HELPING STUDENTS EVALUATE THE STRENGTH OF EVIDENCE IN SCIENTIFIC ARGUMENTS

Thinking About the Inferential Distance Between Evidence and Claims

"Look, the dragonfly nymph is attacking the water snail. He's eating it!"

"Nah, he's just hanging on so he can go for a ride. The snail is carrying him around 'cause they're friends."

by Lauren Brodsky, Andrew Falk, and Kevin Beals

As the exchange on the opening page illustrates, students tend to have no problem using inference to make claims based on what they directly observe. On the contrary, they often have to be reminded to make careful observations and ground their inferences in those observations and other data. Thoughtful science teachers spend considerable time working with students to improve their data-collection and -recording skills, and in maintaining a distinction between what is directly observable or measurable by scientists and what must be inferred. But the act of making inferences also deserves attention. It can be tricky to teach students how to evaluate the strength of particular empirical evidence in relation to the inferences that can be made from it. When evidence and inferences are being tossed around your classroom, how do you help students to become more aware and thoughtful about evaluating them in relation to each other—and be appropriately tentative and skeptical?

The national guidelines and standards for science education emphasize both the importance of students making inferences from evidence to construct arguments and of students critiquing the strength of others' arguments based on evidence (Achieve Inc. 2013; NRC 1996; NRC 2012). How do we as teachers support our students in understanding the relationship between observed evidence and possible inferences? How do we help students to think about how to use investigation, critique, and inductive reasoning to make stronger inferences or to seek stronger evidence for those inferences? In our work as science curriculum developers with middle school students, we have begun to develop what we believe is a promising way of thinking about the relationship between what is observed and what is inferred, and how to become more certain about inferences and resulting arguments.

In this article, we share a description of this proposed way of thinking about observations and inferences, called *inferential distance*. First we describe the context in which we explored the ideas in informal ways with middle school students. Next we describe three dimensions of inferential distance that can be used to critique different kinds of evidence. For each dimension, we describe some specific examples of how we introduced and examined it with students. We end with some more general suggestions about how to integrate this way of thinking into science instruction through classroom activities with students.

The context: An after-school program

We explored and tested our ideas about evidence and inference working with students in an after-school science program in a local middle school. One of us (Beals) met with students after school twice a week for three months. He worked with students to investigate terrarium pond ecosystems that included crayfish, dragonfly nymphs, tubifex worms, mosquito larvae, gambusia fish, freshwater snails, and other organisms. Beals and the students constructed a collective food web that represented the transfer of energy between the organisms in the ecosystem.

A primary focus of the program was developing students' knowledge and skills with scientific argumentation, and students were expected to support the connections they created on the food web with evidence. Students recognized that connections could be more or less strongly supported, and with Beals, they developed a color-coding system for representing how confident they were about a particular connection.

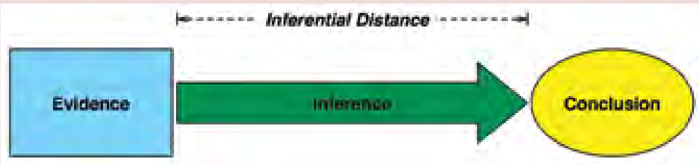
- Orange arrows were inferred relationships without direct evidence. For example, "I think tubifex worms eat water plants, because they are sometimes on them."
- Green arrows were supported by at least one direct observation by a student or an adult in the room. Adding a green arrow also involved some verbal peer review. For example, "Are you sure you saw it eat it?" and "Did you see it going into its mouth?" If students were reasonably sure of their observation, they would add the green arrow to the web, and write their name alongside the arrow.
- Pink arrows represented connections based on secondary information taken from a variety of written sources, and included the name of the source along those arrows, as well.

We did not use the term *inferential distance* explicitly with students. Discussing their connections in terms of their level of certainty accomplished the same goals of pushing them to weigh and evaluate evidence, and the wording was more accessible.

Inferential distance: A definition

To make an *inference* in science is to draw a possible conclusion based on known empirical evidence or information. The *inferential distance* is the size of the

FIGURE 1 Inferring conclusions from evidence



conceptual leap made in going from evidence to conclusion. The shorter the inferential distance, the closer the conclusions are to being a direct description of the evidence. Figure 1 shows a visual representation of the relationship among evidence, conclusion, and inferential distance. However, inferential distance is not unidimensional; it can vary in multiple ways, making particular evidence stronger or weaker in relation to other evidence. We will use similar diagrams to represent those individual dimensions of inferential distance.

Dimension 1: Inferential distance based on a subset of a population or category

In many scientific investigations, the end goal of an experiment or set of observations is to be able to draw a conclusion that can be generalized to all examples of a phenomenon. Scientists, and students conducting classroom experiments, look for patterns or rules that can be applied to all equivalent situations in the natural world. For example, when scientists study the effects of a drug on a group of patients in a clinical trial, they would like to draw conclusions not just about the people in that trial, but about the drug's potential effects on all people. To make these generalizations requires the inference that whatever results are found (in the lab, with a set number of measurements, etc.) is also true of everything in that larger category. Scientists seek to maximize sample sizes or replicate investigations to better support these generalizations.

Inferential distance in this case describes the size of the inference that is made in moving from the actual evidence to the conclusion about the larger group and is related to the number of trials or observations that are made. If an investigation is being done to find out what crayfish eat, the end goal is to be able to conclude that what was observed is indicative of what all crayfish eat. If only one crayfish is observed, this evidence is clearly not very strong. If half of all the crayfish in the world were observed, there would be much more confidence in inferring that the rest of the crayfish eat the same thing. This trend continues—as the number of observations increases, the inferential distance decreases until all crayfish have been observed and no inference needs to be made—the conclusion becomes a description of the evidence. It is seldom, if ever, feasible to make all possible observations, and evidence from a comparatively small subset can still be strong. Being aware of the inference that is made in shifting from the observed subset to the whole population can help students weigh the strength of evidence between two different subsets, or to describe what assumptions are being made and how a given set of evidence could be stronger. Figure 2 represents this dimension of inferential distance.

Beals encouraged students in the after-school program to take into account the number of observations when evaluating the level of certainty of arrows in the food web. Over time, if others made the same observation someone else had already represented with a green arrow, they added their names to the arrow. This created a visual representation of the varying levels of certainty that existed even for interactions that were directly observed. Only one student observed evidence of a dragonfly nymph eating mosquito larvae, so his was the only name along the green arrow between them. The arrow connecting tubifex worms and gambusia fish ultimately had seven names written along it, representing seven different corroborating observations, and a much higher level of certainty. Figure 3 represents the different degrees of inferential distance in this kind of example.

FIGURE 2 Inferential distance based on a subset of a population or category

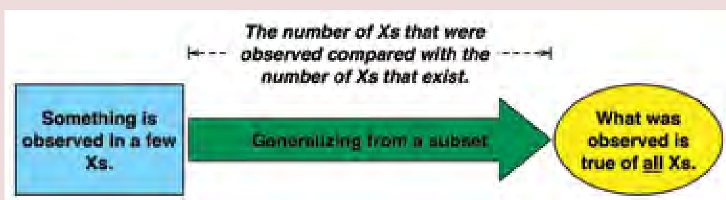
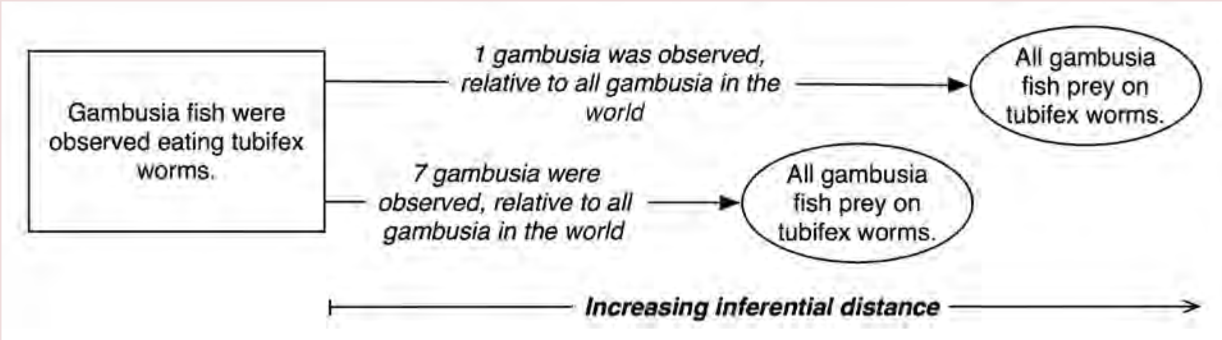


FIGURE 3 Differences in inferential distance based on the number of observations



Dimension 2: Inferential distance based on a model of a phenomenon

Another type of inferential distance that is common both in science and in the science classroom comes with inferring that evidence from a model of a phenomenon is evidence about the phenomenon in the real world. The inference is that the model behaves like the phenomenon it represents for the features being studied. For example, when scientists were studying how particular patterns of wing motion enabled insects to stay airborne, they created large mechanical wing models that could imitate those patterns and used those models to measure the forces on the wings (Dickinson, Lehmann, and Sane 1999). Their inference was that the mechanical wings, which had the same shape and moved in the same ways as insect wings, would experience the same forces and provide the same lift that allows insects to fly.

In creating a model, assumptions must be made about which features of the phenomenon are relevant, and how they can be re-created in ways that behave similarly. The more similar the model is to the phenomenon (the more comprehensive the features, the more similar the scale, the more matched the materials, etc.), the more likely it is that it will behave as the phenomenon would, and the smaller the inference that needs to be made. This can continue until at some point the model is identical to the phenomenon and no inference needs to be made. Figure 4 represents this dimension of inferential distance.

The terrarium in the after-school program was serving as a model of a pond ecosystem. Students were inferring that the organisms in the terrarium would behave in the same way as organisms in a pond. When students were trying to determine what the gambusia fish ate, they had to infer that what they observed in the terrarium is what they would observe in a pond—that the model behaved like the real-world phenomenon. Many assumptions went into the terrarium model that could affect the strength of their evidence about the prey of gambusia fish. Students had to assume, for example, that what gambusia ate in ponds was available to them in the terrarium. Given two terrariums, one that included a more complete set of organisms from the pond, would require fewer assumptions about what a gambusia fish would or would not eat. If students observed gambusia fish eating tubifex worms in this more complete terrarium, they could be more certain that they were observing what a gambusia would eat in a pond environment. The inferential distance from this model to the conclusion about pond interactions would be smaller than if they were making inferences using a model with only a subset of species. Figure 5 represents the different degrees of inferential distance in this kind of example.

FIGURE 4 Inferential distance based on a model of a phenomenon

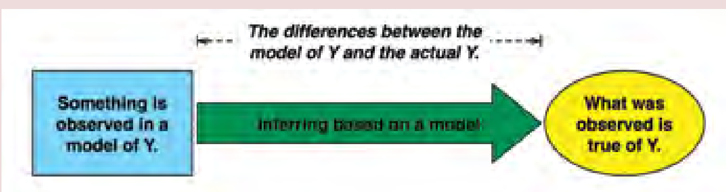
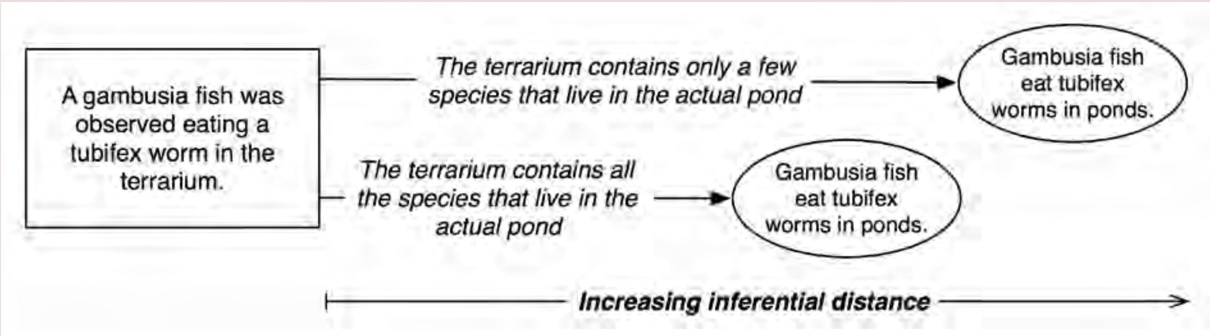


FIGURE 5 Differences in inferential distance based on similarities of a model



Dimension 3: Inferential distance based on alternative explanations

Finally, inferential distance exists whenever there is a conclusion that one thing caused or resulted from another, but the causal interaction was not actually observed. Without direct observation, alternative explanations can always exist. The inference is that, of all the possible explanations, the one proposed is actually the case. Scientists looking for evidence that liquid water could once have existed on the surface of Mars initially had to rely on low-resolution images of the surface. They saw channels that could have been evidence of past surface-water flow but also could have been caused by lava. Scientists inferred, but were not certain, that the channels on Mars were in fact from surface water and decided to continue studying those channels. As higher-resolution images of the surface were collected, the possibility that some of the channels were created by lava was ruled out, and the conclusion that the channels were created by surface-water flow became more certain (Sharp and Malin 1975). Evidence for a proposed explanation becomes stronger as there are fewer possible alternative explanations that could also account for it, until cause and effect are directly observed, and no inference needs to be made. Figure 6 represents this dimension of inferential distance.

In the after-school program, students were unable to directly observe a dragonfly nymph eating anything, so a student set up an investigation in a more controlled environment. He put a dragonfly nymph in a separate container and then added mosquito larvae, counting them as he added them. When he returned the next day, he noticed that almost all of the mosquito larvae were gone. This student inferred that the dragonfly nymph ate the mosquito larvae and was ready to add a green arrow to the food web. Other students challenged his investigation and pointed out that there were other possible explanations for how the larvae disappeared. One student noted that the top of the container was not completely sealed, and instead of being eaten by the dragonfly nymph, the larvae could have escaped. The student who created the investigation decided to put the dragonfly nymph and larvae together overnight again, but this time he made sure the top of the terrarium was sealed. When the larvae were gone again the next day, students decided that this was stronger evidence that the dragonfly nymph ate the larvae. Based on this new evidence, the student added a green arrow to the food web with his name alongside it. There were fewer possible alternative explanations and so the inferential distance was smaller. Figure 7 represents the different degrees of inferential distance in this kind of example.

FIGURE 6 Inferential distance based on alternative explanations

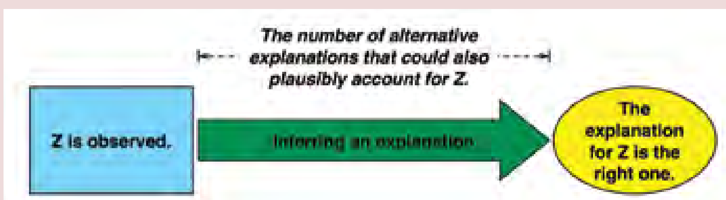
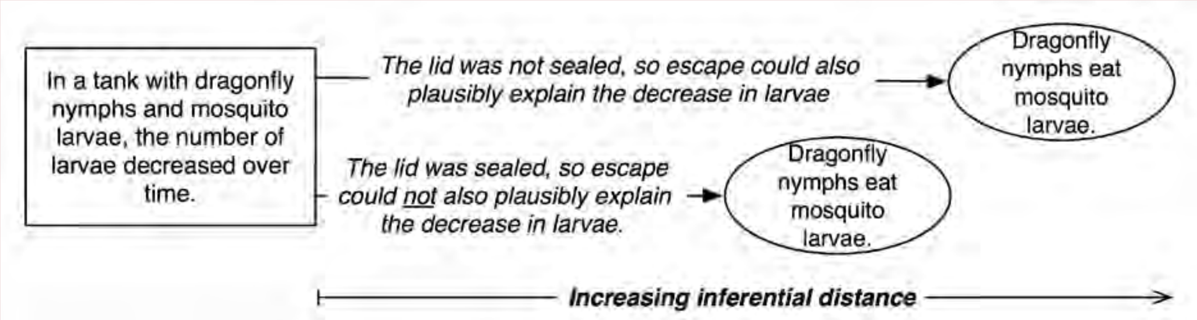


FIGURE 7 Differences in inferential distance based on alternative explanations



How to use inferential distance in the classroom

Helping students to develop an understanding of inferential distance (even if it is an informal understanding) is a powerful way of supporting them in both evaluating and producing scientific reasoning and arguments. Thinking about inferential distance can help students to critique evidence and related conclusions in the investigations and science texts they encounter in their classroom learning, and in their lives outside of school. It can also help them to construct their own conclusions and qualify them as more or less tentative. In Figure 8, we describe a set of strategies that we have found to be useful with students in introducing them to inferential distance and engaging them in using it to evaluate scientific arguments.

Conclusion

Given that the *Next Generation Science Standards* and the *National Science Education Standards* expect students to construct their own explanations based on evidence and evaluate and critique their own and others' arguments, it is important to provide explicit instruction and coaching about what makes evidence weaker or stronger and what makes conclusions more or less certain. Thinking about the inferential distance between evidence and conclusion is powerful because it provides a means for students to weigh different pieces of evidence and to evaluate the strength of evidence in relation to particular conclusions. In this way, it offers a concrete approach to being appropriately tentative in drawing conclusions and making claims and in evaluating and critiquing different arguments in science. ■

Lauren Brodsky (*brodsky@berkeley.edu*) is an assessment developer and ***Andrew Falk*** and ***Kevin Beals*** are science curriculum developers, all at the Lawrence Hall of Science at the University of California, Berkeley, in Berkeley, California.

References

Achieve Inc. 2013. *Next generation science standards. www.nextgenscience.org/next-generation-science-standards.*

Dickinson, M.H., F.-O. Lehmann, and S.P. Sane. 1999. Wing rotation and the aerodynamic basis of insect flight. *Science* 284 (5422): 1954–60.

National Research Council (NRC). 1996. *National science education standards*. Washington, DC: National Academies Press.

National Research Council (NRC). 2012. *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Nussbaum, E.M. 2011. Argumentation, dialogue theory, and probability modeling: Alternative frameworks for argumentation research in education. *Educational Psychologist* 46 (2): 84–106.

Sharp, R.P., and M.C. Malin. 1975. Channels on Mars. *Geological Society of America Bulletin* 86 (5): 593–609.

Resource

Aquatic habitats: Exploring desktop ponds (aquatic ecosystem activity)—*http://lhsgems.org/GEMaquatic.html*

FIGURE 8 Instructional activities for teaching inferential distance

Activity	How we implemented the activity	Example for one dimension of inferential distance
Card sorts Students arrange cards that show different but related information. The cards are useful in helping students develop an understanding of inferential distance.	We provide students with a question, a possible conclusion, and a set of note cards, each with different evidence. Alternatively, we provide a question and a summary of particular evidence, and ask students to sort conclusions in order of least to most tentative. We then ask students to sort the cards in order of most to least certain in relation to the conclusion or evidence.	Dimension 3: Alternative explanations Question: "What are the predator/prey relationships on the African savannah?" Possible conclusion: "Cheetahs prey on wildebeest." The set of cards showed pictures representing various kinds of possible evidence (picture of a cheetah's sharp teeth; cheetah running behind a wildebeest; cheetah pouncing on a wildebeest; and cheetah tearing at a wildebeest carcass).
Critical questions Students use a set of questions as cues in critiquing the strength of evidence based on inferential distance.	Once students have an initial understanding of inferential distance, we provide them with a set of critical questions (Nussbaum 2011) that they can learn to apply. Opportunities for critique include conclusions drawn in scientific writing, scientific work that is reported in the news, or peers' lab reports. Students can also critique the inferential distance involved when using models to investigate phenomena that cannot be brought into the classroom.	Dimension 2: Model of a phenomenon We asked students to critique the ways in which conclusions drawn about a pond based on their aquariums might be more or less strong. Critical questions we provided were "How is the model similar to and different from the situation or process we are drawing conclusions about?" and "How certain do the similarities or differences make us about our conclusions?" Before taking individual responses, we had students hold up one to five fingers to indicate the level of similarity or certainty.
Possible conclusions and new evidence Students draw a range of possible conclusions for themselves and can propose investigations to strengthen conclusions verbally or in writing.	Once students are more comfortable critiquing inferential distance, we ask them to examine evidence that they have gathered or drawn from provided data. From the evidence, they propose multiple conclusions that they consider more and less certain based on that evidence. Alternately, they can propose the conclusion they think is currently the best one, and rate it in terms of certainty. We then invite students to propose further investigations that could make the conclusions more certain.	Dimension 3: Alternative explanations When students shared their observations of their terrariums, we asked them to suggest several conclusions based on the data, some that they thought were more certain, and some that they thought were less certain, because other explanations were possible (e.g., the mosquito larvae were eaten, or they escaped). We then invited the group to propose investigations that would test or control for an alternative explanation (e.g., tightly sealing the lid).