

Chapter Four

CONFIRMATION AND NEW DATA

[Positive economics'] task is to provide a system of generalizations that can be used to make correct predictions about the consequences of any change in circumstances. Its performance is to be judged by the precision, scope, and conformity with experience of the predictions it yields. In short, positive economics is, or can be, an "objective" science in precisely the same way as any of the physical sciences.

--Milton Friedman

1. A Pretty Picture of Science

The key to good science is objectivity. This does not mean that scientists are inhuman, perfectly objective, robots. Objectivity is a product of a procedure -- the scientific method -- not the individuals who carry out the procedure. No matter how much the scientist may want her theory to be true, it will only be accepted by other scientists if it has passed some very stringent empirical tests. It is the theory's success, or failure, in these experimental tests that determine its scientific fate. As with most pretty pictures, this one contains some insight and some truth.

What follows below are two examples -- actually they were used as exercises -- from an excellent recent book, *Understanding Scientific*

Reasoning, by Ronald Giere. First one of scientific success.

The Discovery of Neptune

During the first half of the nineteenth century, astronomers were still working out tables and charts giving the positions of the various planets. In this, they were aided by Newtonian, theoretical models. But then the outermost planet, Uranus, caused some difficulties. Its observed orbit differed from what it should have been according to the then best-fitting Newtonian models. The difference was much too great to be attributed to inaccuracies in measurement. They were forced to conclude that their current models were not correct. They did not, however, give up Newton's theory of celestial mechanics. By that time, there had been so many successful predictions using Newtonian models that they were reluctant to conclude that the general theory could be wrong. Around 1843, the English astronomer J. C. Adams, and somewhat later, the French astronomer Leverier, independently calculated that the observed orbit of Uranus could be explained if there were an additional planet beyond Uranus whose gravitational force produced the deviations from the earlier Newtonian predictions--which, of course, assumed no such planet. Using this more elaborate Newtonian model, Adams and Leverier were able to calculate just where the planet should be at any particular time. The planet, named Neptune by Leverier, was observed in 1846, just where it was predicted to be.

And now a briefer story of scientific failure.

The Missing Planet: The Story of Vulcan

Like the orbit of Uranus, the observed orbit of the innermost planet, Mercury, failed to fit Newtonian models by amounts that could be reliably measured. Fresh from their discovery of Neptune, many astronomers immediately assumed that there must be yet

another planet, closer to the sun than Mercury. Leverrier even named that new planet Vulcan and calculated just where it should be. Although several people claimed to have seen Vulcan, these reports were never substantiated.

The pretty picture of science says that the objective scientist will go out and observe the world. She will find something puzzling, and try to figure out a way of understanding what's going on. She will, if she's lucky and smart enough, propose a theory (or hypothesis, or explanatory account). Here's where the method gets interesting. Her theory, no matter how interesting, potentially profitable, or useful, will have to first be put to experimental tests. Only if it passes these tests, can it be accepted as scientifically respectable.

Both the Uranus theory, and the Vulcan theory, fit the pretty picture. Astronomers are puzzled as to why the observed orbits of Uranus and Mercury fail to fit the Newtonian model. The existence of new, unobserved planets -- Neptune and Vulcan -- are proposed as solutions to the original puzzles. The potential existence of new planets suggests some obvious experimental tests. If they are really there, the Newtonian model should allow us to predict where they will be, and we should ultimately be able to observe them. The Neptune theory passed this test beautifully, and the Vulcan theory crashed and burned.

2. *The Positivists' View of Science*

In the early decades of this century philosophers and scientists attempted to formalize what I have been calling the pretty picture of science. The school of *logical positivism* articulated a view of proper scientific method that dominated thinking about good science until very recently. Logical positivism is somewhat out of date, but its insistence that the most important test of scientific theories is the deductive prediction of empirical consequences in experimental circumstances remains highly influential. According to logical positivism science seeks to discover general regularities in the natural and social world. The potential laws can be confirmed or falsified by straightforward empirical methods. And once these laws are confirmed and accepted within the science they can be used as a source of genuine scientific explanations. The heart of both the *hypothetico-deductive* model of confirmation, and the *deductive-nomological* model of explanation (as the names would imply) is a simple view of deductive syllogism.

The *disconfirmation* (or falsification) of the Vulcan theory might be reconstructed in something like the following way. The existence of the new planet, along with some theoretical calculations from the Newtonian theory *predicted* the observation of a planet at a particular time and place. We can represent this thought in terms of an *if, then* statement, or a *conditional*. Letting V stand for the statement, "Vulcan exists," and O for "We will observe a planet at a particular time and place," we get the following.

$V \supset O$ Conditional stating prediction

Our experiment will either yield positive results, we will observe what was predicted, O, or it will produce negative results, we will fail to observe what was predicted, $\sim O$. As we saw, in the actual historical case, the results were negative. This gives us the following, deductively valid, inference.

$V \supset O$ Conditional stating prediction

$\sim O$ Experimental results

————

$\sim V$ Disconfirmed theory

The "tilde" symbol, \sim , means, "it is not the case." Thus, our experiment deductively "proves" that Vulcan does not exist. Or, so a simple-minded application of logical positivism would lead us to believe.

There is an interesting difference when the experimental results are positive and lead to the *confirmation* of a theory. Letting N stand for "Neptune exists," we get the following.

$N \supset O$ Conditional stating prediction

O Experimental results

————

N

Unfortunately, this inference is *invalid*; it commits the fallacy of *affirming the consequent*. [To see what's going on here, try the following. If today is Tuesday, tomorrow is a weekday; but tomorrow

is not a weekday; therefore today is not Tuesday. This inference is valid. But, notice what happens when we drop the "not." If today is Tuesday, tomorrow is a weekday; but tomorrow is a weekday; therefore today is Tuesday. Invalid, today could just as well be Monday, Wednesday, or Thursday.] Since, N, the truth of the Neptune theory does not follow from a deductively valid argument, positivists speak of it as *partially confirmed*. The scientist has risked having his theory *disconfirmed*, it has passed a tough test.

3. *Some Problems for Logical Positivism*

However insightful the positivist's characterization may be, it seems to admit serious counter-examples from within the natural sciences. In the strict sense of a controlled laboratory experiment, this methodological restriction would rule out most of theoretical physics, cosmology, evolutionary theory, and plate tectonics in geology. All of these theoretical frameworks are designed primarily to explain data we already have, and not predict new data to be discovered in the future. Of course, when the theory does yield a correct prediction -- such as the discovery of the background radiation that helped confirm the big bang cosmology -- that is all to the good. Certainly one thing that good science does is to figure out testable consequences that are implied by interesting hypotheses, and then do the empirical work to see whether those consequences are born out. But to suggest that this procedure is either necessary or sufficient for good science is simply to misread the classical and contemporary history of science.

Positivists claim that scientific reasoning is fundamentally deductive. The first problem, and in many ways the most striking, concerns the formulation of law-like generalizations. There is no pretense that this fundamental step in the process is in any way deductive. Some positivists believed that non-deductive, but systematic procedures -- induction by enumeration -- were the underlying basis for the discovery of causal regularities. Unfortunately, even the most superficial acquaintance with actual procedures in the empirical sciences shows how artificial the imposition of systematicity at this stage of the process is. Thus, even as committed a positivist as Hempel is forced to admit that:

[t]here are, then, no generally applicable "rules of induction", by which hypotheses or theories can be

mechanically derived or inferred from empirical data. The transition from data to theory requires creative imagination. Scientific hypotheses and theories are not *derived* from observed facts, but *invented* in order to account for them. They constitute guesses at the connections that might obtain between the phenomena under study, at uniformities and patterns that might underlie their occurrence. "Happy guesses" of this kind require great ingenuity.

The positivists' models of confirmation also demand re-examination. A given set of experimental conditions, and a given experimental result, always confirms an infinite number of potential predictions. In the Neptune case our experimental finding was the observation of a planet at the time and place predicted. These results also "confirm" the Magic Planet theory which predicts that a planet will be observed at the right time and place in even years, but never in odd years. As a matter of pure logic, an infinite number of these bizarre relationships can be constructed, all of which are "confirmed" by our background assumptions and our experimental results. When we choose the preferred hypothesis as the one confirmed we do so because of elegance, or completeness, or simplicity, or some other "pragmatic" criterion, and not as a matter of deductive logic.

Things get even worse when we consider falsification. According to the the positivists' model, if astronomers had failed to observe Neptune for another several decades, this would have decisively falsified the Neptune theory. This is clearly not true. We could just as well take the negative result as disconfirming our assumptions about the experimental circumstances. All theories and experiments have implicit *ceteras paribus* clauses. Perhaps, in most general terms, all things are not equal in our experimental circumstances. We always can treat the negative result simply as an unexplained anomaly. We can write it off as due to imprecise measuring techniques, or statistical fluke, or a number of other perfectly reasonable alternatives.

4. The Strange Case of the Childbed Fever Deaths

Hempel offers an extended discussion of the work of the great physician and medical researcher Ignaz Semmelweis and his work on childbed fever. I fully agree that this case study of scientific discovery represents "a simple illustration of some important aspects of scientific

inquiry." For Hempel the case is a paradigm of hypothetico-deductive reasoning. I want to use it as example of a detective's reasoning, and to that end I offer the following parody of the genre.

The call came in late; these kind always do. Detective Semmelweis was stuck with the graveyard shift at the robbery-homicide desk. A deep voice of indeterminate age and gender reported that things were not as they should be at the First Maternity Division of the Vienna General Hospital.

The smell of death hovered over this temple of public health, 1840's style. Over two-hundred and fifty women had died in 1844, 8.2 percent of the maternity population. The percentage dipped the next year, down to 6.8, but skyrocketed in 1846 to 11.4. Clearly a homicidal maniac was on the loose! This would be the toughest case of Semmelweis's career.

A careful survey of the crime scene produced a number of intriguing clues. The most interesting of which all centered on the Second Maternity Division. Next door the death rate from childbed fever ranged between 2.0 and 2.7 percent during the same three year period. In the First Division examinations and deliveries were performed by medical students, in the Second by midwives. Both divisions were seriously overcrowded, but because there was an obvious interest in being assigned to the Second Division, it was even more jammed than the First Division. Semmelweis noted with some interest the fact that in the First Division deliveries were performed in the supine position, while in the Second they were lateral. Finally, there was the strange appearance of the death bed priest. In the First Division he had to proceed through five wards, accompanied by ringing bells, in order to administer the last sacrament to the next victim. In the Second Division the priest had direct access to the sickroom. In other relevant respects, however, general care, diet, et cetera, the two Divisions were identical.

Other interesting entries in the detective's notebook included the following observations. Women who delivered in transit to the First Division -- the "street births" -- and were subsequently admitted had much lower death rates. Also was the curious fact that those babies who contracted the disease almost all came from mothers who were already ill. The perpetrator was as shifty as he was bloodthirsty.

Semmelweis ordered that the usual rogue's gallery of suspects be brought down to the station house and be given the third degree. These low-lifes all seemed to have airtight alibis. The most ruthless of the bunch, of course, was airborne infection. He looked to be clean, however, since it was hard to see how he could break into the First

Division without disturbing things in the Second Division, and the rest of Vienna for that matter. Although things were at epidemic proportions in the First Division, the death rate was too low in the immediate surroundings. Another serious suspect was overcrowding; he had a rap sheet as long as Semmelweis's right arm. But he cleared himself by pointing out that if his gang of thugs was involved, their fingerprints would be all over the Second Division.

Some of the boys in the squad room were convinced that overly rough examinations by the medical students had to be responsible. Semmelweis didn't buy this. For one thing, the midwives next door used much the same examination techniques, and besides, the natural birth process was much more traumatic than any examination. But the clincher was that rough examinations avoided the trap they set. The First Division cut its medical student staff in half, and ordered that examinations be kept to a minimum. You guessed it, after a momentary dip in the death rate, it shot up higher than ever before.

One of the most curious suspects was the death bed priest. The proposed *modus operandi* went something like this. The sight of the priest as he marched through the five wards, literally with bells ringing, was so terrifying to the patients that they were contracting the disease. Semmelweis offered the priest a deal; by altering his behavior he could clear himself. He agreed to enter the sick room by back corridors, and without the bells, thus avoiding the other patients. He had to be scratched from the list of viable suspects when the murders went on just as before.

One final suspect seemed very unlikely. But, it was so easy to set a trap for him, Semmelweis would have been crazy not to give it a shot. Perhaps birthing in the supine position was the culprit. Semmelweis was sure that if this was so, the perpetrator would tip his hand with an order to institute lateral deliveries in the First Division. The change in delivery position, unfortunately, had no effect of the mounting deaths.

After months on the case Semmelweis was no where near making an arrest. This string of grizzly murders was beginning to look like a good candidate for the unsolved case file. Fortunately, even the craftiest and most diabolical of killers occasionally slip up. The break in the case came when a physician in the First Division received a laceration from the scalpel of a medical student as they were performing an autopsy. The doctor became ill, and died from a disease with strikingly similar symptoms to childbed fever. Semmelweis reasoned that this death was caused by "cadaveric matter" introduced into the bloodstream that resulted in a kind of blood poisoning.

In an instant the detective saw the killer's plan in all its horrible detail. The crucial clue was in his notebook all along. The physicians and medical students routinely performed autopsies, and would often perform examinations immediately following, with only the most superficial cleansing of their hands. Here was the relevant difference between the two maternity divisions. The midwives in the Second Division did not perform autopsies! Other things fell into place as well. It was now easy to see why the women who delivered on the way to the First Division and were only later admitted had a better chance of escaping the killer. Also beautifully explained was the fact that the newborn infants in the First Division who occasionally contracted childbed fever came almost exclusively from mothers who were already suffering from the disease. Obviously the common circulatory system was responsible.

The trap to catch cadaveric matter was easy to set. If it was being introduced into the women's bloodstreams from the hands of the doctors and medical students, the murders could be stopped by insuring that those who performed examinations thoroughly disinfected their hands. Indeed, Semmelweis issued an order requiring disinfection in chlorinated lime before all examinations. The results were startling. In 1848 the death rate from childbed fever was down to 1.27 percent, even lower than the Second Division.

The arrest, arraignment, and subsequent conviction followed as a matter of course. The case did not end here, however. It turned out that cadaveric matter had an accomplice. He might have gotten off scot free except for Semmelweis's sharp eye. A woman suffering from cervical cancer was examined in the First Division. Twelve other women were subsequently examined by the physicians, and eleven of them contracted and died from childbed fever. Although disinfection in chlorinated lime was now standard practice after autopsies, no such precaution was taken between examinations. Semmelweis now saw that "putrid matter derived from living organisms" was also in on the murders.

The killers, sadly, managed to get their revenge. Semmelweis was lacerated during an autopsy, and died from childbed fever.

5. Inference to the Best Explanation and New Data

Our detective story can be productively retold from the perspective of inference to the best explanation. Semmelweis, in his investigation of childbed fever, was confronted with a great deal of initial data.

- e₁. Childbed fever death rates in the First Division were significantly greater than in the Second Division.
- e₂. Women who experienced "street birth" on the way to the First Division had lower death rate.
- e₃. Priests had direct access to sick room in the Second Division.
- e₄. Second Division was more overcrowded than the First.
- e₅. The two divisions were virtually identical in terms of general care and diet.
- e₆. Midwives in Second Division used same examination techniques as medical students in the First.
- e₇. First Division deliveries were in the supine position, Second Division deliveries were lateral.
- e₈. General care, diet, etc., were the same in both divisions.

Semmelweis begins by considering four causal explanations, two of which were assigned extremely low initial plausibility simply on the basis of the data at hand.

- t₁. Airborne transmission is the cause of childbed fever.
- t₂. Overcrowding is the cause of childbed fever.

We might speak of these as being incompatible with the data. After all, Semmelweis notes that airborne transmission would surely infect the Second Maternity Division as well as the first, especially since they are right next door. And the Second Division was even more overcrowded than the first, thus ruling out overcrowding as the relevant causal factor. Neither of these hypotheses logically contradict the data; some

appropriately baroque story could be told showing the logical possibility of t_1 or t_2 without denying any of the data. What happens, I believe, is that it is obvious to the casual observer that these hypotheses have a very difficult time explaining all of the data. Each faces a particularly embarrassing fact -- what Wright calls an explanatory hurdle -- that can only be accounted for by an extremely complicated and *ad hoc* explanatory story.

The two other explanatory rivals that Semmelweis begins with face similar explanatory hurdles, but can be so easily tested that the plausibility judgment is delayed until the new data can be gathered.

- t₃. The "terrifying and deliberating effect" of the priest passing through the First Division was the cause of the disease.

But that was easy to test. By convincing the priest to alter his routine, Semmelweis gained some relevant new data.

- e₈. First Division deaths did not decrease by the change in the priest's behavior.

Thus, another causal account had to be rejected. A similar test was possible for the fourth possible hypothesis.

- t₄. Delivery in the supine position was the cause of the childbed fever epidemic in the First Division.

This possible cause was eliminated by the introduction of the lateral position in the First Division, and the consequent accumulation of more new data.

- e₉. The change to lateral position delivery had no effect on death rate.

Positivists tell a tale of deduction. The priest hypothesis entails that the change in behavior will result in a declining death rate. The empiricist is handed an observable, testable prediction. When the death rate remained high, this entailed, by the rule of *modus tollens*, that the hypothesis was falsified. Inference to the best explanation allows us to bypass all of this. What we have is relevant new data. Merely articulating the possible explanation was enough to suggest a

straightforward way to go find some things out. We realized -- not deductively, but through explanatory reasoning -- that if the death rate went down after the change in the priest's route to the sickroom, every other hypothesis would have a very difficult time accounting for this. Likewise, t_3 has a very difficult time explaining e_8 . Neither t_3 nor t_4 are logically falsified, nor disconfirmed, on the basis of the new data we have gathered during the testing phase of the investigation. They are simply assigned very low plausibility rankings, as things stand at this stage in the investigation.

Significant new data was uncovered quite by accident. A colleague received a laceration during an autopsy, and died from symptoms very similar to childbed fever.

e_{10} . Circumstances and symptoms of colleague's death.

All of this suggested an exciting new hypothesis.

t_5 . The cause of the disease was the introduction of "cadaveric matter" into the bloodstream.

The test of this hypothesis was to introduce more sanitary practices in the First Division following autopsies. Once again we have classic experimental design. Circumstances are so structured that we are confident that new data will be uncovered that is directly relevant to the explanatory controversy at hand. t_5 will be significantly weakened if the death rate remains unchanged; on the other hand, all of the other rivals will have a very difficult time accommodating a decline in the death rate. Indeed, the new data confirmed the hypothesis.

e_{11} . Following the order of washing in chlorinated lime after autopsies, the death rate fell dramatically.

The investigation might have ended there, but fortuitous new data presented itself. A woman with cervical cancer was examined. Without disinfecting their hands, examinations were subsequently performed on several other women; almost all of them died.

e_{12} . Details concerning cancerous examination.

From all of this data, Semmelweis concluded that t_5 was incomplete and

needed to be expanded.

- t₆. Childbed fever was caused by "cadaveric matter" as well as "putrid matter derived from living organisms."

The moral to this story is straightforward -- always be on the lookout for new data. Sometimes new data will be engineered in a carefully constructed experiments, like Simmelweis' altering of the priest's path into the sickroom. Other times it will be discovered serendipitously, like the colleague's bad luck with the autopsy. Sometimes new data will strengthen a theory like the experiment with the chlorinated lime, while other new data, like the change in birthing position, will hurt a specific hypothesis. Occasionally, like the case of the cancerous examination, it will suggest completely unthought of new theories that will then need to be tested. New data is always possible, and on the basis of new data, new rank orderings of explanations may be called for. Good evidence, therefore, is merely a static judgment in the middle of a dynamic process.