necessary, and in Unit 6, a sampling distribution of the model's statistics, usually its median and IQR, show the sample-to-sample variability that the model predicts. We can use these sampling distributions to decide whether or not our model tends to fit (but not copy) our real world data. Sometimes when we look at our sampling distribution, we find that the first one or two simulated samples that we thought corresponded well with our sample were indeed relative outliers in the sampling distribution—so we may need to adjust our model.

Setting Up the Eye Illusion as a Game

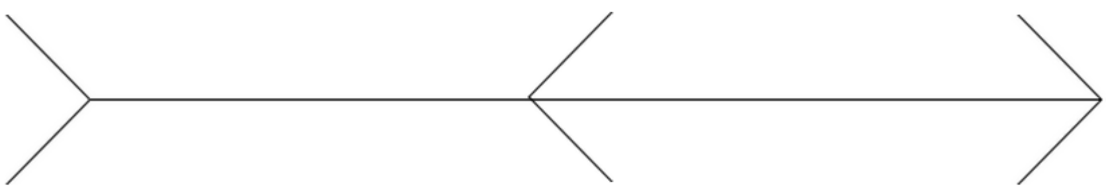
I set the experiment up as a game, emphasizing the goal of marking the mid-point. I tell them that in the first condition (see below), they get to make the mark without any modification to the line segment.. I elicit predictions about whether or not everyone will mark the exact center. Usually students predict variability about the target, as they have seen before. Students know that their task is to drag the vertical line segment to the exact center of the line.



Then we consider the second and third conditions of the experiment. I mention that now arrows are placed on the line, with the intention of fostering an illusion. When the arrows face inward and we move the second arrow to mark the line, our eyes tell us that the line segment between the inner arrows is shorter than it really is, so we tend to over-compensate and move the arrow further to the right. This tendency is to over-shoot or over-estimate the target. I ask students to consider whether or not they think they can overcome this illusion. They usually think they can.



In the other condition, the arrows face outward. Now the tendency is to believe that the line segment from the left to middle is longer than it really is. So, when we mark the center, we tend to under-estimate or under-shoot the target. Students usually believe they can overcome this illusion too.



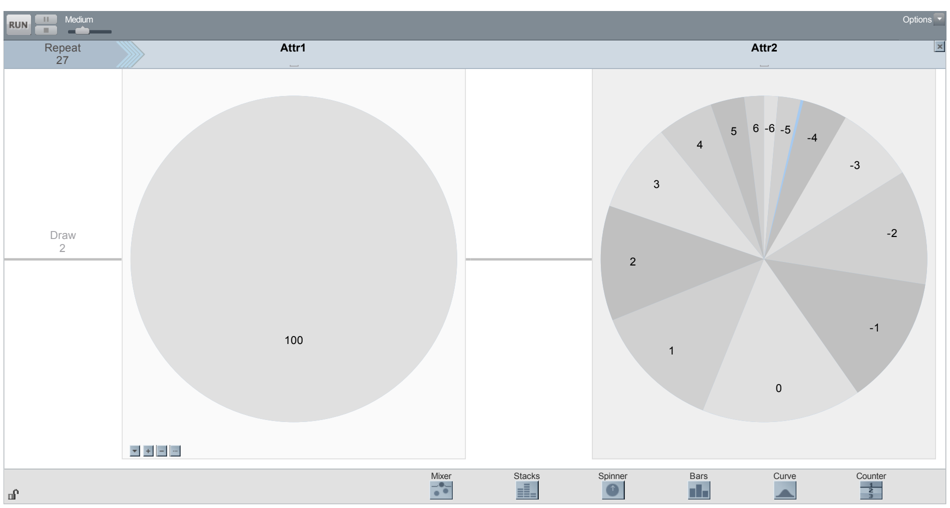
Now comes the critical part. I ask how we can tell whether or not someone overcame the illusion. Overcoming the illusion means that the class behaves the same way in conditions two and three as it does in condition one. The critical insight is that we expect individual estimates to vary just by chance, so if we overcome the illusion, then we behave just as we would expect with chance variation.. We discuss that we might even expect the marks of center to be similar for everyone in the sample, but we would not expect every individual mark to be identical. Hence, we might expect that a statistic of center will inform us about the target value, and a statistic of variability will inform us about our tendency to cluster about that target value.

Then I let students try out the experiment on line and I select only the cases from the first, no arrow condition. I ask students to use the data to make an inference about the target value (100 mm) and to measure our consistency in meeting this target. I remind them that we know of statistics that will help us (generally, I encourage them to use the median and the IQR). Then I ask them to build models to reproduce these key characteristics of our class sample, reminding them that we want to fit our data while keeping in mind that our estimates are only one sample and there could be many, many samples (and in fact are if you look at the website's database).

Modeling Individual Difference Variation

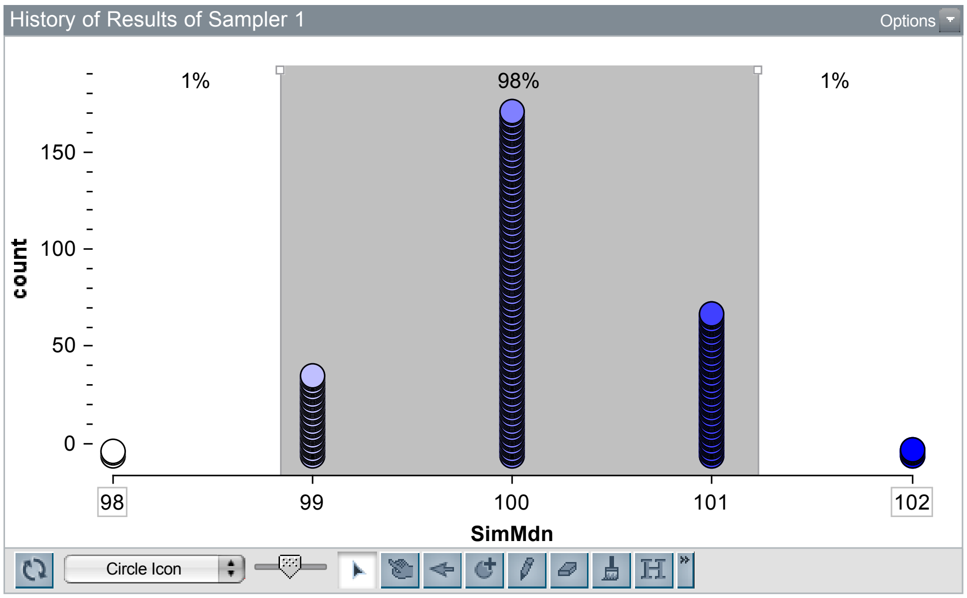
Students usually build one of two kinds of models. One kind uses the data of the sample and resamples those data with replacement. A variation on this model represent the errors as the difference between each estimate and the median, so that the model has the familiar form of signal (the median) and error, but still relies on resampling. We talk about models like these as reasonable approximations or representations when we do not know or understand very well the sources of error. We believe that chance is at play, but unlike modeling repeated measures, we have a hard time deciding on the nature of the sources of error (error here is the difference between the target value and the mark made by each individual).

A second kind of model also represents signal and noise, but these students typically reason that small errors are much more likely than large errors. And, because they have in mind that our class is only one sample, there are other "possible values" that should be generated by the model that cannot be generated by the re-sampling model. For example, in the model below, values such as 106 were possible in the model but were not observed in our class sample.



Models and Inference

Students generate a sampling distribution to decide whether or not their model is a good representation of the process and of our data. I ask students to use the divider tool in TinkerPlots to decide on what values of the sample medians their model suggest are likely just by chance, and what values of the sample medians could happen by chance but are very unlikely. For example, the sampling distribution of the sample medians indicated by the previous model for a sample size of 27 is displayed below. It indicates that sample medians between 99 and 101 are very likely and expected given the model of errors that happen just by chance. In contrast, values of 98 or 102 are very unlikely, when all that is happening is random error.



Some students may suggest that a sample median of 98 or 102 "is only 2 away," so they may need to be reminded that the consideration is how likely, not how far away. Students then consider other values not in the sampling distribution, such as 97 or 103. I suggest that if we ran the model more times, generating perhaps 10,000 samples instead of 300 (or however many times we ran the model), we might see these values but they would be even more unlikely.

Before looking at the data for the other two conditions, I ask:

"If the arrows had no effect on our ability to mark the midpoint of the line, then we overcame the illusion. If that was the case, what are some values of the sample's median would you expect to find just by chance? What are values of the median that would indicate that the illusion *did* work, and we were *NOT* able to over come it?" With these in mind, we then look at our data in the other two conditions. Students are usually disappointed to find that they too responded to the illusion, despite their best efforts. The sample medians are well outside of the range of chance. I emphasize that we can't tell for sure. Maybe it happens by chance 1 billionth of the time, but it seems highly unlikely that it is just chance at play. Our eyes can indeed deceive us!

Because the model of target + chance did not describe our responses in the other two conditions, students adjust the chance model so that it models these other conditions, by adding or subtracting a fixed term to represent the bias introduced by the illusion. We complete our discussion by comparing the models, again with the intention of clarifying that if there was no illusion, the first model would fit the data well for all three conditions. Because it does not, we infer that something other than pure chance is at work—and so we have to develop models that include other effects—here the effect of an illusion, which introduces a bias to our judgments. I remind them that previously, when we measured the length of an attribute, we distinguished between accuracy and precision. The illusion influences accuracy of judgment. The sampling distribution of the first condition model IQR or average deviation might be used to address whether or not the illusion also affects the precision or consistency of the judgments.