Teaching Introductory Statistics Using Simulation-Based Inference (SBI)

AMATYC 2018

Barbara Dolansky

Dutchess Community College

Poughkeepsie, NY
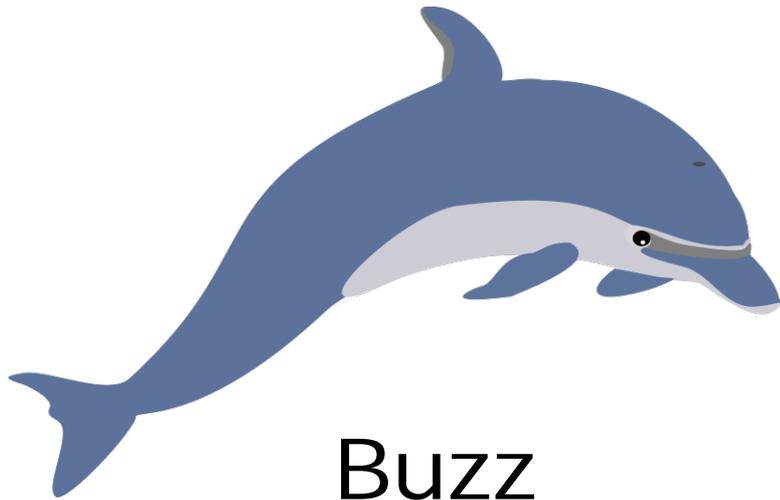
Can Dolphins Communicate?

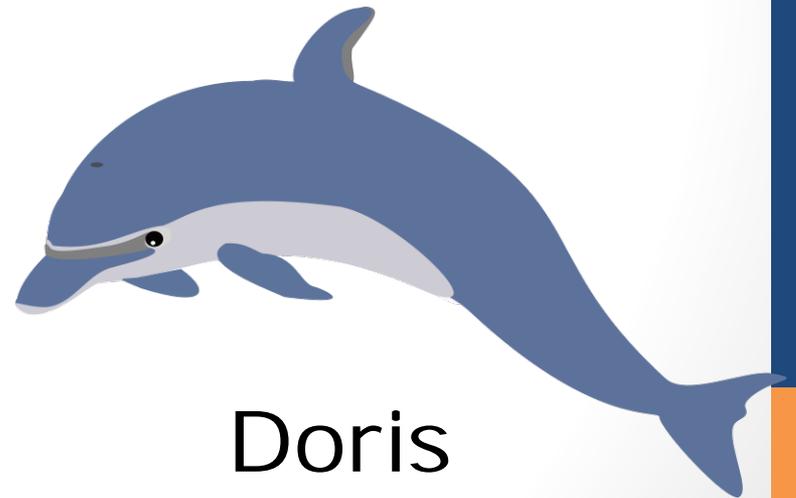# U.S. Navy Marine Mammal Program

The dolphin communication study that we will be looking at was done under a contract with the Navy in the 1960s. The Navy still has a marine mammal program.

- Dolphins and sea lions
- Mine detection
- Equipment recovery
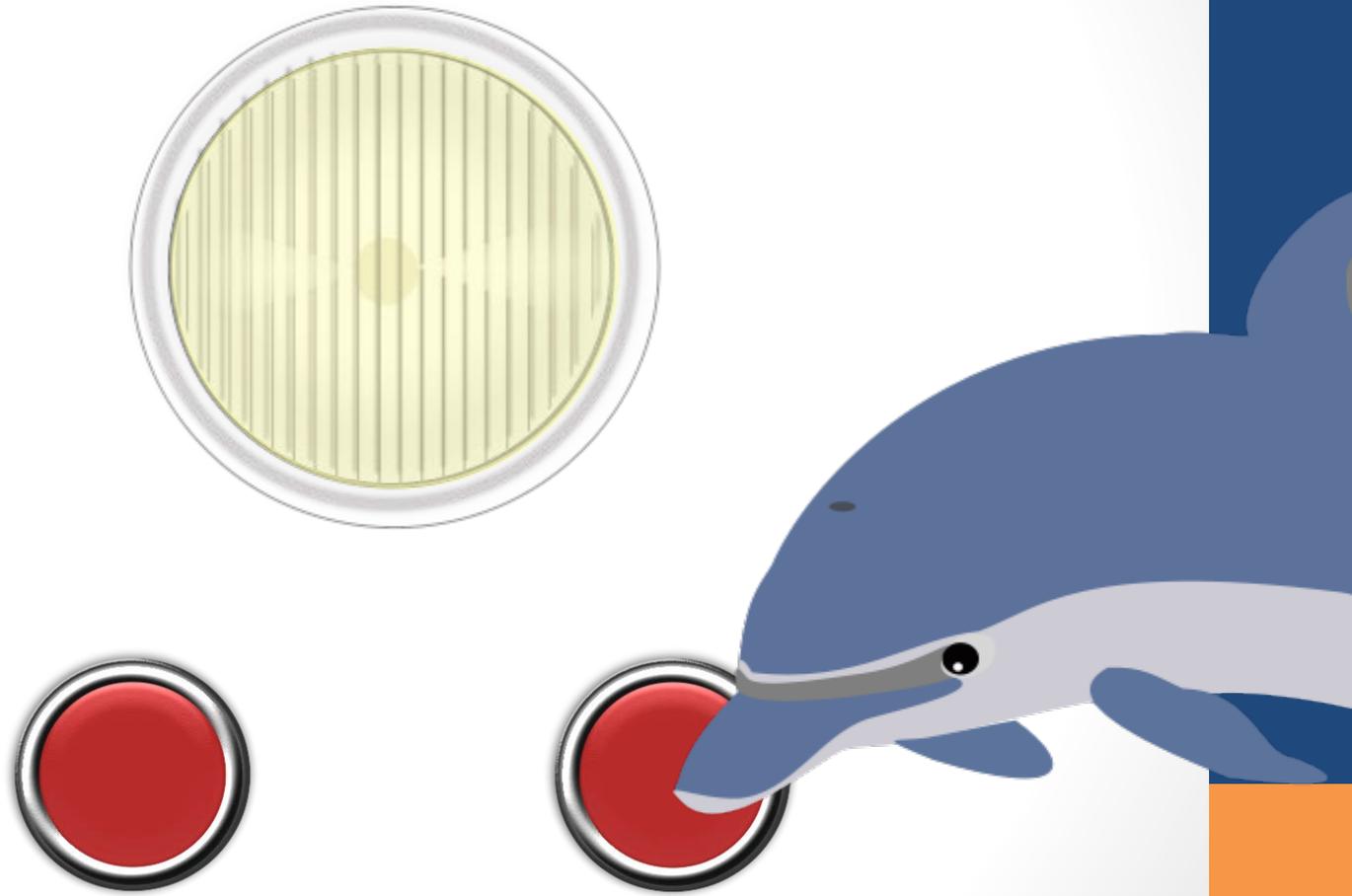
Can dolphins communicate abstract ideas?
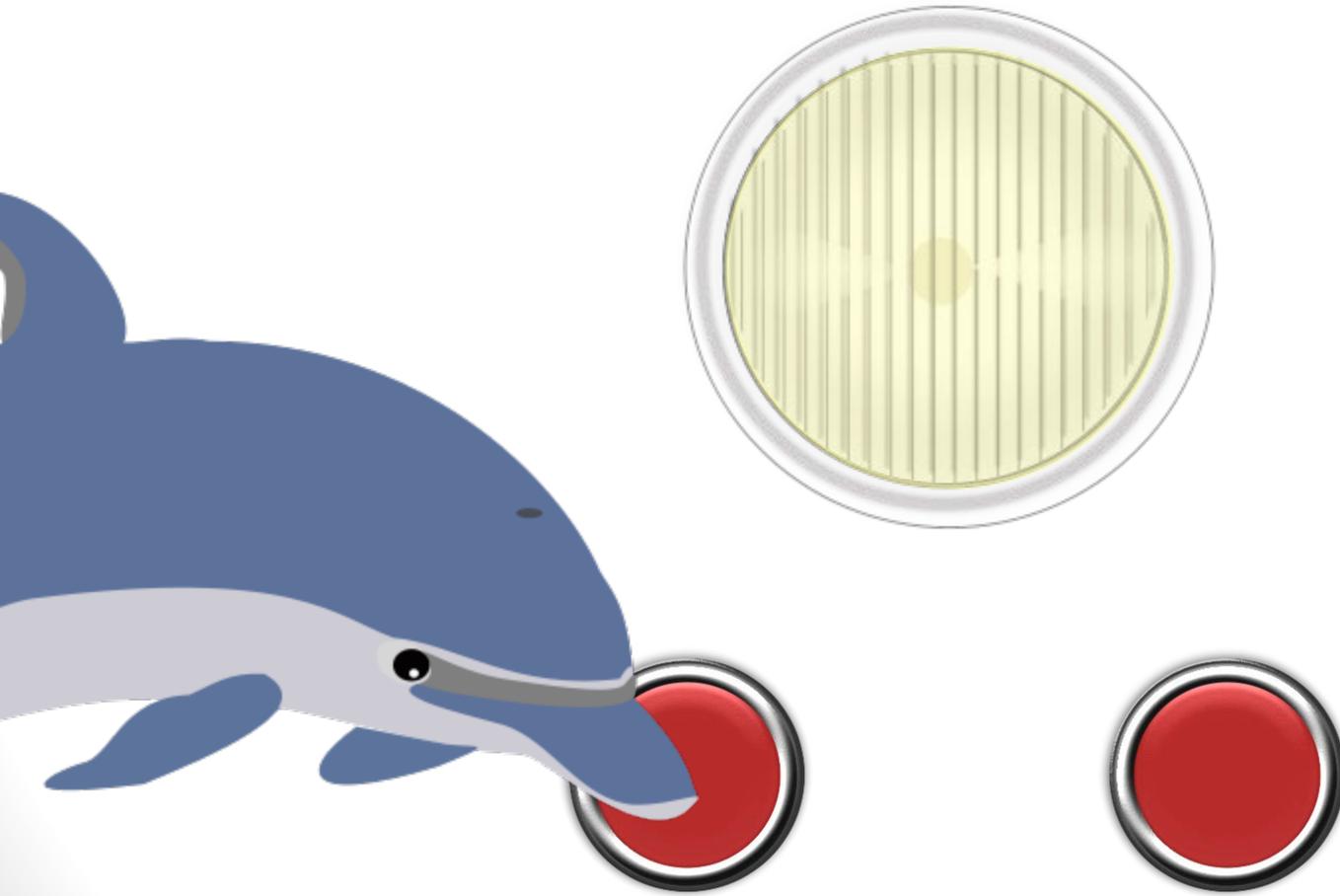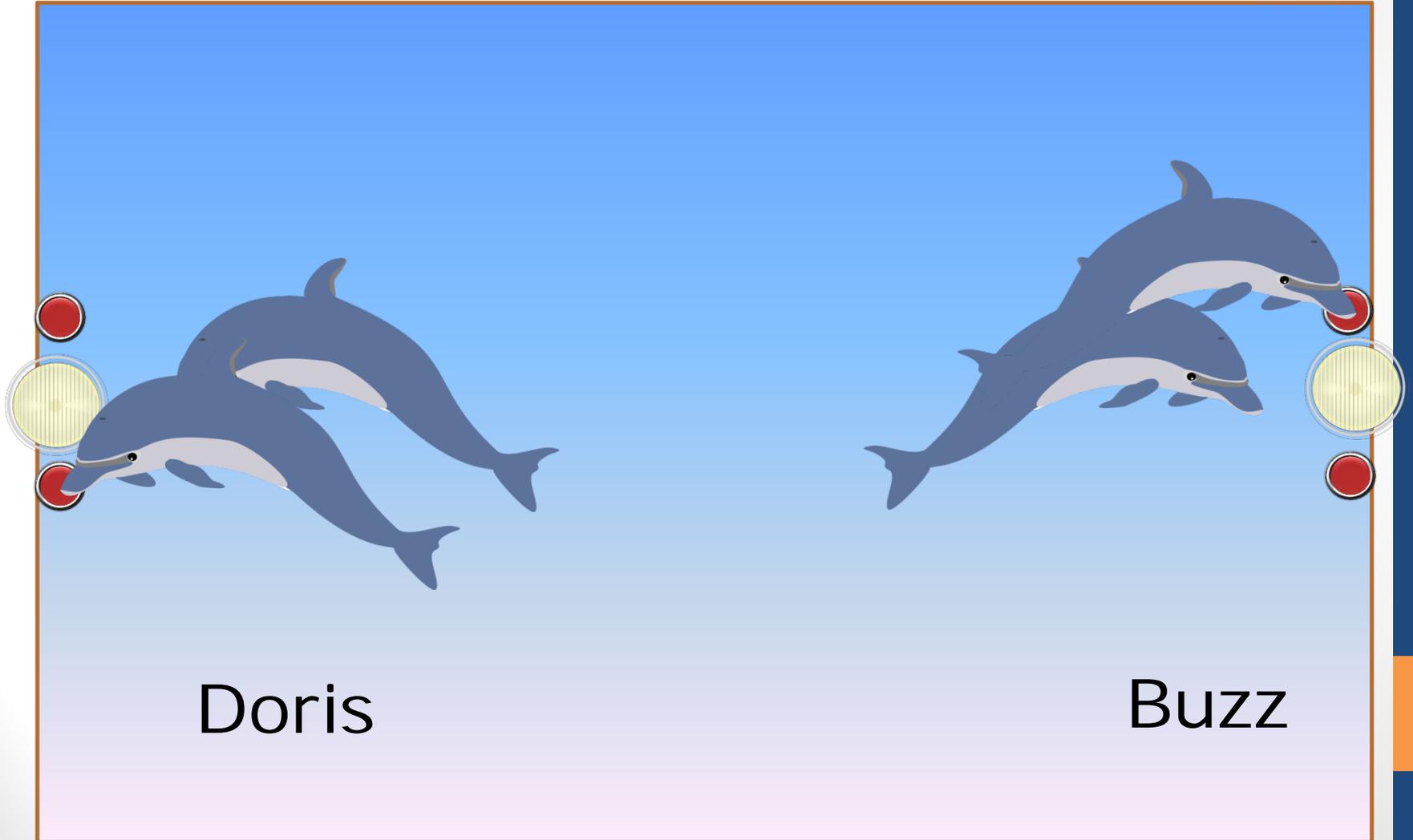
Buzz

Doris

Step 1: Learn the Signals

Step 1: Learn the Signals

Step 2: Learn the Order

Doris

Buzz

Step 3: Communicate!
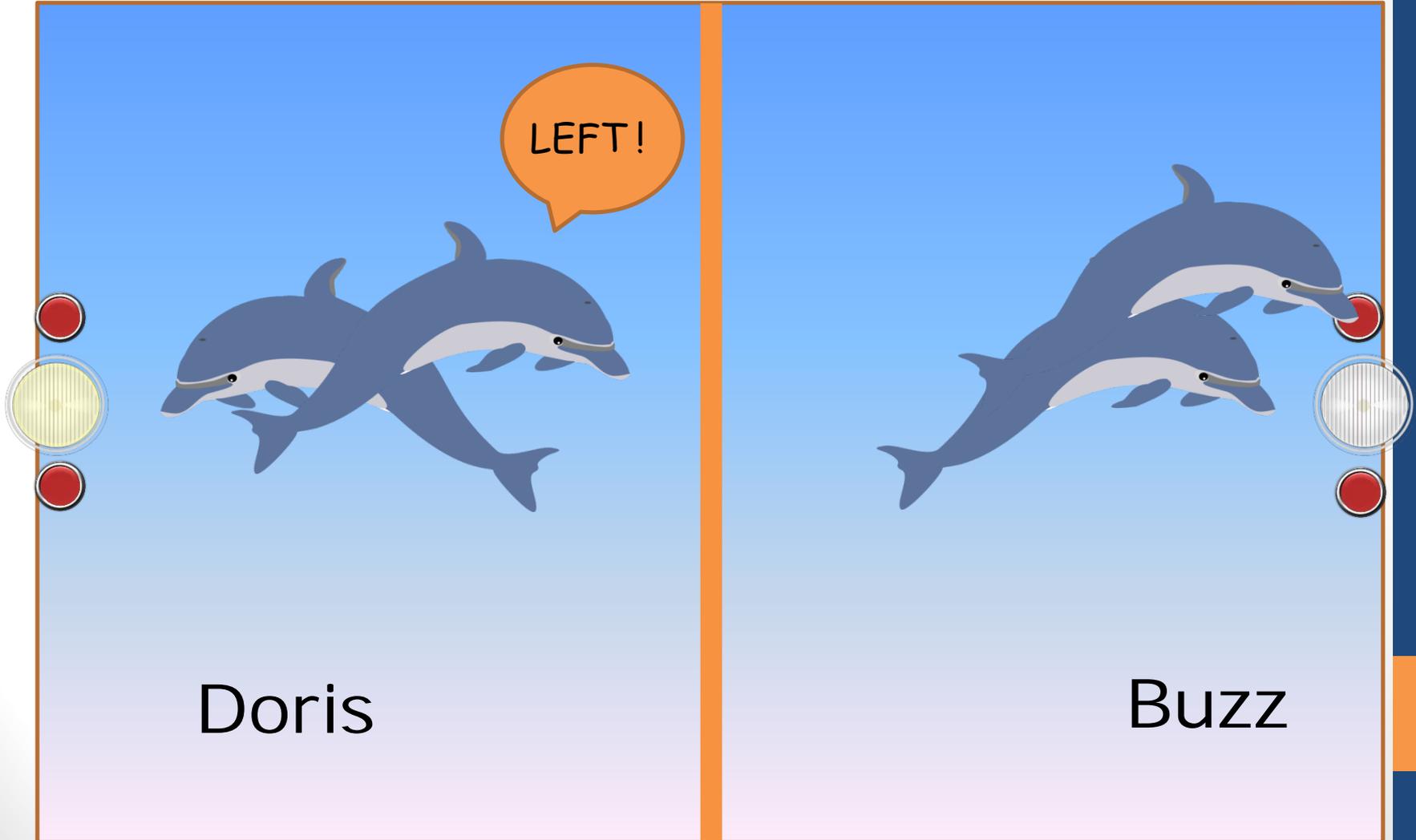
The results

- In one set of trials, Buzz chose the correct button 15 out of 16 times

- Based on the results, do you think Buzz knew which button to push?

Vocabulary

- What are the **observational units** in the Buzz and Doris experiment? How many are there? This is called the **sample size**

- What is the **variable**? Is it categorical or quantitative?

- The **statistic** is the proportion of times Buzz pushed the correct button (15/16)

- The **parameter** is Buzz's long-run probability of choosing the correct button (unknown)

Possible Explanations

- There are two possible reasons why Buzz chose the correct button so many times
  - He is just randomly guessing and got 15 out of 16 correct just by chance
  - He was doing something other than just guessing and was understanding what Doris was telling him
- We want to model the random guessing and see how Buzz's result fits in this model
- How might we model the situation where Buzz is just guessing (our **chance model**)?

Simulation vs. Real Study

TABLE 1.1 Parallels between real study and physical simulation		
Coin flip	=	guess by Buzz
Heads	=	correct guess
Tails	=	wrong guess
Chance of heads	=	1/2, probability of correct button when Buzz is just guessing
One repetition	=	one set of 16 simulated attempts by Buzz

Modeling Doris and Buzz

- Flip a coin 16 times and record the number of heads in 16 tosses

- Enter results at https://www.statcrunch.com/5.0/survey.php?surveyid=21917&code=OOURI

Modeling Doris and Buzz

- We can use an applet to simulate many more sets of 16 coin tosses
- Go to http://www.rossmanchance.com/ISIapplets.html
- Open the **One Proportion** applet

Three S Strategy

- **Statistic:** Compute the statistic from the observed data

- **Simulate**: Identify a model that represents a chance explanation; form a distribution by repeatedly simulating values of the statistic that could have happened when the chance model is true

- **Strength of evidence**: Consider whether the value of the observed statistic is likely to occur when the chance model is true

Three S Strategy

- **Statistic:** Compute the statistic from the observed data

- Buzz pressed the correct button 15/16 times, so the statistic is 15/16 = 0.9375

Three S Strategy

- **Simulate**: Using the **One Proportion** applet we can simulate 1000 sets of 16 tosses to get a sense of what we might expect if Buzz is just guessing

Mean = 0.505
SD = 0.124

Proportion of heads
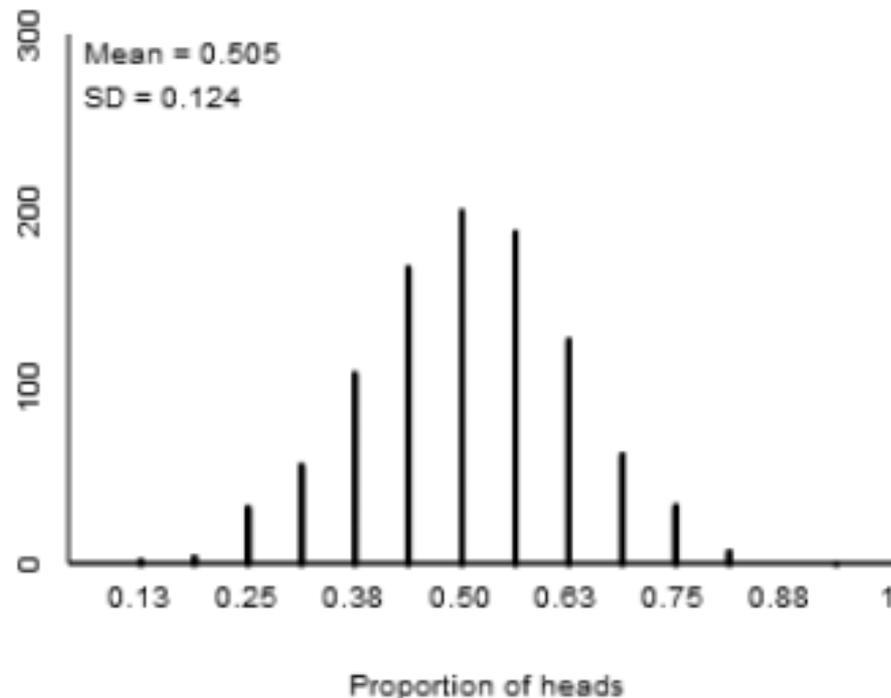
Three S Strategy

- **Strength of evidence**: Consider whether the value of the observed statistic is likely to occur when the chance model is true
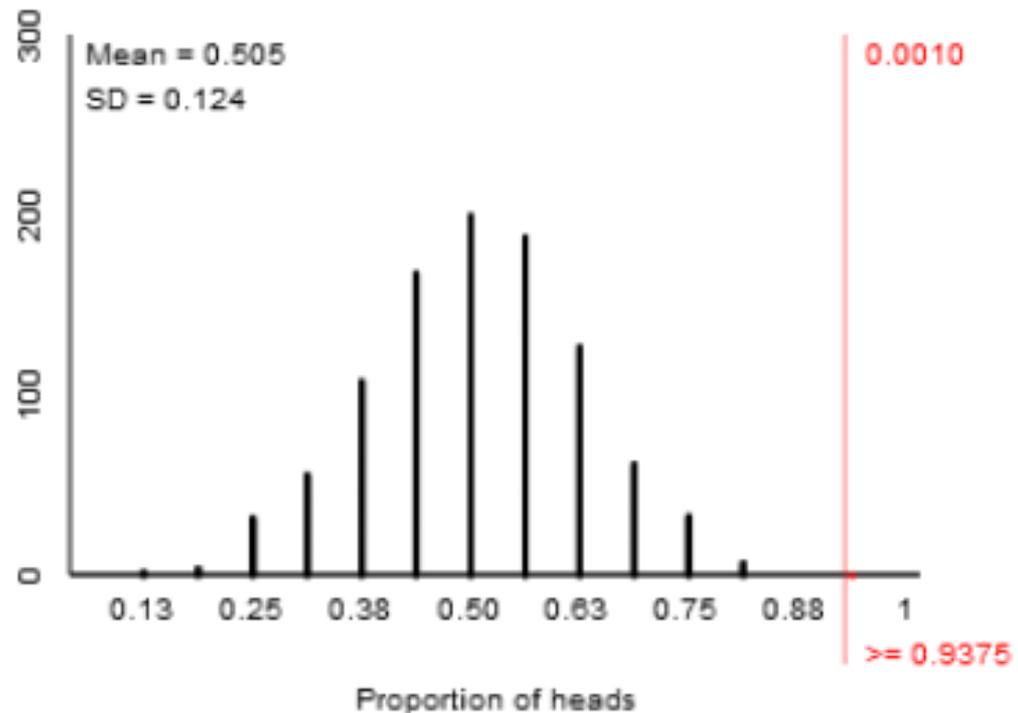
Conclusions

- The simulation analysis provided strong evidence that Buzz wasn't just guessing

- "Buzz was just guessing" is not a **plausible** explanation for the results

- Have we proven that Doris and Buzz can communicate?

- Next steps?

Doris and Buzz Redo

- Instead of a canvas curtain, Dr. Bastian constructed a wooden barrier between Buzz and Doris

- When tested, Buzz pushed the correct button 16 out of 28 times

- Are these results statistically significant?

- Let's go to the applet to check this out

Questions?

Thank you!

dolansky@sunydutchess.edu

Session S054

Thursday, November 15

1:50 – 2:40 pm

Presenter: Barbara Dolansky

Can Dogs Understand Human Cues?

Dogs have been domesticated for about 14,000 years. In that time, have they been able to develop an understanding of human gestures such as pointing or glancing? How about similar nonhuman cues? Researchers Udell, Giglio, and Wynne tested a small number of dogs in order to answer these questions.

In this exploration, we will first see whether dogs can understand human gestures as well as nonhuman gestures. To test this, the researchers positioned the dogs about 2.5 meters from the experimenter. Two cups were placed, one on each side of the experimenter. The experimenter would perform some sort of gesture (pointing, bowing, looking) toward one of the cups or there would be some other nonhuman gesture (a mechanical arm pointing, a doll pointing, or a stuffed animal looking) toward one of the cups. The researchers would then see whether the dog would go to the cup that was indicated. There were six dogs tested. We will look at one of the dogs in two of his sets of trials. This dog, a four-year-old mixed breed, was named Harley. Each trial involved one gesture and one pair of cups, with a total of 10 trials in a set.

We will start out by looking at one set of trials where the experimenter bowed toward one of the cups to see whether Harley would go to that cup.

STEP 1: State the research question.

1. Based on the description of the study, state the research question.

STEP 2: Design a study and collect data.   Harley was tested 10 times and 9 of those times he chose the correct cup.

2. What are the observational units?

3. Identify the variable in the study. What are the possible outcomes of this variable? Is this variable quantitative or categorical?

STEP 3: Explore the data.   With categorical data, we typically report the number of "successes" or the proportion of successes as the statistic.

4. What is the number of observational units (sample size) in this study?

5. Determine the observed statistic and produce a simple bar graph of the data (have one bar for the proportion of times Harley picked the correct cup and another for the proportion of times he picked the wrong cup).

Definitions

The set of observational units on which we collect data is called the **sample**. The number of observational units in the sample is the **sample size**. A **statistic** is a number summarizing the results in the sample.

6. If the research conjecture is that Harley can understand what the experimenter means when they bow toward an object, is the statistic in the direction suggested by the research conjecture?

7. Could Harley have gotten 9 out of 10 correct even if he really didn't understand the human gesture and so was randomly guessing between the two cups?

8. Do you think it is likely Harley would have gotten 9 out of 10 correct if he was just guessing randomly each time?

STEP 4: Draw inferences beyond the data. There are two possibilities for why Harley chose the correct cup 9 out of 10 times:

- He is merely picking a cup at random and in these 10 trials happened to guess correctly in 9 of them. That is, he got more than half correct just by random chance alone.
- He is doing something other than merely guessing and perhaps understands what the experimenters mean when they bow towards the cup.

The unknown long-run proportion (i.e., probability) that Harley will choose the correct cup is called a *parameter*.

We don't know the value of the parameter, but the two possibilities listed above suggest two different possibilities.

9. What is the value of the parameter if Harley is picking a cup at random? Give a specific value.

Definition

For a random process, a *parameter* is a long-run numerical property of the process.

10. What is the possible range of values (greater than or less than some value) for the parameter if Harley is not just guessing and instead understands the experimenter?

We will show you how statisticians use simulation to make a statement about the strength of evidence for these two possible statements about the parameter's value.

The chance model

Statisticians often use *chance models* to generate data from random processes to help them investigate the process. In particular, they can see whether the observed statistic is consistent with the values of the statistic simulated by the chance model. If we determine that Harley's results are not consistent with the results from the chance model, we will consider this to be evidence against the chance model and in favor of the research conjecture, that he understands the bowing gesture. In this case, we would say Harley's results are *statistically significant*, meaning unlikely to have occurred by chance alone.

We can't perform the actual study more times in order to assess the second possibility, but we can simulate the behavior of Harley's choices if we were to assume the first possibility (that he is simply guessing every time).

11. Explain how you could use a coin toss to represent Harley's choices if he is guessing between the two cups each time. How many times do you have to flip the coin to represent one set of Harley's attempts? What does heads represent?

12. If Harley was guessing randomly each time, on average, how many out of the 10 times would you expect him to choose the correct cup?

13. Simulate one repetition of Harley guessing randomly by flipping a coin 10 times (why 10?) and letting heads represent selecting the correct cup ("success") and tails represent selecting the incorrect cup ("failure"). Count the number of heads in your 10 flips. Combine your results with the rest of the class to create a *dotplot* of the distribution for the number of heads out of 10 flips of a coin.
    a. Where does 9 heads fall in the distribution? Would you consider it an unusual outcome or a fairly typical outcome for the number of heads in 10 flips?

    b. Based on your answer to the previous question, do you think it is plausible (believable) that Harley was just guessing which cup to choose?

Using an applet to simulate flipping a coin many times

To really assess the typical values for the number of heads in 10 coin tosses (number of correct picks by Harley assuming he is guessing at random), we need to simulate many more outcomes of the chance model. Open the **One Proportion** applet from the textbook webpage.

Notice that the probability of heads has been set to be 0.50, representing Harley guessing between the two cups. Set the number of tosses to 10 and press the **Draw Samples** button. What was the resulting number of heads?

Notice that the number of heads in this set of 10 tosses is then displayed by a dot on the graph. Uncheck the **Animate** box and press the **Draw Samples** button 9 more times. This will demonstrate how the number of heads varies randomly across each set of 10 tosses. Nine more dots have been added to your dotplot. Is a pattern starting to emerge?

Now change the **Number of repetitions** from 1 to 990 and press **Draw Samples**. The applet will now show the results for the number of heads in 1,000 different sets of 10 coin tosses. So each dot represents the number of times Harley chooses the correct cup out of 10 attempts assuming he is just guessing.

Remember why we conducted this simulation: to assess whether Harley's result (9 correct in 10 attempts) would be unlikely to occur by chance alone if he were just guessing between the pair of cups for each attempt.

14. Locate the result of getting 9 heads in the dotplot created by the applet. Would you consider this an unlikely result in the tail of the distribution of the number of heads?

15. Based on the results of 1,000 simulated sets of 10 coin flips each, would you conclude that Harley would be very unlikely to have picked the correct cup 9 times in 10 attempts

if he was randomly guessing between the two cups each time? Explain how your answer relates to the applet's dotplot.

16. Do the results of this study appear to be statistically significant?

17. Do the results of this study suggest that Harley just guessing is a plausible explanation for Harley picking the correct cup 9 out of 10 times?

Summarizing your understanding

18. To make sure that you understand the coin-flipping chance model, fill in Table 1.2 indicating what parts of the real study correspond to the physical (coin-flipping) simulation.

Table 1.2 Parallels between real study and physical simulation		
Coin flip	=	
Heads	=	
Tails	=	
Chance of heads	=	
One repetition	=	one set of _____ simulated attempts by Harley

The 3S strategy

We will call the process of simulating could-have-been statistics under a specific chance model the *3S strategy*. After forming our research conjecture and collecting the sample data, we will use the 3S strategy to weigh the evidence against the chance model. This 3S strategy will serve as the foundation for addressing the question of statistical significance in Step 4 of the statistical investigation method.

3S Strategy for Measuring Strength of Evidence

1. **Statistic:** Compute the statistic from the observed sample data.
2. **Simulate:** Identify a "by-chance-alone" explanation for the data. Repeatedly simulate values of the statistic that could have happened when the chance model is true.
3. **Strength of evidence:** Consider whether the value of the observed statistic from the research study is unlikely to occur if the chance model is true. If we decide the observed statistic is unlikely to occur by chance alone, then we can conclude that the observed data provide strong evidence against the plausibility of the chance model. If not, then we consider the chance model to be a plausible (believable) explanation for the observed data; in other words what we observed could plausibly have happened just by random chance.

Let's review how we have already applied the 3S strategy to this study.

19. Statistic. What is the statistic in this study?

20. **Simulate.** Fill in the blanks to describe the simulation. We flipped a coin _____ times and kept track of how many times it came up heads. We then repeated this process _____ more times, each time keeping track of how many heads were obtained in each of the _____ flips.

21. **Strength of evidence.** Fill in the blanks to summarize how we are assessing the strength of evidence for this study. Because we rarely obtained a value of _____ heads when flipping the coin _____ times, this means that it is _____ (believable/unlikely) that Harley is just guessing, because if Harley was just guessing he _____ (rarely/ often) would get a value like _____ correct out of _____ attempts.

STEP 5: Formulate conclusions.

22. Based on this analysis, are you convinced that Harley can understand human cues? Why or why not?

Another study

One important step in a statistical investigation is to consider other models and whether the results can be confirmed in other settings.

23. In a different study, the researchers used a mechanical arm (roughly the size of a human arm) to point at one of the two cups. The researchers tested this to see whether dogs understood nonhuman gestures. In 10 trials, Harley chose the correct cup 6 times.

    a. Using the dotplot you obtained when you simulated 1,000 sets of 10 coin flips assuming Harley was just guessing, locate the result of getting 6 heads. Would you consider this an unlikely result in the tail of the distribution?

    b. Based on the results of 1,000 simulated sets of 10 coin flips each, would you conclude that Harley would be very unlikely to have picked the correct cup 6 times in 10 attempts if he was randomly guessing between the two cups each time? Explain how your answer relates to the applet's dotplot.

    c. Is this study's result statistically significant?

    d. Do the results of this study suggest that Harley just guessing is a plausible explanation for Harley picking the correct cup 6 out of 10 times?

    e. Does this study prove that Harley cannot understand the mechanical arm?

STEP 6: Look back and ahead.

24. Compare the analyses between the two studies. How does the unusualness of the observed statistic compare between the two studies? Does this make sense based on the value of the observed statistic in the two studies? Does this make sense based on how the two studies were designed? Explain. (*Hint:* Why might the results differ for human and mechanical arms? Why would this matter?)

25. A single study will not provide all of the information needed to fully understand a broad, complex research question. Thinking back to the original research question, what additional studies would you suggest conducting next?