Chapter Seven

NEW DATA AND EXPERIMENTATION

> Science is best defined as a careful, disciplined, logical search for knowledge about any and all aspects of the universe, obtained by the best available evidence and always subject to correction and improvement upon discovery of better evidence. What's left is magic. And it doesn't work.
>
> James Randi

THE CRAZY PHILOSOPHER'S EVIDENCE

As you will remember, Johnson thought he had discovered evidence that there was glitch in his iPod software. His schematized argument was as follows.

e_1. Johnson went to a Pink Martini concert, planning to ask for a specific encore.
e_2. "Que Sera Sera" was played during the concert.
e_3. He never got a chance to ask for "Lily."
e_4. On the ride home the next morning he set his iPod to play all 36 of the Pink Martini songs.
e_5. He set the iPod to "shuffle songs."
e_6. He listened to all 36 songs.

e_7. The last two songs played were "Lily" and "Que Sera Sera" -- the imagined encore from the night before!

e_8. "Lily" and "Que Sera Sera" are the two Pink Martini songs he listens to most often.

---

t_0. There is a glitch in the iPod software -- rather than playing the songs in completely "random" order, it is weighting things according to how often songs are listened to.

There are 36 Pink Martini songs in Johnson's iPod. What are the odds of his imagined encore occurring on the drive home? Let's spend just a minute and figure that out. "Lily" came up as the next to the last song played. The odds of this happening are straightforward. Any one of 36 songs could have come up here, so the odds are 1/36. But to have the encore you had to also have "Que Sera Sera" come up last. So what are the odds of that happening? It's actually easy to figure out. We already know the odds of "Lily," so it's a question of "Lily" ***and*** "Que Sera Sera." Since "Lily" has already been played, the odds of "Que Sera Sera" are 1/35, and the odds of "Lily" ***and*** "Que Sera Sera" are 1/36 x 1/35, or 1/1260. But, of course, I would have also had my encore if the last two songs had been "Que Sera Sera" and then "Lily." The odds of this happening figure out exactly the same -- 1/1260. So the odds of my encore popping up -- ["Lilly" and "Que Sera Sera"] ***or*** ["Que Sera Sera" and "Lily"] are 1/1260 + 1/1260, or 1/630.

Certainly one thing that would explain that 1/630 shot coming up on the ride home is that my imagined encore was composed of my two favorite (and most listened to) Pink Martini songs, and the program was illegitimately taking this into account in generating the "random" play order. But, I hope it's obvious by now, it's easy enough to think of lots of rival explanations.

t_1. This was just a true, 1/630, coincidence.
t_2. This is not a software glitch; the iPod software is designed to do exactly this.
t_3. The iPod software is illegitimately weighting things, not by number of times played, but something else – length of the songs, where they occur in the album, etc.
t_4. The philosopher set his iPod incorrectly.
t_5. The philosopher dosed in and out on the drive home and only thought that these two songs came up last.
t_6. The problem is in Johnson's iPod, the hardware, not the software.

My students have been worrying about what happened for the last couple of years on quizzes, ever since this really happened on a drive back from the Oregon League of Cities. They pretty generally rank the coincidence hypothesis as a much better explanation, though they are often surprised once they see the math that the odds are really 1/630. They also don't seem to have too much confidence in their professor, since explanations like t_4 and t_5 are consistently ranked ahead of t_0. So, according to the Inference to the Best Explanation recipe, these students are committed to saying that Johnson's evidence for the glitch theory is pretty weak.

WHY DON'T YOU JUST TEST IT?

I've told you this little anecdote for two very different reasons. One, of course, is I wanted a little exercise that would allow you to apply the inference to the best explanation test from the Chapter Five to an argument. The other, though, is to tell you about a very common feature that my students have felt compelled to add to their discussions. There is almost a sense of frustration, or least the need to lecture their professor.

They suggest, indeed insist on, a very simple test of the glitch hypothesis.

Look, isn't there an obvious way to settle this matter? Turn off the iPod, reset everything, play Pink Martini's songs again and see what happens. What is being proposed here is a classic little experiment, the kind of thing that some philosophers and scientists say is the defining condition of real science. I hope to convince you in the next couple of chapters that there is something brilliantly right about this claim, but at the same time, dangerously misleading.

A PRETTY PICTURE OF SCIENCE

Here is an idealization about the natural sciences. The scientist is really smart and is trained to go about her business in a very special, almost ritualized, way. She goes out and observes the world. Being smart, and being trained to be a careful observer, she notices things. Sometimes she is puzzled by the things she observes and she asks questions -- why am I observing this? She starts looking for an ***explanation***. Being smart and creative she thinks about this really hard and comes up with a possible answer -- a ***hypothesis*** or a ***theory***. This is all fine and good, but according to the pretty picture, it's only now that the rules of science kick in. It's not good enough to just have a theory, the theory must now be tested. The scientist must devise an experiment, and let the results of the experiment determine the fate of her theory.

Bear with me for a bit of technical stuff in symbolic logic. Logicians talk about ***conditionals***, "if … then" sentences. There are two ***valid*** inferences that follow directly from a true conditional.

1. ***If*** the figure is a plane right triangle, ***then*** the interior angles total 180^0.
2. The figure is a plane right triangle.

---

3. The interior angles total 180^0.

This inference is called ***modes ponens***. A kind of mirror image inference is called ***modes tollens***.

1. ***If*** the figure is a plane right triangle, ***then*** the interior angles total 180^0.
1. The interior angles ***do not*** total 180^0.

2. The figure is ***not*** a plane right triangle.

Finally there is a tempting inference that is not valid, but is rather a logical fallacy, ***affirming the consequent***.

1. ~~***If*** the figure is a plane right triangle, ***then*** the interior angles total 180^0.~~
2. ~~The interior angles total 180^0.~~

3. ~~The figure is a plane right triangle.~~

You can easily spot the fallacy by noting that figure might total 180^0 because it's a triangle, but at the same time not be a right triangle, but rather, say, an equilateral triangle.

OK, so what does all of this have to do with the pretty picture of science, and maybe Johnson's iPod? Well, suppose the conditional sets up something we might expect to see in an experimental circumstance, given the theory we are testing is true.

1. If the theory is true, we will see … in the experiment.

By the inference of *modes tollens* we will be able to ***falsify*** theory by ***disconfirming*** it in an experiment

1. If the theory is true, we will see … in the experiment.

2. We ***do not*** see … in the experiment.
3. The theory is ***not*** true.

Experiments, according to the pretty picture, provide tests that can show us that theories are false. They cannot, however, show us that theories are true. Remember, it is a fallacy to affirm the consequent.

1. ~~If the theory is true, we will see … in the experiment.~~
2. ~~We see … in the experiment.~~
3. ~~The theory is true.~~

A BETTER, BUT UNTIDY, PICTURE OF SCIENTIFIC DISCONFIRMATION

Now, the theory about the iPod hardly counts as deeply scientific, but suppose we imagine an experiment, nonetheless. The conditional that sets all this up looks something like the following.

1. If there is a glitch in the software, so that when the iPod is set to play all the songs by an artist, and is set to "shuffle" these songs, then rather than playing them in random order, it will play the most often listened to tracks last.

I could test my theory by reprogramming everything with the Pink Martini tracks, but since I've offered a general theory, let's test it with a different artist. I have lots of Lucinda Williams' albums, and I'm certain I listen to two of her songs, "Right in Time," and "Essence," the most. So if I set my iPod to play all of her tracks, and to shuffle them, I am predicting that the two songs will be played last.

Suppose I do all of this with my iPod, and listen to all of her songs, more than a hundred, I'd say. We can imagine four different out comes to the experiment. Focusing on the last two songs, we might observe any of the following.

e_n^a. The two songs come up as the last two played.
e_n^b. Neither song is in the last two.
e_n^c. Only "Right in Time" is in the last two.
e_n^d. Only "Essence" is in the last two.

e_n^c and e_n^d are interesting and deserve further study, but let's set them to the side and focus on the "pure" experimental outcomes. According to the pretty picture, e_n^b conclusively establishes that the glitch theory is false. But isn't that a little extreme? We've already honed our skills at rival explanations – surely we can imagine scenarios where the glitch hypothesis is (was) true, but neither song played last.

t_1. Between the drive home and the experiment, iTunes downloaded a newer (debugged) version of the software.
t_2. The glitch only occurs in playlists shorter than 50 songs.
t_3. There is a countervailing glitch when any of the songs are classified as "country."

It's doubtful in the extreme that a negative experimental outcome can ***falsify*** a theory, though it certainly can provide strong evidence that there is something wrong with the theory.

The problem here goes back to the original conditional that set up the experiment in the first place. Remember the difference between a ***sound*** argument and a ***valid*** one? The if … then sentence that gets our inference going in the first place states an absolute connection between the glitch theory

and the predicted outcome of the experiment. But the rival explanations we have just considered above seem to show that this connection is not so absolute after all. Almost always the conditional that sets up our experiment contains what Larry Wright calls a ***weasel word***. A more modest, but also more accurate, statement of the predicted experimental outcome will look more like this.

> If the theory in question is true, then ***all things being equal*** we will see … in our experiment.

We predict that we will observe an as yet undiscovered planet at such and such location in the night sky. But certainly not, if the observatory is socked in by clouds. We expect the solution to turn a certain color in our chemistry experiment, but not if the test tube is contaminated.

When we include this suppressed, but understood, *ceteris paribus* clause[1], our inference looks a little more problematic.

1. If there is a glitch in the software, so that when the iPod is set to play all the songs by an artist, and is set to "shuffle" these songs, then, ***all things being equal***, rather than playing them in random order, it will play the most often listened to tracks last.
2. "Essence" and "Right in Time" did not play last.

Two valid conclusions can be derived from these premises. One, of course, is that the glitch hypothesis is mistaken. But as a matter of pure logic, it is equally legitimate to infer that all things in our experimental circumstances were ***not*** equal.

Does any of this mean that the "scientific method" and the requirement that we experimentally test our theories is a waste of time? Nothing could be further from the truth. Let's

go back to our original "evidence" for the glitch theory, but add to it the new data from our experiment.

e_1. Johnson went to a Pink Martini concert, planning to ask for a specific encore.
e_2. "Que Sera Sera" was played during the concert.
e_3. He never got a chance to ask for "Lily."
e_4. On the ride home the next morning he set his iPod to play all 36 of the Pink Martini songs.
e_5. He set the iPod to "shuffle songs."
e_6. He listened to all 36 songs.
e_7. The last two songs played were "Lily" and "Que Sera Sera" -- the imagined encore from the night before!
e_8. "Lily" and "Que Sera Sera" are the two Pink Martini songs he listens to most often.
e_9. When Johnson tried the shuffle all songs routine for Lucinda Williams, his most listened to songs did not come up last.

---

t_3. There is a glitch in the iPod software -- rather than playing the songs in completely "random" order, it is weighting things according to how often songs are listened to.

We've already imagined some rivals to e_9, but I assume that you would all agree with me that t_0 has been seriously weakened by our experiment, and that the random fluke hypothesis, or the operator error rivals, look even better.

The moral here is straightforward. When a theory suggests that we can expect to see something as yet undiscovered, and we go out and look for this thing, but don't find it, this is highly relevant ***new data*** that almost always hurts the status of the

original explanation as being the best explanation of everything, including, of course, the experimental results.

A BETTER, BUT UNTIDY, PICTURE OF SCIENTIFIC CONFIRMATION

None of what I have just told you is earthshaking, and not known by careful scientists and philosophers. Still, the pretty picture, partly because is so pretty, can allow us to lose sight of the subtleties of experimental design and protocol. Maybe even more problematic for the pretty picture is the evidential value of experimental confirmation.

Suppose I program my iPod to play all 116 Lucinda Williams tracks. I set the iPod to shuffle the songs, and then sit back for a really long time and wait to see what the last two songs are. Sure enough, up pops "Essence" and "Right in Time" as the last two played. What do you think of my glitch hypothesis now?

According to the pretty picture, my theory has been put to the test, and perhaps surprisingly, has survived the test. But, it would be the fallacy of affirming the consequent to say that the experiment has confirmed my theory. We've already seen that if confirmation means "logically derived" from the experimental setup and results, that's exactly right. But none of this means that the experiment hasn't produced very strong evidence that the songs are not playing in purely random order.[2]

What is the best explanation of e_1 through e_8 when we add the positive experimental result below?

> e_9. When Johnson tried the shuffle all songs routine for Lucinda Williams, his most listened to songs did indeed come up last.

All of the rivals we thought of with Pink Martini are still possible, but hardly any seem plausible any longer. One of the most seriously misleading features of the pretty picture is that it sets up an asymmetry between experimental confirmation and disconfirmation. We've seen why as a matter of deductive logic why this asymmetry exists. But no such asymmetry exists when we see experimental results as additional data that the tested theory and its rivals must explain.

THE SIGNIFICANCE OF NEW DATA

One of the remarkable things about the natural sciences is that we can devise experiments and go looking for highly relevant new data. But new data can cause us to rethink our evidence, or feel even more confident about it, in any of the arguments we've been thinking about, not just the scientific ones. If we find out that Dick's been in the hospital with pneumonia, and that he loaned his car to his buddy, Sam, things are going to seem much more promising for Dick and Jane. If we find a copy of Sarah's midterm on Charlie's laptop, the case for cheating is obviously strengthened. And if a law journal article surfaces that was published before Ms Daily's with the same wording, it's looks bad for both Mr. O'Neill and Ms Daily.

Three very important things follow from all of this. The first is that evidence evaluation is always relative to what we presently know. If we learn new things, and assemble them in new arguments, there will be times when our original conclusion will be strengthened, times when it will be weakened, and times when it will be pretty much left untouched. The second is that new data is ***always*** possible. The fact that we could imagine rival explanations means that we can imagine new evidence for these rivals. But this last fact leads to our third moral. Just because new data is possible, does not mean that our assessment of the current evidence is unreliable. If all the rivals are farfetched, then the

chances of finding new data that supports them is pretty slim. We do, of course, need a certain kind of intellectual modesty. We concede that things ***could change*** on the basis of new discoveries. But at the same time, for some kinds of evidence, we can be confident that they ***won't change***.

Endnotes

[1] The on-line *Merriam-Webster Dictionary* defines *ceteris paribus* as, "if all other relevant things, factors, or elements remain unaltered."

[2] It is, of course, true that digital computers like iPods do not truly generate anything randomly. But their random number generating algorithms simulate randomness for all practical purposes.