Science Reasoning

Supplement

to

The Pocket Guide to Critical Thinking

Richard L. Epstein

Distributed *FREE*

Advanced Reasoning Forum

Acknowledgements
I am grateful to four students who read and commented on an earlier draft: Lauren Smith, Thaddeus Sonnenfeld, Sarah Wilson, and Amber Woodyatt. I am also grateful to Carolyn Kernberger who collaborated on some of this material in earlier publications.

For more information contact:
Advanced Reasoning Forum
P. O. Box 635
Socorro, NM 87801 USA
www.AdvancedReasoningForum.org

ISBN 978-0-9815507-9-4

Science Reasoning

supplement
to

The Pocket Guide to Critical Thinking

Chapter numbers refer to chapters in
The Pocket Guide to Critical Thinking.

Preface

Science is a way of understanding the world. To be a scientist is, above all, to make claims that you're willing to test, to leave open to disproof, to make public. Science is a public kind of knowledge and a verbal kind of knowledge, based on specific skills of reasoning.

Those skills are not an arcane mystery that requires years of practice in science to master. They are skills we use and need every day in our lives, refined and applied carefully to the subject matter of each science. We can and should master them before we begin studying deeply any science.

Those basic skills of reasoning are presented in *The Pocket Guide to Critical Thinking*. In this supplement you'll see applications of them to many sciences. The skills build bit by bit, just as in the study of any science, from the most basic to the most complex, from analyses of claims and arguments to the heart of scientific reasoning: generalizing, cause and effect, explanations, models and theories.

But unlike the study of a particular science, you can't master the ideas bit by bit, being sure of your understanding of each part before going on. Reasoning well requires judgment, and that can be developed only slowly with experience as you see the ideas of each chapter applied in many different contexts through many examples. As you come to the final chapters, the scope of how to reason should begin to be clear and you should be able to read and understand better the material you encounter in your study of any science you choose.

2 Definitions

Science examples

Example 1 Zoe: This ecology stuff is driving me crazy. Did you see how our professor wasn't sure whether he should call that plant a tree or a bush? He said it was "borderline." What kind of answer would that be on a test?

Analysis It looks like Zoe is asking for more precision than is reasonable. Nothing deep or important hangs on whether a particular plant is classified as a bush or a tree. But if it does, a botanist could arbitrarily stipulate that a tree is any plant that, say, typically has a main trunk with diameter greater than 40 cm.

Example 2 "The so-called 'respiratory center' is a widely dispersed group of neurons located bilaterally in the reticular substance of the medulla oblongata and pons."

Arthur C. Guyton, *Basic Human Physiology*

Analysis For people who work in this area this is a good enough definition, though for most of us the words doing the defining aren't clearer and better understood than what's being defined. What counts as a good definition may depend on the background of the people who are meant to be using it.

Example 3 "A true binge is an episode of eating marked by two particular features. First, the amount eaten is, by 'normal' standards, excessively large. Second, the eating is accompanied by a subjective sense of loss of control."

P. J. Cooper, *Bulimia Nervosa and Binge-Eating*

Analysis Given the variability in human behavior, this is good enough for most classifying work in this area.

Example 4 "Communication is the process by which the behavior of one animal affects the behavior of others; that is, it changes the probability distribution of other animals' behavior."

Claud A. Bramblett, *Patterns of Primate Behavior*, 2nd ed.

Analysis This is a bad definition because it's too broad. For example, a dog walks by a tree, and a squirrel she isn't aware of on another tree hides in its hole. You can call that something, but you can't call it "communication" because that word has a clear use that requires an intention to affect others' behavior.

Example 5 Communication is when one animal intentionally affects the behavior of another.

Analysis This seems to assume that animals other than humans have intentions, which would make it a persuasive definition. But it could be used without that assumption in order to investigate whether animals do communicate.

Example 6 "Since the time of thalidomide, it has become widely recognized that drugs consumed by a mother during pregnancy can alter the development of the fetus. Drugs that cause such malformations are called teratogens (literally, 'monster makers')."

William A. McKim, *Drugs and Behavior*: *An Introduction to Behavioral Pharmacology*, 5th edition

"Teratogens are environmental agents (such as drugs or viruses), diseases (such as German measles), and physical conditions (such as malnutrition) that impair physical development and lead to birth defects and even death."

Richard A. Griggs, *Psychology*

Analysis These are very different definitions. The first suggests that only drugs are teratogens, while the second includes a wider variety of "agents." Both inadequately clarify what counts as the kind of defect: "such malformations" is too vague, and "birth defects" is probably too wide, including minor conditions such as tiny birthmarks.

Example 7 Eugenics is the science of improving the human gene pool by selective breeding.

Analysis This is a persuasive definition because it assumes that eugenics is a science.

Example 8 "Species are groups of interbreeding natural populations that are reproductively isolated from each other."

E. Mayr, *Populations, Species and Evolution*

Analysis Consider two kinds of crows in Europe: the black crow and the hooded crow. The former is completely black, while the latter is part black and part grey. These used to be called different species, but some intermediate forms occur due to interbreeding in various places; biologists now classify the two kinds as subspecies. It seems that they take the word “isolated” in Mayr’s definition to mean “total isolation”; by that definition species can’t interbreed.

Other biologists point to species that are highly isolated but do interbreed, just in very restricted areas and only rarely. They are interpreting Mayr’s definition in terms of relative isolation. For them, species can interbreed.

The “deep” question about whether species can interbreed turns into a question about how to define “species.”

3 Arguments

Science examples

Example 1 Maria mixes some chemicals in a flask. She gets a new compound that is pink, just as she predicted. She goes to Lee and says, "See, I was right" and points to the liquid in the flask.

Analysis Maria is attempting to convince Lee. But not every attempt to convince is an argument. Pointing by itself isn't convincing: Lee has to know what Maria is pointing at and what claim she is trying to show is true, and why the pointing matters. The attempts to convince that we are interested in are those that are in language, the language of truth and falsity: claims.

Example 2 "The feeding strategies of amphibians include their choice of prey and the ways in which they locate, capture, and ingest prey. Amphibians generally are considered to be feeding opportunists with their diets reflecting the availability of food of appropriate size. This may be true for some, but results of field and laboratory studies show that some species are selective in their feeding."

W. E. Duellman and L. Trueb, *Biology of Amphibians*

Analysis The authors of this textbook intend for you to believe what they're saying. But they present no argument here, no reason to believe the claims they are making beyond their supposed authority.

Example 3 "The causative agents of some diseases may be found in all parts of the carcasses of animals dying from those diseases. Therefore, carcasses should be disposed of as quickly and thoroughly as possible, by burning, sterilizing with heat, or burying deeply to prevent feral animals from disturbing the carcasses and further spreading the disease."

J. H. Galloway, *Farm Animal Health and Disease Control*

Analysis This is an argument: The word "therefore" cues us to that. The conclusion is "Carcasses (of animals dying from those diseases) should be disposed of as quickly and thoroughly as possible, by burning, sterilizing with heat, or burying deeply." One premise is "The causative agents of some diseases may be found in all parts of the carcasses of animals dying from those diseases." Another premise is

"Disposing of carcasses prevents feral animals from disturbing the carcasses of animals dying from diseases and further spreading the disease" because that's meant as a reason to believe the conclusion. There's some rewriting here, but it doesn't change our understanding of what's said; it just allows us to see the organization of the argument better.

4 What Is a Good Argument?

Science examples

Example 1 The liver is not the seat of the soul, as was believed by many of the ancients. The proof is that the liver can be removed and another transplanted and the person's soul remains the same. Indeed, every part of the human body can be removed, or removed and transplanted, and, except for the degradations of suffering, the person's soul remains unchanged. All that is, save one: the brain. Damage that even a little, and you will see the person's soul in throes. Thus, the brain is the seat of the soul.

Analysis This may look like a great argument. But what's meant by "soul"? If it just means "personality" the argument is good. If not, then we don't even know what the argument is about.

Example 2 "Friction is not a conservative force since the work done in, say, pushing a crate across a floor from one point to another depends on whether the path taken is straight, or is curved or zigzag. (In the two latter cases more work is required since the distance is greater and, unlike the gravitational force, the friction force is always directed precisely opposite to the direction of motion—so the work done depends on the particular path taken.)"

D. C. Giancoli, *Physics*, 2nd ed.

Analysis There are three arguments here:

Conclusion 1 Friction is not a conservative force. *Premises* The work done in pushing a crate across a floor from one point to another depends on whether the path taken is straight, or is curved, or is zig-zag. In the latter two cases more work is required.

Conclusion 2 In the latter two cases more work is required. *Premise* The distance is greater, and unlike, gravitational force, the friction force is always directed precisely opposite to the direction of motion.

Conclusion 3 The work done depends on the particular path taken. *Premise* In the latter two cases more work is required.

It's because science textbooks are written this densely that they're hard to read.

5 Evaluating Premises

Science examples

Example 1 The following appeared in a major newspaper:

> "A DePauw university computer study sometime back turned up a remarkable finding. Coeds were asked to submit their grade point averages plus their bust, waist, and hip measurements. No significance was found in the upper-body measurements. But the larger the hips, the better the grades."

That's amazing. But is it true? It was worth checking. Here's the response I got when I wrote to DePauw.

> "As far as I am aware, I think the study to which you refer can best be classified as an 'urban legend.' No such study has been done during my 12 years at DePauw, and I can find no one here who has ever heard of such a study."
>
> Larry Anderson, Director of Public Relations, DePauw University

Example 2 "Such basic skills [forging links between different sensory perceptions] are essential for children because they pave the way for more complex tasks later. To appreciate this, take the seemingly simple job of understanding what someone is referring to when they point at an object. Only other apes and dolphins are able to grasp that there's something 'over there" worth looking at, and then find the object of interest. This is one form of a skill, called joint attention, which [the robot] needs to have because so much social interaction depends on it, says Scasselati [a cognitive scientist and one of the robot's principal architects]."

"Booting up baby," *New Scientist*, 5/22/1999

Analysis I read this and said: "That's false!" Dogs understand pointing very well, if we judge by their behavior. That's how they hunt in packs, following the gaze of other dogs. With a little effort I got my dog to look at my gaze and pick out what I wanted him to fetch. It took more effort to get him to look where I pointed with my hand. Yet the prestige of the journal made me doubt myself until I went out and confirmed it again. Fetch!

Example 3 The camera lies

"Vicki Bruce of the University of Stirling and Mike Burton of the University of Glasgow tested the ability of 230 Open University students to match pictures of faces grabbed from video with still photographs of ten similar faces. The faces, which were all young, clean-shaven short-haired males, were pulled from a Home Office [British Ministry of Interior] database of 200 trainee police officers.

To their surprise, Bruce and Burton found that even in ideal conditions—using high quality pictures, full-frontal faces and neutral expressions—only 70 per cent of identifications were correct. When the face grabbed from video was smiling, the proportion of accurate matches dropped to 64 per cent. When it was shown at an angle of 30 degrees, the figure was only 61 per cent.

In a second experiment, 60 students at the University of Stirling watched short high-quality video clips of unknown faces and tried to match them with 10 photographs. Even when they were told that one of the ten photographs matched, and they could rewind and pause the video as much as they wished, only 79 per cent of the identifications were accurate.

However, other experiments showed that when faces caught on poor quality video were familiar to the student volunteers over 90 per cent of their identifications were correct. Bruce and Burton will publish their findings later this year in two American journals, *Psychological Science* and the *Journal of Experimental Psychology: Applied.*"

Rob Edwards, *New Scientist*, March 27, 1999, p. 27

Experiments

Observational claim An observational claim is one established by personal experience or observation in an experiment.

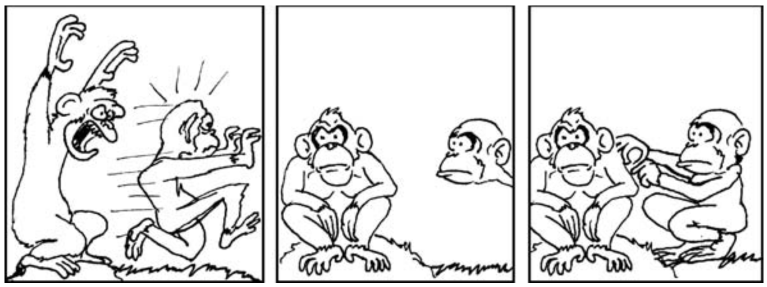
Evidence Evidence is usually the observational claims used as premises of an argument. Sometimes the term refers to all the premises.

What do we mean by "observation in an experiment"?

A physicist may say she saw an atom traverse a cloud chamber, when what she actually saw was a line made on a piece of photographic film. A biologist may say he saw the nucleus of a cell, when what he saw was an image projected through a microscope. In both cases they are not reporting on direct personal experience but on deductions made from their experience. However, those claims made by deduction from the perceptions arising from certain types of experiments are, by consensus in that area of science, deemed to be observations.

Within any one area of science there is a high level of agreement on what counts as an observational claim. But from one area of science to another that standard may vary. A physicist beginning work in biology may well question why certain claims are taken as obvious deductions from experience, such as the reality of what you see through a microscope. But after the general form of the inference is made explicit once or twice—from such direct claims about personal experience to the observational claims—he is likely to accept such claims as undisputed evidence. If he doesn't accept such deductions, he is questioning the basis of that science.

When new techniques are introduced into a science or a new area of science is developing there is often controversy about what counts as an observational claim. Galileo's report of moons around Jupiter was received with considerable skepticism because telescopes were not assumed to be accurate, and indeed at that time they distorted a lot. In ethology, the study of animal behavior in natural settings, there is no agreement yet on what counts as an observational claim, and you can find different journal articles using different standards. For example, consider:



Some would describe this as an incident of the first chimpanzee getting angry and chasing the second one away, and then the second returning to pacify and re-establish bonds with the first. That's what they saw. But others say that such a description is loaded with assumptions that have not been established, such as that chimpanzees have emotions sufficiently similar to humans to label as anger, and that chimpanzees intend to accomplish certain ends rather than operating instinctually.

One constraint we impose on reports of observations is that they should be replicable. We believe that nature is uniform. What can happen once can happen again, *if* the conditions are the same. Scientists typically won't accept reports on observations that they are unable to reproduce.

> ***Duplicable and replicable experiments*** An experiment is *duplicable* means it is described clearly enough that others can follow the method to obtain observations. It is *replicable* if when it is duplicated the observations of the new experiment are in close agreement with the observations of the original experiment.

The difficulty is to specify exactly what conditions are required and what counts as close enough agreement. It's fairly easy in chemistry and physics; less so in biology; much more difficult in psychology or ethology. It's virtually impossible in history and economics, which means history and economics are not sciences, except to the extent that we can describe very general conditions that may recur.

Some examples will illustrate these ideas.

Examples

Example 1   A recipe from a famous coffeehouse

"Vegetarian Chile

2 cans each (include liquid) of:

Pinto beans   Chili beans   Great Northern beans
Red beans   Kidney beans

1 # 10 can diced Tomatoes
Garlic   6-8 cloves chopped
Bell Pepper   1 chopped
Jalapeño Peppers   3 chopped
Chili Powder   2 soup spoons
Onions   2 chopped or in food processor
Paprika   1 soup spoon

Put in soup tureen and heat to boil for 1 hour. Take care the beans don't stick to the bottom."

Analysis   Any expert in the subject (any person who has worked in a commercial kitchen) will know what a #10 can of tomatoes is. Though "chopped" and "soup spoon" may be unclear, anyone who saw the chile being made would be able to duplicate the preparation.

Example 2   "Feeding behavior of primates

General Methodology

Data were collected simultaneously on both the activity of the animals and the forest strata at which this activity took place. Counts were made at five-minute intervals of the numbers of individuals engaged in each of the six activities and the level of the forest in which the activity was performed. The following activities were recorded: feeding—the animal actually in the process of ingesting or picking a food item; grooming—mutual and self-grooming were distinguished for certain analyses; resting—no body displacement, or feeding, or grooming, sunning, etc.; moving—movement of an individual, including individual foraging; travel—movement of the group; and other—e.g., sunning, play, fighting. These data were collected only after the animals under observation were reasonably habituated to the observer. Each observation of an animal constituted an individual activity record (IAR) collected in a given five-minute time sample. Because of the focus of the study and the difficulty in keeping continuous contact with an individual animal, no attempt was made to follow individual animals nor to collect statistical data on specific age or sex classes. Statistical analyses of the data were complicated by the fact that some of the activity records were not independent of each other. The methods used for the statistical analyses are reported in Sussman *et al*.

To determine levels of the forest, I used Richards' (1957) categories of forest stratification as a model and assigned numbers of one to five to the forest layers. Level 1 is the ground layer of the forest; it includes the herb and grass vegetation. Level 2 is the shrub layer, from one to three metres above the ground. This layer is usually found in patches throughout the continuous canopy forest, but is much more dense and is the dominant layer in the brush and scrub regions. Level 3 of the forest consists of small trees, the lower branches of larger trees, and saplings of the larger species of trees. This layer is about three to seven metres high. Level 4 is the continuous or closed canopy layer. It is about five to 15 metres high. The dominant tree of the closed canopy, at all three forests, is the kily (*Tamarindus indica*). Level 5 of the forest is the emergent layer and consists of the crowns of those trees which rise above the closed canopy. It is usually over 15 metres high.

All three forests in which I made intensive studies were primary forests and the tree layers were quite distinct. In most cases, the particular level in which an animal was observed could be distinguished easily. If I could not determine the forest level unambiguously, I did not record it.

Observations recorded in this manner may be biased because animals that are active in certain levels of the forest may be more difficult to see than those active at other levels. I attempted to minimize this problem by following a relatively small number of animals (usually from five to ten) throughout a period of continuous observation, keeping track of all the animals. For *Lemur fulvus* this usually included the whole group, which was small and, for the most part, moved together. It was more difficult to do this when observing *Lemur catta*, for which it was often necessary to follow and observe subgroups of the larger group. The larger group would disperse, especially during foraging and feeding, and during afternoon rest periods.

Day ranges were mapped by following a group from one night resting site in the morning to the time it settled in another night resting site in the evening. The location of the group was plotted throughout the day on a prepared map of the forest and the amount of time the group spent in each location was recorded. Home ranges include the sum of all the day ranges. The data on home ranges are limited, however, and probably do not represent total home ranges of the groups, since the study in each area was limited to a few months."

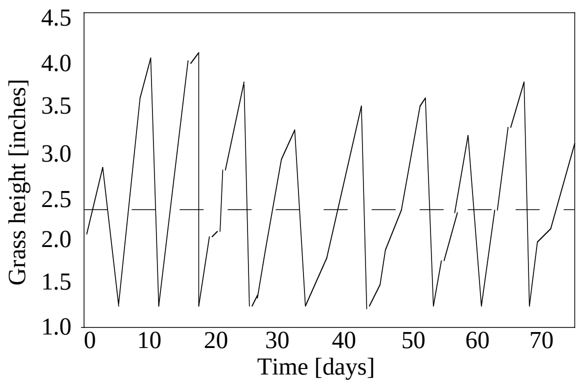
R.W. Sussman, "Feeding behaviour of Lemur Catta and Lemur Fulvus" in *Primate Ecology*, ed. T. H. Clutton-Brock

Analysis It is difficult to be more precise than this in ethology. The description of the methodology is clear enough to count as duplicable, perhaps even by someone who isn't an expert in the subject. Whether the observations are replicable will depend on how closely we expect them to agree with the ones in this paper.

Note that the author has not stated what time of year the observations were made, nor the percentage of males versus females in the groups he studied. These are not part of the conditions that need to be duplicated; implicitly, the author is saying they don't matter. If it turns out in trying to duplicate this experiment that different observations are obtained at different times of the year, then the time of year would have to be added as part of the conditions that are important and which have to be duplicated.

Example 3 "*Cyclic Variations in Grass Growth*
Grass exhibits a cyclical growth pattern surprisingly different from any other known plant. In this study, average grass blade heights have been measured, on a daily basis, over a 10 week period. Measurements were taken, utilizing vernier calipers, of the height of one hundred individual grass blades randomly chosen in a 10 foot square area positioned in front of an apartment complex in the Lexington, Kentucky area. (Measurements were also repeated with a different set of calipers to ensure reproducibility on a different apparatus.) The average of these measurements was computed and experimental error was taken as the standard deviation of the mean divided by the square root of the number of grass blades in the average. The procedure was repeated on a daily basis for a period of 10 weeks.

Figure 1: Experimental measurements of average grass height are plotted versus time. Solid line represents experimental data. Short dashed line indicates a 'constant grass height' calculation and is normalized to the experimental data to produce the best fit.



Results and Discussion The average grass heights, measured in this work, are plotted as a function of time in Figure 1. As one can readily see, there exists a periodic variation in average grass height with an approximate cycle of 7 to 10 days. Another intriguing observation is that there exists a minimum grass height, or ‘grass baseline,’ of about 1.3 inches.

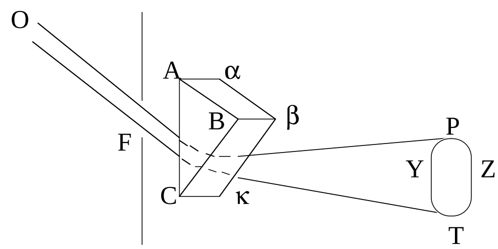
Since the cyclic period of the grass is 7 to 10 days, one may conclude that grass height varies on a ‘week-about’ basis. The physical mechanism responsible for this cyclic grass height phenomenon is not clearly understood at this time.”

V. D. Irby, M. S. Irby, Dept. of Physics and Astronomy, University of Kentucky, *Annals of Improbable Research*, Vol. 1, no. 4, 1995

Analysis The authors take great care that their experiment can be duplicated, and almost certainly it is replicable. But you should realize by now that this doesn’t make it a good experiment.

Example 4 The refraction of light rays

“In the wall or window of a room let *F* be some hole through which solar rays *OF* are transmitted, while other holes elsewhere have been carefully sealed so that no light enters from any other place. The darkening of the room, however, is not necessary; it only enables the experiment to turn out somewhat more clearly. Then place at that hole a triangular glass prism AαBβCκ that refracts the rays *OF* transmitted through it toward *PYTZ*.”



Isaac Newton, *Optica*, Part 1, Lecture 1, 1670, translated from the Latin in *The Optical Papers of Issac Newton*, ed. Alan E. Shapiro

Analysis This is very clear because of the diagram. It can be and often was duplicated, and the observations were replicable.

Example 5 Testing for anomalous cognition (*ESP*)
"The vast majority of anomalous cognition experiments at SRI [Stanford Research Institute] and SAIC [Science Applications International Corporation] used a technique known as remote viewing. In these experiments, a viewer attempts to draw or describe (or both) a target location, photograph, object, or short video segment. All known channels for receiving the information are blocked. Sometimes the viewer is assisted by a monitor who asks the viewer questions; of course, in such cases the monitor is blind to the answer as well. Sometimes a sender is looking at the target during the session, but sometimes there is no sender. In most cases the viewer eventually receives a feedback in which he or she learns the correct answer, thus making it difficult to rule out precognition [knowing the future] as the explanation for positive results, whether or not there was a sender.

Most anomalous cognition experiments at SRI and SAIC were of the free-response type, in which viewers were asked simply to describe the target. . . .

The SAIC remote-viewing experiments and all but the early ones at SRI used a statistical evaluation method known as rank-order judging. After the completion of a remote viewing, a judge who is blind to the true target (called a blind judge) is shown the response and five potential targets, one of which is the correct answer and the other four of which are 'decoys.' Before the experiment is conducted, each of those five choices must have had an equal chance of being selected as the actual target. The judge is asked to assign a rank to each of the possible targets, where a rank of 1 means it matches the response most closely, and a rank of 5 means it matches the least.

The rank of the correct target is the numerical score for that remote viewing. By chance alone the actual target would receive each of the five ranks with equal likelihood, since, despite what the response said, the target matching it best would have the same chance of selection as the one matching it second best and so on. The average rank by chance would be 3. Evidence for anomalous cognition occurs when the average rank over a series of trials is significantly lower than 3. (Notice that a rank of 1 is the best possible score for each viewing.)

This scoring method is conservative in the sense that it gives no extra credit for an excellent match. A response that describes the target almost perfectly will achieve the same rank of 1 as a response that contains only enough information to pick the target as the best choice out of the five possible choices."

Jessica Utts, An assessment of the evidence for psychic functioning
The Journal of Parapsychology, vol. 59, n. 4, p. 289, 1995

Analysis What does "All known channels for receiving information are blocked" mean? We need to know the exact layout of the room where the experiment was done. "In most cases the viewer eventually receives feedback"—how often, under what circumstances, exactly when? We need to know how close the "decoys" were to the true target. Who are the judges? This is crucial because different judges from different backgrounds may classify differently.

The experiment is not duplicable. Even if you watched the experiment being done you couldn't duplicate it, for it's not clear what the author considers important and what not important in the set-up.

Even if it were possible to duplicate the experiment and get the same results, it's not clear that by chance alone the actual target would not receive each of the five ranks with equal likelihood. Perhaps this experiment would show the opposite.

Example 6 The growth of living nerve cells in vitro
"The immediate object of the following experiments was to obtain a method by which the end of a growing nerve could be brought under direct observation while alive, in order that a correct conception might be had regarding what takes place as the fibre extends during embryonic development from the nerve center out to the periphery.

The method employed was to isolate pieces of embryonic tissue known to give rise to nerve fibres, as for example, the whole or fragments of the medullary tube or ectoderm from the branchial region, and to observe their further development. The pieces were taken from frog embryos about three mm. long, at which stage, i.e. shortly after the closure of the medullary folds, there is no visible differentiation of the nerve elements. After carefully dissecting it out the piece of tissue is removed by a fine pipette to a cover slip upon which is a drop of lymph freshly drawn from one of the lymph sacs of an adult frog. The lymph clots very quickly, holding the tissue in a fixed position. The cover slip is then inverted over a hollow slide and the rim sealed with paraffin. When reasonable aseptic precautions are taken, tissues will live under these conditions for a week and in some cases specimens have been kept alive for nearly four weeks. Such specimens may be readily observed from day to day under highly magnifying powers."

Ross Harrison, *Proceedings of the Society for Experimental and Medicine Biology,* vol. 4, 1907 (as quoted in *The Origins and Growth of Biology,* ed. Arthur Rook, pp. 159–160)

Analysis This is the first method ever recorded for maintaining living cells outside the body. It is very much like the recipe from the *Dog & Duck*. Even for an expert it would have been difficult to duplicate it from just reading this.

The morals of these examples

- It's very hard to describe an experiment clearly enough to duplicate it.
- What is described in an experiment is what needs to be duplicated. What is not described is deemed irrelevant to obtaining similar observations.
- What counts as duplicable is going to be relative to the particular scientific discipline. Expert knowledge in the area may make some descriptions very clear.
- What counts as close enough agreement for observations to be deemed replicable is going to depend on the particular scientific discipline.
- New experimental designs are often sketchily described, but they are accepted anyway because people go to the lab, see how it is done, then go back to their labs and do the experiment, and then pass that on to other people.

6 Repairing Arguments

Science examples

Example 1 "I wish that everybody could feel the confidence of being alive in a fairly benign world, because I know that most of the advice and conclusions that scientists—scientists not being very scientific about it—have made concerning our planet and the perilous journey we're on, is all bullshit. Because the place is very good at taking care of itself. We don't have to take care of it. And only by being able to read and understand the articles in places like *Science* and *Nature* and *Scientific American* can you really come to that conclusion. You're surrounded by this total barrage of faulty information that's been driven for economic reasons, information that's been basically made up for purposes of getting grants."

Kary Mullis, 1993 Nobel Laureate in chemistry,
Annals of Improbable Research, vol. 5, no. 2, 1999

Analysis It's not clear that this is an argument, though it seems that Mullis is trying to convince us of something. Nor is there reason to think that just because he's a Nobel Laureate he's an expert on this subject. When scientists spout off like this, just treat it as noise.

Example 2 "In eclipses the outline is always curved: and, since it is the interposition of the earth that makes the eclipse, the form of this line will be caused by the form of the earth's surface, which is therefore spherical. Again, our observations of the stars make it evident, not only that the earth is circular [spherical], but also that it is of no great size. For quite a small change of position to south or north causes a manifest alteration of the horizon. There is much change, I mean, in the stars which are overhead, and the stars seen are different, as one moves northward or southward. Indeed, there are some stars seen in Egypt and the neighborhood of Cyprus which are not seen in the northerly regions; and stars, which in the north are never beyond the range of observation, in those regions rise and set. All of which goes to show not only that the earth is circular in shape, but also that it is a sphere of no great size: for otherwise the effect of so slight a change of place would not be so quickly apparent."

Aristotle, *On the Heavens*, II.14.297, translated by Richard McKeon

Analysis There are two arguments here for the earth being spherical. The first is by reference to the shape of the form of an eclipse. But that is not valid or strong: the earth could be a flat disk. The second is by consideration of the position of the stars. This needs a lot of unstated premises to be good, none of which are obvious.

Example 3 Lee: See that little protozoan there? It must be alive because it's moving around.

Analysis The obvious repair to make this into a strong or valid argument is to add the claim "If a protozoan is moving around, it's alive." We can infer that from Lee's remarks; he has implied it.

Example 4 "An electron is no more (and no less) hypothetical than a star. Nowadays we count electrons one by one in a Geiger counter, as we count the stars one by one on a photographic plate. In what sense can an electron be called more unobservable than a star? I am not sure whether I ought to say that I have seen an electron; but I have just the same doubt whether I have seen a star. If I have seen one, I have seen the other. I have seen a small disc of light surrounded by diffraction rings which has not the least resemblance to what a star is supposed to be; but the name 'star' is given to the object in the physical world which some hundreds of years ago started a chain of causation which has resulted in this particular light-pattern. Similarly, in a Wilson cloud chamber I have seen a trail not in the least resembling what an electron is supposed to be; but the name 'electron' is given to the object in the physical world which has caused this trail to appear. How can it possibly be maintained that a hypothesis is introduced in the one case and not in the other?"

Sir Arthur Eddington, *New Pathways in Science*

Analysis This is a clear argument, and it's a good exercise to provide additional premises to make it valid or strong.

8 Concealed Claims

Science examples

Example 1 "Despite the fact that [Benjamin] Franklin was out of touch with the centers of European thought, his ideas on electricity were truly original and fundamental."

Gordon S. Wood, *The New York Review of Books*, 9/26/2003

Analysis That "despite" conceals a claim about where original and fundamental ideas in science come from

Example 2 "Dr. Rajendra K. Pachauri, the chairman of the United Nations Intergovernmental Panel on Climate Change (IPCC), compared Bjørn Lomborg, Danish statistician and author of *The Skeptical Environmentalist*, to Adolf Hitler in an interview with Jyllandsposten, a leading Danish newspaper (Apr. 21). Pachauri said, 'What is the difference between Lomborg's view of humanity and Hitler's? You cannot treat people like cattle. You must respect the diversity of cultures on earth. Lomborg thinks of people like numbers. He thinks it would be cheaper just to evacuate people from the Maldives, rather than trying to prevent world sea levels from rising so that island groups like the Maldives or Tuvalu just disappear into the sea. But where's the respect for people in that? People have a right to live and die in the place where their forefathers have lived and died. If you were to accept Lomborg's way of thinking, then maybe what Hitler did was the right thing.' "

Cooler Heads Coalition, April 28, 2004,
http://www.globalwarming.org/article.php?uid=637

Analysis The gross exaggeration (*hyperbole*) comparing Lomborg to Hitler makes us ready to discount the significance of anything else that Dr. Pachauri says about global warming.

Example 3 Jared Diamond in "The Religious Success Story" in the *New York Review of Books* relates how he and his "hyper-rational" classmates at Harvard in 1955 were stymied by the theologian Paul Tillich's simple question: "Why is there something, when there could have been nothing?"

Analysis This loaded question only seems mystifying until you ask, "What makes you think there could have been nothing?" To assume there could have been nothing—for which we have no evidence whatsoever in our experience—is to assume a cause for the existence of of everything.

10 Compound Claims

Science examples

Example 1 To take issue with the assumptions of evolution is to be a creationist. To be a creationist is to reject science. To reject science is to be irrational. So to take issue with the basis of evolution is to be irrational.

Analysis This is a slippery slope argument, which you can see by rewriting it using conditionals. The first premise is false.

Example 2 Lee: We had an interesting case at St. Spiridon's where I'm volunteering. We admitted a lady who's pregnant with triplets. She comes in a couple times every week, always super worried. Well, we explained to her that if she's OK, then her blood pressure and temperature would be normal, and her eyes wouldn't show any yellow. We checked all of those and they were all right, so she was OK.

Analysis Lee is affirming the consequent, confusing "if" with "only if." There are lots of other ways she could be ill.

Example 3 Saying that chimpanzees can't reason is stupid, because that would mean they couldn't make tools, which we know they can.

Analysis This is valid, the indirect way of reasoning with conditionals. But the first premise, "If chimpanzees can't reason, then they couldn't make tools" is dubious.

Example 4 If the theory of evolution is right, then the fossil record will always show transitions that are gradual. But that isn't the case. There are some periods where there are big transitions in species in little time. So evolution is just a wrong theory.

Analysis This is is valid, the indirect way of reasoning with conditionals. But it's not good because the first premise is dubious: it's likely that the fossil record is incomplete.

Example 5 Lee: You're always wondering what we do in the biology labs. C'mon in with me. We figured that that this bologna is infected with that nasty strain of Salmonella only if our lab mouse would die from injecting him with the culture we made from the bologna. Look!

Harry: Boy, he's deader than a doornail.

Tom: So the bologna was contaminated.

Analysis Lee is confusing "only if" with "if" by affirming the consequent. "Only if" = "then". The argument is weak: the mouse could have died of pneumonia.

Example 6 If this were once the bed of an ancient sea, like you claim, it would be sandy and there would be fossils of sea creatures here. But in years and years of looking here, neither I nor anyone on my team has ever found fossils. So this simply isn't the floor of an ancient sea bed.

Analysis A premise is needed for this argument to be valid or strong: "If neither I nor anyone on my team has found fossils, then it's unlikely there are fossils." From that by the indirect way we can conclude "There aren't any fossils here." Then the whole argument is valid, an example of the indirect way of reasoning with conditionals. It's good if the premises are plausible.

Example 7 Maria: Listen to this argument I read in Steen's *Practical Philosophy for the Life Sciences*,

> "If the population density of a species is high in some area, then the species will not reproduce in that area. If a species doesn't reproduce in some area, it will go extinct in that area. Therefore, if the population density of a species is very high in some area, it will go extinct in that area."

Lee: Gosh, that explains why there aren't any alligators in New York: There used to be too many of them.

Analysis Steen's argument is reasoning in a chain with conditionals and is valid and good, if the premises are true. Lee's argument is affirming the consequent and is bad: he's overlooked other possibilities.

Example 8 Air has weight

"It is no longer open to discussion that the air has weight. It is common knowledge that a balloon is heavier when inflated than when empty, which is proof enough. For if the air were light, the more the balloon were inflated, the lighter the whole would be, since there would be more air in it. But since, on the contrary, when more air is put in, the whole becomes heavier, it follows that each part has a weight of its own, and consequently that the air has more weight."

Blaise Pascal, *The Physical Treatises*, (Treatise on the Weight of the Mass of Air) in *The Origins and Growth of Physical Science*

Analysis Pascal is using the indirect way of reasoning with conditionals, and his argument is good.

11 General Claims

Science examples

General claims

Example 1 Electrons have spin.

Analysis This is the same as “All electrons have spin.” Some laws in science are meant as “all” claims, even though they don’t use “all” or “every.”

Example 2 Donkeys can breed with horses.

Analysis This is not equivalent to “All donkeys can breed with horses.” The claim is true, but some donkeys are castrated. This general law means that *it is possible* for some donkeys to breed with some horses. Only the context of the subject can tell you how you should understand a stated law of a science.

Contradictories

Example 3 Electrons never have a positive charge.

Analysis This claim is equivalent to “No electron has a positive charge.” So a contradictory is “Some electron has a positive charge.”

Example 4 “Butterflies go through the following stages in their lifetime: egg ⟶ caterpillar ⟶ pupa ⟶ adult ⟶ butterfly.”

Analysis *Contradictory*: Some butterfly does not go through these stages in its lifetime.

Example 5 When a steady current is flowing through a conductor, the strength of the current is proportional to the potential difference between its ends.

Analysis *Contradictory*: Sometimes when a steady current is flowing through a conductor, the strength of the current is not proportional to the potential difference between its ends.

Example 6 *The gas law*
$PV = kT$ where P stands for the pressure, V for the volume, and T for the absolute temperature of a fixed volume of gas, and k is a constant.

Analysis *Contradictory*: Some gas at some time and place does not satisfy $PV = kT$.

Example 7 Newton's law of universal gravitation
Every particle in the universe attracts every other particle with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between them.

Analysis *Contradictory*: Some particle in the universe does not attract some other particle in the universe with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between them.

Example 8 Given a quantity of radium, after 1620 years, approximately half the radium atoms in the quantity will have transmuted into radon atoms.

Analysis *Contradictory*: There is some quantity of radium which after 1620 years nowhere near half the atoms in the sample have transmuted into radon atoms. "Given" in science writing normally means "for any."

Example 9 When a large number of pea plants having round, yellow seeds are crossed with a large number of pea plants having wrinkled, green seeds, the second generation of round to wrinkled and of yellow to green is approximately 3:1.

Analysis *Contradictory*: Sometimes, when a large number of pea plants having round, yellow seeds are crossed with a large number of pea plants having wrinkled, green seeds, the second generation of round to wrinkled and of yellow to green is not approximately 3:1.

Arguments with general claims

Example 10 All scientists are honest. Ralph is a scientist. So Ralph is honest.

Analysis This is valid, an example of the direct way of reasoning with "all." But it's not good, as the first premise is false: a Korean group working on human cloning was exposed for publishing fake data—in good journals, too.

Example 11 All chemistry labs have a bunsen burner. The lab where Lee works has a bunsen burner. So the lab where Lee works is a chemistry lab.

Analysis This is weak, arguing backwards with "all." Lee might work in a biology lab that has a bunsen burner.

Example 12 All chimpanzees are related to simians. No lemur is related to a simian. So no lemur is a chimpanzee.

Analysis This is valid, the direct way of reasoning with “no.”

Example 13 Other than humans, only chimpanzees can communicate with signs. George is a chimpanzee. So George can communicate with signs.

Analysis This is weak: “only” does not mean “all.” George could be a chimpanzee that’s never been trained.

Example 14 All mammals have both a heart and a liver. The fossil remains of this animal show that it had a heart and a liver. So it must have been a mammal.

Analysis This is invalid, reasoning backwards with “all.” The animal could have been a bird.

12 Prescriptive Claims

Science examples

Example 1 If you do the experiment right, the liquid should turn blue.

Analysis This is a descriptive claim, not prescriptive, because "should" here means "this is what is most likely to happen."

Example 2 "So we understand the reasons for the ongoing evolution of matter toward more complex and more perfected forms. Progress is not accomplished 'actively,' guided by some predetermined 'plan' or 'goal,' but by the elimination of the least well-adapted structures by what could be called an upward leveling. Natural selection applies not only to living organisms but also to molecules, even small ones: any chemical entity exists only if the conditions in its environment allow it. The chemical development of matter was easier in that Nature did not have to show a great deal of creativity in this area."

Martin Olomucki, *The Chemistry of Life*

Analysis Science does not tell us that some forms are "more perfected" than others: that is a value judgment. Capitalizing the word "nature" and ascribing creativity to it is a bad metaphor and an odd substitute for theology. This passage has not even one descriptive claim.

13 Numbers

Science examples

Example 1 “Apples and Oranges: A Comparison

Scott A. Sandford, NASA/Ames Research Center,
Moffett Field, California

We have all been present at discussions (or arguments) in which one of the combatants attempts to clarify or strengthen a point by comparing the subject at hand with another item or situation more familiar to the audience or opponent. More often than not, this stratagem instantly results in the protest that “you’re comparing apples and oranges!” This is generally perceived as a telling blow to the analogy, since it is generally understood that apples and oranges cannot be compared.

However, after being the recipient of just such an accusation, it occurred to me that there are several problems with dismissing analogies with the comparing apples and oranges defense.

First, the statement that something is like comparing apples and oranges is a kind of analogy itself. That is, denigrating an analogy by accusing it of comparing apples and oranges is, in and of itself, comparing apples and oranges. More important, it is not difficult to demonstrate that apples and oranges can, in fact, be compared (see Figure 1).

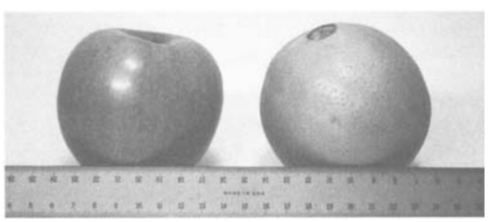


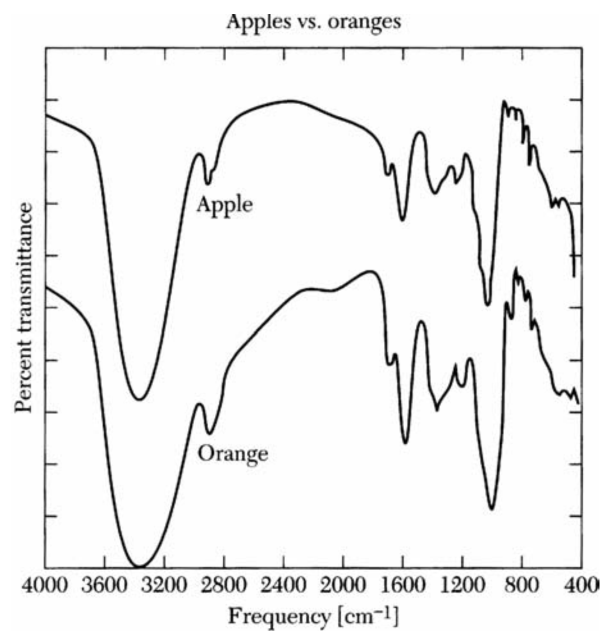
Figure 1: A Granny Smith Apple and a Sunkist Navel Orange.

Material and Methods

Figure 2 shows a comparison of the 4000–400 cm^{-1} (2.5–24 mm) infrared transmission spectra of a Granny Smith Apple and a Sunkist Navel Orange.

Both samples were prepared by gently desiccating them in a convection oven at low temperature over the course of several days. The dried samples were then mixed with potassium bromide and ground in a small ball-bearing mill for two minutes. One hundred milligrams of each of the resulting powders were then pressed into a circular pellet having a diameter of 1 cm and a thickness of approximately 1 mm. Spectra were taken at a resolution of 1 cm^{-1} using a Nicolet 740 FTIR spectrometer.

Figure 2: A comparison of the 4000–400 cm^{-1} (2.5–24 mm) infrared transmission spectra of a Granny Smith Apple and a Sunkist Navel Orange.



Conclusions Not only was this comparison easy to make, but it is apparent from the figure that apples and oranges are very similar. Thus, it would appear that the comparing of apples and oranges defense should no longer be considered valid. This is a somewhat startling revelation. It can be anticipated to have a dramatic effect on the strategies used in arguments and discussions in the future."

Annals of Improbable Research, vol. 1, no. 3, 1995

15 Analogies

Science examples

Example 1 Dick: Our diet should be similar to that of cavemen—that's what our genes are programmed for.

Analysis This is a bad analogy. There are a lot of generations between "cavemen" and us, so natural selection may have prepared us for a different diet. Besides, "cavemen" typically didn't live past forty, so their diet might be OK if you don't want to live past forty.

Example 2 We should take claims about extrasensory perception seriously. Look, suppose no one in the world had a sense of smell except one person. He would walk along a country road where there is a high stone wall and tell his friend, "There are roses there." Or he would walk into a home and say, "Someone cooked onions here yesterday." These would seem extraordinary extrasensory perceptions to his friends and acquaintances. Similarly, just because we don't understand and can't imagine a mechanism that would explain extrasensory perception, we shouldn't stop the investigation.

Analysis The analogy is bad. The person with the sense of smell would be right most of the time, in many different situations, and clearly so. No magician is going to find him out. Eventually, using physical examinations, including brain scans, we could determine to some extent the mechanism behind his predictions, even if we ourselves couldn't experience them. But to date, claims about ESP can't be duplicated, even by people claiming to have the powers; they are often debunked; they aren't right almost always, but just a bit more than average. It's not just that we have lost motivation to investigate ESP because of so many false claims about it. We haven't even found a good candidate to study. (Compare Example 5 of Experiments.)

Example 3 "Biology as it is taught in elementary schools has retained somewhat of an 'object lesson' style based on observation of animals and the mechanics of their function. This mechanistic representation of the living world actually bears the mark of an archaic cultural context, that of the late 19th century; it has aged more than that of the world of physics.

The metaphors applied to physiology at that time nearly always referred to machines. Some fifty years later, this mechanistic vision was tempered by electrical metaphors, with the brain being conceived as a very powerful telephone system. And, of course, the metaphors used in the writings of today's biologists come from computer science, with the 'software systems' of cerebral communication programming the 'hardware' of the brain, in this instance the 'wiring' of the nervous system.

No metaphor is really explanatory; rather, it reflects the cultural references through which we have been conditioned to decipher reality. But these cultural references play an important role in the way we look at the world."

Claude Kordon, *The Language of the Cell*

Analysis The brain as a machine, the brain as a telephone system, the brain as a computer. When you see such comparisons, ask whether they are really used for reasoning by analogy or whether they are, as Kordon says, only metaphors, suggestive comparisons, which may help or hinder our understanding of the subject matter.

Example 4 God must exist. The way everything works together in nature, the adaptation of means to ends, the beauty, resembles but far exceeds what humans do. Everything works together as a fine piece of machinery, like a watch. So there must be some maker with intelligence behind all of nature. That is, God exists and is similar to human mind and intelligence.

Analysis This is a bad analogy because of the differences. We determine that a watch was made by someone because it differs from what we find in nature that is not crafted, such as rocks or trees. And we can deduce from its construction that it has a purpose. We can't do that for all of nature. This doesn't prove that God doesn't exist: as always, when we show an argument is bad, we've shown only that we have no more reason to believe its conclusion than we had before.

16 Generalizing

Science examples

Example 1 Every time I or anyone else has looked into my refrigerator, the light is on. Therefore, the light is always on in my refrigerator.

Analysis Most folks will say the generalization is bad. They'll point out that when you put your finger over the little button which the door hits, you can see the light go out. That is, the sample isn't representative. But what if the refrigerator's light is controlled by a metal connection or a magnet that they can't manipulate? Why should we still believe the generalization is false? We can check the wiring diagram to see that the light is designed to go out when the door is shut.

Example 2 Every time I or anyone I know or have read about has seen a tree fall in the forest, it makes a sound. Therefore, any time a tree falls in the forest it makes a sound.

Analysis Why should we believe this generalization and not the the previous one? Our best wiring diagram (our current physics) says that the tree makes a sound. (We have to understand "make a sound" to mean "sound waves are produced" and not define "sound" as something that depends on someone hearing, for then the generalization would be trivially bad.) Modern science says that the laws of physics are the same whether there is an observer or not (though the observer might change what's being observed). But that's an untestable assumption and is not needed to do physics.

Finding a Cause

The best way to determine cause and effect is to experiment.

Example 1 I have a waterfall in my backyard in Cedar City. The pond has a thick rubberized plastic pond liner, and I have a pump and hose that carry water from the pond along the rock face of a small rise to where the water spills out and runs down more rocks with concrete between them. Last summer I noticed that the pond kept getting low every day and had to be refilled. You don't waste water in the desert, so I figured I'd better find out what was causing the loss of water.

I thought of all the ways the pond could be leaking: The hose that carries the water could have a leak, the valve connections could be leaking, the pond liner could be ripped (the dogs get into the pond to cool off in the summer), there could be cracks in the concrete, or it could be evaporation and spray from where the water comes out at the top of the fountain.

I had to figure out which (if any) of these was the problem. First I got someone to come in and use a high pressure spray on the waterfall to clean it. We took the rocks out and vacuumed out the pond. Then we patched every possible spot on the pond liner where there might be a leak.

Then we patched all the concrete on the waterfall part and water-sealed it. We checked the valve connections and tightened them. They didn't leak. And the hose wasn't leaking because there weren't any wet spots along its path.

Then I refilled the pond. It kept losing water at about the same rate.

It wasn't the hose, it wasn't the connections, it wasn't the pond liner, it wasn't the concrete watercourse. So it had to be the spray and evaporation.

I reduced the flow of water so there wouldn't be so much spray. There was a lot less water loss. The rest I figured was probably evaporation, though there might still be small leaks.

In trying to find the cause of the water leak at my waterfall and pond I was using a method scientists often use.

> *Finding a cause* Conjecture possible causes. By experiment eliminate them by showing they don't make a difference until there is only one. Check that one: Does it make a difference? If the purported cause is eliminated, is there still the effect? Is there a common cause?

I assumed there was a cause, then by a process of elimination on some conjectured causes, I fixed on one. When that occurred, the suspected effect always did, too, and it made a difference, and I knew I could fill in the normal conditions.

But why should I assume that there is a cause? Does this mean that I'm assuming everything has a cause? No, I'm assuming that there is some way to stop the leak, which in this case amounts to assuming that the leak has a cause. The assumption that a particular effect has a cause is sometimes just an expression of our desire to find a way to manipulate the world.

But doesn't this method rest on a false dilemma?

> A or B or C is the cause of E. It's not A. It's not B.
> Therefore, it's C.

No. We still have to check that C satisfies all the conditions for cause and effect, not just that it makes a difference. We must be willing to accept that our experiments will show that none of the conjectured causes satisfies all the conditions. This method cannot find the cause from nothing, but only, if we guess right, isolate it from a range of conjectured causes.

Example 2 Recently Lee found out that he has hepatitis B. None of his friends has hepatitis. He wonders how he could have gotten it.

He reasons: Since he wants to be a nurse he volunteers to work at a hospital three times per week. Some of the patients there have hepatitis, and he often washes their bedpans and comes in contact with their body fluids, though he's always careful to wear gloves. Or at least he thought he was. A recent study he read said that 25% of health care workers exposed to hepatitis B get it. So, he figures, he got hepatitis B from working at the hospital.

Analysis How strong is this argument? We start with the conjecture: "Lee contracted hepatitis B from working at the hospital." We rule out all other causes we can think of. We can imagine conditions under which he could have gotten hepatitis, but we can't specify the exact conditions that occurred that would give us the normal conditions. Eliminating all other possible causes (that we can think of) doesn't mean that we can conclude we've found the cause unless we also have:

(*) The only ways Lee could have gotten hepatitis B are P, Q, R, S, T, U, or V.

There are very strong arguments that he didn't get it from Q, R, S, T, U, or V. Therefore (reasoning by excluded middle), he got it from P. This reasoning to a cause is just as strong as (*) is plausible.

Example 3 "In my backyard, indeed throughout the neighborhood where I live, the abundance of birds is limited. In other neighborhoods there are many more birds. The most important difference I can think of concerns cats. Many cats are around where I live; elsewhere there are less of them. It is probable that there will be other differences between neighborhoods which differ in bird abundance. However, in view of background information it is reasonable to infer that cats will be a causal factor. Cats eat birds and birds are afraid of cats. An experiment could provide more confirmation. If I would shoot the cats near my place and bird abundance would subsequently increase, I would feel confident that cats do influence the abundance of birds. . . .

If an experiment of this kind were indeed performed with positive results (for the birds I mean), the evidence would be telling. However, we should realize that the situations compared—before and after the shooting—may differ in other respects. Thus it is possible that, from a bird's point of view, there happens to be a long-lasting improvement of the weather after the shooting.

In view of this the following experiment would be more decisive. Suppose we identify ten neighborhoods with many cats. We could remove the cats from five randomly chosen neighborhoods, and let the cats be in the remaining ones. If bird abundance would increase in the cat-free areas, not elsewhere, that would be something. It is improbable that the two groups of neighborhoods will systematically differ in another factor that influences birds."

Wim J. van der Steen, *A Practical Philosophy for the Life Sciences*

These examples test particular causal claims. Scientists, though, are usually interested in general causal claims, for which we have the methods of cause-in-population studies.

18 Cause in Populations

Science examples

Example 1 “Several studies indicate that people who smoke cigarettes have an increased risk for low back pain and prolapsed disk [references given]. Individuals who have not smoked for more than a year, however, do not appear to have an increased risk, as least for prolapsed lumbar disk [reference given]. Table 6 shows that current smokers have almost twice the risk for prolapsed lumbar disk as those who have never smoked or who are former smokers. In the same study [reference given] it was estimated that the risk in current smokers is increased by about 20% for every 10 cigarettes smoked per day on the average. Possible mechanisms for the association between smoking and low back pain and prolapsed disk include decreased diffusion of nutrients into the intervertebral disk among smokers [reference given], and increased pressure on the low back from the frequent coughing experienced by many smokers.

Table 6. Estimated Relative Risk for Prolapsed Lumbar Intervertebral Disk
According to Cigarette Smoking Status, Connecticut

Smoking Status	Estimated Relative Risk	95% Confidence Limits
Never smoked (referent group)	1.0	—
Current smoker (smoked in past year)	1.7	1.0–2.5
Former smoker (smoked, but not in past year)	1.0	0.6–1.7

*Relative risk = risk in those exposed to factor divided by risk in those not exposed (referent group).”

Jennifer L. Kelsy, Anne L. Golden, Diane J. Mundt
Rheumatic Disease Clinics of America, vol. 16, no. 3, 1990

Analysis The authors suggest that the cause in population studies they cite show smoking causes lower back pain. But perhaps they’ve got cause and effect reversed: people who have back pain might want to smoke to take their minds off the pain, or possibly even to alleviate the pain. Or there could be a common cause: people who do manual labor might smoke more, and the manual labor also causes back problems.

Until further cause in population studies rule out those possibilities, this is just a correlation-causation fallacy.

Example 2 *"Bad hair can give self-esteem a cowlick, study says*
People's self-esteem goes awry when their hair is out of place, according to a Yale University researcher's study of the psychology of bad-hair days. People feel less smart, less capable, more embarrassed and less sociable, researchers said in the report released Wednesday.

And contrary to popular belief, men's self-esteem may take a greater licking than women's when their hair just won't behave. Men were more likely to feel less smart and less capable when their hair stuck out, was badly cut or otherwise mussed.

'The cultural truism is men are not affected by their appearance,' said Marianne LaFrance, the Yale psychology professor who conducted the study. '(But) this is not just the domain of women.'

The study was paid for by Proctor & Gamble, which makes hair-care products. The Cincinnati-based company would not discuss how much the study cost or what they planned to do with their newfound knowledge about the psychology of hair.

Janet Hyde, a psychology professor at the University of Wisconsin at Madison who studies body image and self-esteem, said personal appearance can have an enormous effect on people, especially adolescents.

But Hyde said she was surprised to hear bad hair had a stronger effect on men than on women in some cases.

For the study, researchers questioned 60 men and 60 women ages 17 to 30, most of them Yale students. About half were white, 9 percent were black, 21 percent were Asian and 3 percent were Hispanic.

The people were divided into three groups. One group was questioned about times in their lives when they had bad hair. The second group was told to think about bad product packaging, like leaky containers, to get them in a negative mind-set. The third group was not asked to think about anything negative.

All three groups then underwent basic psychological tests of self-esteem and self-judgment. The people who pondered their bad-hair days showed lower self-esteem than those who thought about something else. . . .

LaFrance, who has also studied the psychology of smiles, facial expressions and body language, said she would continue to look into the effects of bad hair. 'We all do research that at first pass might seem quite small,' she said. 'Yes, some of my colleagues said, 'That's interesting, ha, ha.' But then, when we talk about it, people are interested.' "

Associated Press, 1/27/2000

Analysis The causal claim here is "Having bad hair causes people to lack self-confidence." The type of cause-in-population experiment is sort of a cause-to-effect controlled experiment, except the subjects weren't interviewed on days they actually had bad hair but about times when they had bad hair.

There's no reason to think the sample is representative even of students that age. There's no reason to think the subjects remembered accurately. And then there's the possibility that cause and effect have been reversed. The research is sponsored by a company that benefits from the results that were obtained, so the authority of the researcher is called into question (possible conscious or unconscious bias). And what were those other 17% of the students who weren't white, black, hispanic, or Asian? An address at a respected university is not a guarantee of good research.

Example 3 A good causal analysis from the Web
"The Japanese eat very little fat and suffer fewer heart attacks than the British or Americans. On the other hand, the French eat a lot of fat and also suffer fewer heart attacks than the British or Americans.

The Japanese drink very little red wine and suffer fewer heart attacks than the British or Americans. The Italians drink excessive amounts of red wine and also suffer fewer heart attacks than the British or Americans.

Conclusion Eat and drink what you like. It's speaking English that kills you."

Example 4 A researcher writing in the January/February issue of Australasian Science magazine reported that the Toxoplasma gondii parasite, carried by many cats, not only can harm pregnant women (as was previously known) but also can lower the IQ of men and make women more promiscuous.

News of the Weird, 1/28/2007

Analysis So cats aren't all bad: only two out of three.

The Placebo Effect

Example 5 "In studying new drugs, there is a problem that some people will report improvement of symptoms if given only sugar pills. That is why the control group is administered a ***placebo***. Neither the subjects in the experiments nor those administering the drug or placebo are told which is a placebo and which a drug—that's the definition of a ***double blind trial***.

The anecdotal and empirical accounts of the potency of the placebo effect are legion. For example, in one study, 30% of a large number of patients reported decreased sex drive, 17% increased headache, 14% increased menstrual pain, and 8% increased nervousness and irritability. These were all side effects of the administration of a placebo in a double-blind study of oral contraceptives [reference given]. In a double-blind study of a cold vaccine, 7% of patients in both groups reported toxic side effects requiring additional medical intervention. Double-blind studies will often list iatrogenic [i.e., induced by medical procedure] side effects found in the placebo group, but these symptoms will differ markedly from study to study. In contrast to the study of oral contraceptives, it is not surprising that in double-blind studies with antihistamines, fatigue and sleepiness are reported. Obviously the target symptoms monitored are different. In an antihistamine study, it is unlikely that the investigators would inquire about decreased sex drive and headaches among females."

Frederick J. Evans, "Expectancy, therapeutic instructions and the placebo response" in *Placebo*, eds. White, Trusky, and Schwartz.

Analysis There is a serious ambiguity here. What does Evans mean by "placebo effects"? If he means that the people actually had those symptoms, then those studies do not show that. The particular study he cites shows that the placebos had the effect of making people *say* they had those symptoms. That's a big difference. No one really knows how to define "placebo" and "placebo effects" in a useful way. All we can do is compare the responses of the two groups to the questions they are asked and see if there is a statistically significant difference. But what that difference means is often difficult to say.

19 Explanations

Science examples

Example 1 The gas has temperature 83ºC because it has pressure 7 kg/cm^2 and volume 807 cm^3.

Analysis Explanations that invoke a law that gives a correlation, such as this one that assumes Boyle’s law, can be construed as causal only if it is possible for some of the quantities to vary before the others, for the cause must precede the effect. So if the pressure and volume can first be increased or decreased and the temperature follows, it’s causal; otherwise it’s inferential and not causal.

Example 2 “Like multiple sclerosis, poliomyelitis in its paralytic form was a disease of the more advanced nations rather than of the less advanced ones, and of economically better off people rather than of the poor. It occurred in northern Europe and North America much more frequently than in southern Europe or the countries of Africa, Asia or South America. Immigrants to South Africa from northern Europe ran twice the risk of contracting paralytic poliomyelitis than South-African-born whites ran, and the South-African-born whites ran a much greater risk than nonwhites did. Among the Bantu of South Africa paralytic poliomyelitis was rarely an adult disease. During World War II in North Africa cases of paralytic poliomyelitis were commoner among officers in the British and American forces than among men in the other ranks. At the time various wild hypotheses for the difference were proposed; it was even suggested that it arose from the fact that the officers drank whiskey whereas men in the other ranks drank beer!

We now understand very well the reason for the strange distribution of paralytic poliomyelitis. Until this century poliomyelitis was a universal infection of infancy and infants hardly ever suffered paralysis from it. The fact that they were occasionally so affected is what gave the disease the name “infantile paralysis.” With the improvement of hygiene in the advancing countries of the world more and more people missed infection in early childhood and contracted the disease for the first time at a later age, when the risk that the infection will cause paralysis is much greater.

This explains why the first epidemics of poliomyelitis did not occur until this century and then only in the economically advanced countries."

G. Dean, The multiple sclerosis problem, *Scientific American*, 1970

Analysis This explanation is fairly strong if it's supplemented by some plausible premises. So the explanation is as good as the plausibility of the claims doing the explaining. And we need evidence for those, unless we're willing to rely on this author's word and the reputation of the magazine.

Example 3 "Consider the explanation offered by Torricelli for a fact that had intrigued his teacher Galileo; namely, that a lift pump drawing water from a well will not raise the water more than about 34 feet above the surface of a well. To account for this, Torricelli advanced the idea that the air above the water has weight and thus exerts pressure on the water in the well, forcing it up the pump barrel when the piston is raised, for there is no air inside to balance the outside pressure. On this assumption the water can rise only to the point where its pressure on the surface of the well equals the pressure of the outside air on that surface, and the latter will therefore equal that of a water column about 34 feet high.

The explanatory force of this account hinges on the conception that the earth is surrounded by a 'sea of air' that conforms to the basic laws governing the equilibrium of liquids in communicating vessels. And because Torricelli's explanation presupposed such general laws it yielded predictions concerning as yet unexamined phenomena. One of these was that if the water were replaced by mercury, whose specific gravity is about 14 times that of water, the air should counterbalance a column about 34/14 feet, or somewhat less than $2^1/_2$ feet, in length. This prediction was confirmed by Torricelli in the classic experiment that bears his name. In addition, the proposed explanation implies that at increasing altitudes above sea level, the length of the mercury column supported by air pressure should decrease because the weight of the counterbalancing air decreases. A careful test of this prediction was performed at the suggestion of Pascal only a few years after Torricelli had offered his explanation: Pascal's brother-in-law carried a mercury barometer (i.e., essentially a mercury column counterbalanced by the air pressure) to the top of the Puy-de-Dôme, measuring the length of the column at various elevations during the ascent and again during the descent; the readings were in splendid accord with the prediction."

Carl G. Hempel, *Aspects of Scientific Explanation*

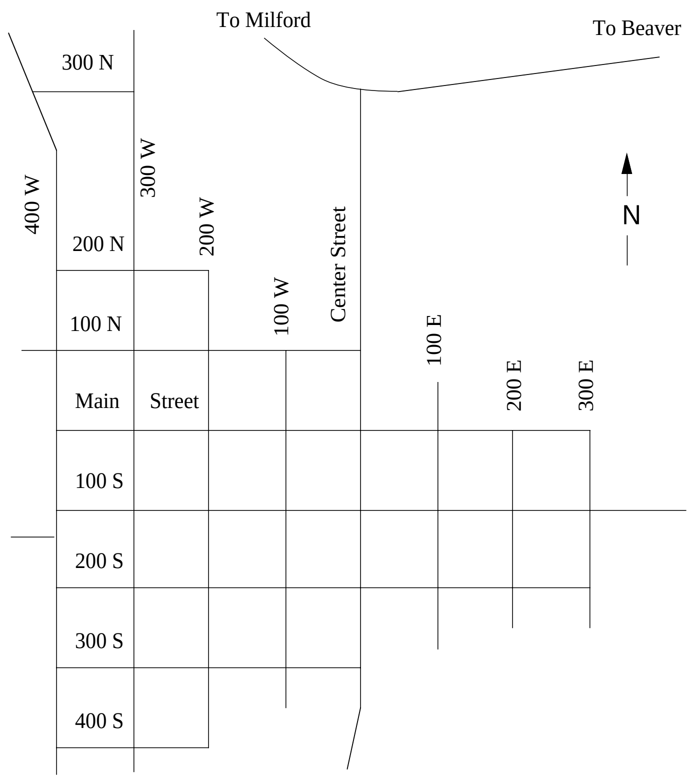
Analysis Torricelli offered an explanation, but the only evidence he had for the premise, which was a generalization, was the associated argument. So he made predictions: further instances of the generalization or of consequences of it. Those were shown to be true. The claim thus became more plausible because the associated argument for it was strengthened.

Models and Theories

What is a model? How do we determine if a model is good? How can we modify a model in the light of new evidence?

Examples of models and theories

Example 1 A map of Minersville, Utah—reasoning by analogy



Analysis This is an accurate map of Minersville, Utah. Looking at it we can see that the streets are evenly spaced. For example, there is the same distance between 100 N and 200 N as between 100 E and 200 E. The last street to the east is 300 E. There is no paved road going north beyond Main Street on 200 E.

That is, from this map we can deduce claims about Minersville, even if we've never been there. But there is much we can't deduce:

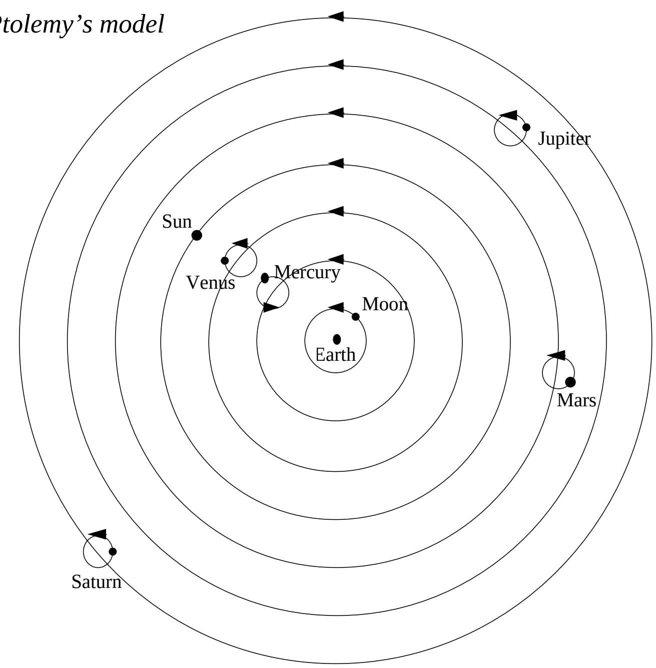
Are there hills in Minersville? Are there lots of trees? How wide are the streets? How far apart are the streets? Where are there houses? The map is accurate for what it pays attention to: the relative location and orientation of streets. But it tells us nothing about what it ignores.

The differences between the map and Minersville aren't important when we infer that the north end of 200 W is at 200 N. In contrast, a scale model of a city or a mountain abstracts less from the actual terrain: height and perhaps placement of rivers and trees are there. The map of Minersville *abstracts more* from the actual terrain than a scale model of the city would, that is, *it ignores more*.

To use this model is to reason by analogy: We can draw conclusions when appropriate similarities are invoked and the differences don't matter. The general principle, in this example, is not stated explicitly. The discussion above suggests how we might formulate one, but it hardly seems worth the effort. We can "see" when someone has used a map well or badly.

Example 2 Models of the solar system
Here is a sketch of the model of the universe the Egyptian astronomer Ptolemy proposed in the Second Century A.D.

Ptolemy's model

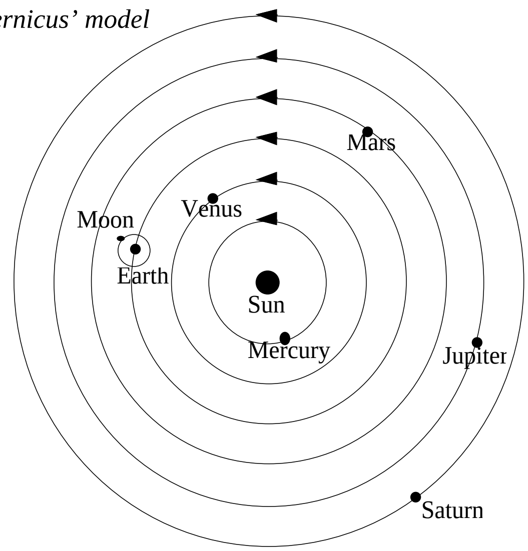


It's meant to show the relative positions of the planets, sun, and moon, and the ways they move. We can't deduce anything about, say, the size of the planets, the distances between them, nor the speeds at which they move, because this model ignores those. According to this model, each of the moon, sun, and planets revolves around the Earth in a circular orbit, all moving in the same direction. Along that orbit, each planet also revolves in a smaller circle, called an "epicycle." The sun, Earth, and Venus are always supposed to be in a line as shown in the picture.

Ptolemy made a lot more claims about the planets, Earth, and sun that were to be used in making predictions, but for our purposes this sketch will do.

Ptolemy's model accorded pretty well with observations of the movements of the planets and was the generally accepted way to understand the universe for many centuries. But in 1543 the Polish astronomer Copernicus published a book with a different model of the universe.

Copernicus' model



This sketch, too, abstracts a lot from what is being modeled. The sun is shown to be larger than the planets, but that's all we see about their relative sizes. We can't tell from the picture whether the orbits are all on the same plane or on different planes. We do see that the planets all revolve in the same direction, and that the Earth, sun, and Venus do not always stay lined up.

Ptolemy accounted for the motion of the sun, planets, and stars in the sky by saying they revolved around the Earth every 24 hours. Copernicus accounted for those motions by saying that the Earth revolved around its own axis every 24 hours. How could someone in the late 16th Century decide between these two models? Both were in accord with the observations that had been made.

In the early 1600s the telescope was invented, and in 1610 Galileo built his own telescope with a magnification of about 33 times, using it to study the skies. One of his students suggested an experiment that might distinguish between the Ptolemaic and Copernican models. Venus was too far from the Earth to be seen as anything other than a spot of light. But according to Ptolemy's model, viewed from the earth at most only a small crescent-shaped part of Venus will be illuminated by the sun. From Copernicus' model, however, we can deduce that from the earth Venus should go through all the phases of illumination, just like the moon: full, half, crescent, dark, and back again. Galileo looked at Venus through his telescope for a period of time and saw that it exhibited all phases of illumination, and this he took to be proof that Copernicus' model was correct.

Not a lot of other people were convinced, however. Telescopes were rare and not very reliable: they introduced optical illusions, such as halos, from the imperfections in the glass and the mounting. Why should astronomers have trusted Galileo's observations?

It was more due to Newton that something like Copernicus' model of the universe was finally accepted. Newton deduced from his laws of motion that the orbits of the Earth, sun, and the planets would have to be ellipses, not circles. And the distances between them would have to be much greater than supposed. Using Newton's laws, Edmond Halley predicted correctly the return of a comet that had been observed in 1682. Telescopes were better, with fewer optical illusions, and they were common enough that most astronomers could use one, so better and better observations of the planets and stars could be made. Those observations could be deduced from the Copernican-Newtonian model, while new epicycles had to be invented to account for them in the Ptolemaic model.

Note that each model is supposed to be similar to the universe in only a few respects, ones that would have an effect on how we could see the objects in the universe from the earth. Differences, such as whether Venus is rocky or gaseous, are not supposed to matter for those observations. If the model is correct, then reasoning by analogy—very precise analogy—certain claims can be deduced.

Example 3 The kinetic theory of gases—getting true predictions doesn't mean the model is true

"This theory is based on the following postulates, or assumptions.

1. Gases are composed of a large number of particles that behave like hard, spherical objects in a state of constant, random motion.

2. The particles move in a straight line until they collide with another particle or the walls of the container.

3. The particles are much smaller than the distance between the particles. Most of the volume of a gas is therefore empty space.

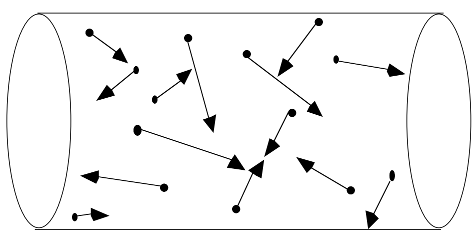
4. There is no force of attraction between gas particles or between the particles and the walls of the container.

5. Collisions between gas particles or collisions with the walls of the container are perfectly elastic. Energy can be transferred from one particle to another during a collision, but the total kinetic energy of the particles after the collision is the same as it was before the collision.

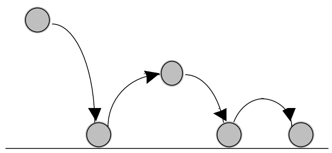
6. The average kinetic energy of a collection of gas particles depends on the temperature of the gas and nothing else."

J. Spencer, G. Bodner, and L. Rickard, *Chemistry*

Analysis Here is a picture of what is supposed to be going on in a gas in a closed container. The molecules of gas are represented as dots, as if they were hard spherical balls. The length of the line emanating



from a particle models the particle's speed; the arrow models the direction the particle is moving. The kinetic energy of a particle is defined in terms of its mass and velocity: kinetic energy = .5 mass x velocity^2. The model defines what is meant for a collision to be elastic. In contrast, here is a picture of what happens in an inelastic collision between a rubber ball and the floor.



Each time the ball hits the ground, some of its kinetic energy is lost either through being transferred to the floor or in compressing the ball.

What are we to make of these assumptions? Some are false: molecules of gas are not generally spherical and are certainly not solid; the collisions between molecules and the walls of a container or each other are not perfectly elastic; there is some gravitational attraction between the particles and each other and also with the container. How can we use false claims in a model?

The model proceeds by abstraction, much like in analogy: To the extent that we can ignore how molecules of gases are not spherical, and ignore physical attraction between molecules, and ignore . . . we can draw conclusions that may be applicable to actual gases. To the extent that the differences between actual gases and the abstractions don't matter, we can draw conclusions. But how can we tell if the differences matter?

The model suggests that the pressure of a gas results from the collisions between the gas particles and the walls of the container. So if the container is made smaller for the same amount of gas, the pressure should increase; and if the container is made larger, the pressure should be less. So the pressure should be proportional to the inverse of the volume of the gas. That is, the model suggests a claim about the relationship of pressure to volume in a gas. Experiments can be performed, varying the pressure or volume, and they are close to being in accord with that claim.

Other laws are suggested by the model: Pressure is proportional to the temperature of the gas, where the temperature is taken to be the average kinetic energy of the gas. The volume of the gas should be proportional to the temperature. The amount of gas should be proportional to the pressure. All of these are confirmed by experiment.

Those experiments confirming predictions from the model do not mean the model is more accurate than we thought. Collisions still aren't really elastic; molecules aren't really hard spherical balls. The kinetic theory of gases is a model that is useful, as with any analogy, when the differences don't matter.

Example 4 The acceleration of falling objects— an equation can be a model

Galileo argued that falling objects accelerate as they fall: they begin falling slowly and fall faster and faster the farther they fall. He didn't need any mathematics to show that. He just noted that a heavy stone dropped from 6 feet will drive a stake into the ground much farther than if it were dropped from 6 inches.

He also said that the reason a feather falls more slowly than an iron ball when dropped is because of the resistance of air. He argued that at a given location on the earth and in the absence of air resistance, all objects should fall with the same acceleration. He claimed that the distance traveled by a falling object is proportional to the square of the time it travels. Today, from many measurements, the equation is given by:

(*) $d = {}^{1}/_{2}\ 9.82\ \text{meters/sec}^2 \cdot t^2$, where t is time in seconds

Analysis The equation (*) is a model by abstraction: We ignore air resistance and the shape of the object, considering only the object's mass and center of gravity. If the differences don't matter, then a calculation from the equation, which is really a deduction, will hold. But often the differences do matter. Air resistance can slow down an object: if you drop a cat from an airplane, it will spread out its legs and reach a maximum velocity when the force of the air resistance equals the force of acceleration.

With this model there is no visual representation of that part of experience that is being described. There is no point-to-point conceptual comparison, nor are we modeling a static situation. The model is couched in the language of mathematics; equations can be models, too.

Example 5 Newton's laws of motion and Einstein's theory of relativity—how a false theory can be used

Newton's laws of motion are taught in every elementary physics course and are used daily by physicists. Yet modern physics has replaced Newton's theories with Einstein's and quantum mechanics. Newton's laws, physicists tell us, are false.

But can't we say that Newton's laws are correct relative to the quality of measurements involved, even though Newton's laws can't be derived from quantum mechanics? Or perhaps they can if a premise is added that we ignore certain small effects. Yet how is that part of a theory?

A theory is a schematic representation of some part of the world. We draw conclusions from the representation (we calculate or deduce). The conclusion is said to apply to the world. The reasoning is legitimate so long as the differences between the representation and what is being represented don't matter. Newton's laws of motion are "just like" how moderately large objects interact at moderately low speeds; we can use those laws to make calculations so long as the differences don't matter. Some of the assumptions of that theory are used as conditions to tell us when the theory is meant to be applied.

Example 6 Ether as the medium of light waves—
a prediction can show that an assumption of a theory is false
In the 19th century light was understood as waves. In analogy with waves in water or sound waves in the air, a medium was postulated for the propagation of light waves: the ether. Using that assumption, many predictions were made about the path and speed of light in terms of its wave behavior. Attempts then were made to isolate or verify the existence of an ether. The experiments of Michaelson and Morley showed those predictions were false. When a better theory was postulated by Einstein, one which assumed no ether and gave as good or better predictions in all cases where the ether assumption did, the theory of ether was abandoned.

Example 7 Euclidean plane geometry—a model that can't be true
Euclidean plane geometry speaks of points and lines: a point is location without dimension, a line is extension without breadth. No such objects exist in our experience. But Euclidean geometry is remarkably useful in measuring and calculating distances and positions in our daily lives.

Points are abstractions of very small dots made by a pencil or other implement. Lines are abstractions of physical lines, either drawn or sighted. So long as the differences don't matter, that is, so long as the size of the points and the lines are very small relative to what is being measured or plotted, we can deduce conclusions that are true.

No one asks (anymore) whether the axioms of Euclidean geometry are true. Rather, when the differences don't matter, we can calculate and predict using Euclidean geometry. When the differences do matter, as in calculating paths of airplanes circling the globe, Euclidean plane geometry does not apply, and another model, geometry for spherical surfaces, is invoked.

Euclidean geometry is a deductive theory: A conclusion drawn from the axioms is accepted only if the inference is valid. It is a purely mathematical theory, which taken as mathematics would appear to have no application since the objects of which it speaks do not exist. But taken as a model it has applications in the usual way, arguing by analogy where the differences don't matter.

Models, analogies, and abstraction

We've seen models of static situations (the map) and of processes (acceleration of falling objects). We've seen examples of models that are entirely visual and of models formulated entirely in terms of mathematical equations. We've seen models in which the assumptions of the model are entirely implicit (the map), and we have seen models in which the assumptions are quite explicit (Newton's laws of motion).

In all the examples either the reasoning is clearly reasoning by analogy or can be seen to proceed by abstraction much as in reasoning by analogy. We do not ask whether the assumptions of a theory or model are true, even if that was the intention of the person who created the theory. Rather, we ask whether we can use it in the given situation: Do the similarities that are being invoked hold and do the differences not matter? Even in the case of Newton's laws of motion, where it would seem that what is at stake is whether the assumptions are true, we continue to use the model when we know that the assumptions are false in those cases where, as in any analogy, the differences don't matter. In only one example (the ether) did it seem that what was at issue was whether a particular assumption of the theory was actually true of the world.

The assumptions of theories in science are false when we consider them as representing all aspects of some particular part of our experience. The key claim in every analogy is false in the same way. When we say that one side of an analogy is "just like" the other, that's false. What is true is that they are "like" one another in some key respects which allow us to deduce claims for the one from deducing claims for the other.

The term ***model*** is typically applied to what can be visualized or made concrete, while ***theory*** seems to be used for examples that are fairly formal with explicitly stated assumptions. But in many cases it is as appropriate to call an example a theory as to call it a model, and there seems to be no definite distinction between those terms.

Confirming a theory

From theories we can make predictions. When a prediction turns out to be true, we say it ***confirms*** (to some degree) the theory. This is not the same as confirming an explanation, for it rarely makes sense to say the claims that make up a theory—the assumptions of the theory—are true or false.

We cannot say that verifications of the relation of pressure, temperature, and volume in a gas confirm that molecules are hard little balls and that all collisions are completely elastic. Nor does fitting a carpenter's square exactly into a wooden triangle that is 50cm x 40cm x 30cm confirm the theorem of Pythagoras. Nor can we say that finding a tree at the corner of 100 W and 100 S in Minersville disconfirms the model given by the map. The map wasn't meant to give any information about trees, so it doesn't matter that it shows no tree there.

Except in rare instances where we think (usually temporarily) that we have hit upon a truth of the universe to use as an assumption in a theory, we do not think that the assumptions of a theory are true or false. We can only say of a theory such as Euclidean plane geometry or the kinetic theory of gases whether it is ***applicable*** in a particular situation we are investigating.

To say that a theory is applicable is to say that, though there are differences between the world and what the assumptions of the theory state, those differences don't matter for the conclusions we wish to draw. Often we can decide if a theory is applicable only by attempting to apply it. We use the theory to draw conclusions in particular cases, claiming that the differences don't matter. If the conclusions—the predictions—turn out to be true (enough), then we have some confidence that we are right. If a prediction turns out false, then the model is not applicable there. We do not say that Euclidean plane geometry is false because it cannot be used to calculate the path of an airplane on the globe; we say that Euclidean plane geometry is inapplicable for calculating on globes.

When we make predictions and they are true, we confirm a range of application of a model. When we make predictions and they are false, we disconfirm a range of application, that is, we find limits for the range of application of a model. More information about where the model can be applied and where it cannot be applied may lead, often with great effort, to our describing more precisely the range of application of a model. In that case, the claims describing the range of application can be added to the theory. We often use mathematics as a language to make this art of analogy precise. But in many cases it is difficult to state precisely the range of application. Reasoning using models is reasoning by analogy, which is likely to require judgment.

Sometimes it's said that a theory is valid, or is true, or that a theory holds, or a theory works for. These are just different ways to assert that a particular situation or class of situations to which we wish to apply a theory is within the range of that theory.

Good theories, and modifying theories in the light of new evidence

We've seen that the criteria for whether one theory is good or better than another cannot in general include whether the assumptions of the theory are true or "realistic." So besides getting true predictions, what criteria can we use to evaluate theories? Consider what we do when we discover that a prediction made from a theory is false.

When Newton's laws of motion result in inaccurate predictions for very small objects, we note that the theory had been assumed true for all sizes and speeds of objects and then restrict the range of application. But when the theory of the ether resulted in false predictions, no modification was made to the theory, for none could be made. That theory did not abstract from experience, ignoring some aspects of situations under consideration, but postulated something in addition to our experience, which we were able to show did not exist. The theory was completely abandoned.

If a theory has been made by abstraction, that is, many aspects of our experience are ignored and only a few are considered significant, then tracing back along that ***path of abstraction*** we can try to distinguish what difference there is between our model and our experience that matters. What have we ignored that cannot in this situation be ignored? If we cannot state generally what the difference is that matters, then at best the false prediction sets some limit on the range of applicability of the model or theory. We cannot use the theory here —where "here" means this situation or ones that we can see are very similar.

But our goal will be to state precisely the difference that matters and try to factor it into our theory. We try to devise a complication of our theory in which that aspect of our experience is taken into account. As with Einstein's improvement of Newton's laws, we get a better theory that is more widely applicable and which explains why the old theory worked as well as it did and why it failed in the ways it failed. We improve the map: by adding more assumptions, we can pay attention to more in our experience, and that accounts for the differences between the theories.

True predictions are never enough to justify a theory. Indeed, we do not "justify" a theory, nor show that it is "valid." What we do in the process of testing predictions is show how and where the theory can be applied. And for us to have confidence in that, either we must show that the claims in the theory are true, or show in what situations the differences between what is being modeled and the abstraction of it in the theory do not matter. True (enough) predictions help in that. But equally crucial is our ability to trace the path of abstraction so that we can see what has been ignored in our reasoning and why true predictions serve to justify our ignoring those aspects of experience. Without that clear path of abstraction, all we can do is try to prove that the claims in the theory are actually true.

Sometimes we are confronted with two theories that both yield good predictions for a class of situations and both of which have a clear path of abstraction. In that case, we say that one theory is ***better*** than another if (1) its assumptions are simpler; (2) it yields clearer derivations of the claims it is meant to explain; (3) it has a wider the range of application; and (4) it yields better the explanations of the archetypal claims it is meant to explain.

Example 8 "Consider the density of leaves around a tree. I suggest the hypothesis that the leaves are positioned as if each leaf deliberately sought to maximize the amount of sunlight it receives, given the position of its neighbors, as if it knew the physical laws determining the amount of sunlight that would be received in various positions and could move rapidly or instantaneously from any one position to any other desired and unoccupied position. Now some of the more obvious implications of this hypothesis are clearly consistent with experience: for example, leaves are in general denser on the south than on the north side of trees but, as the hypothesis implies, less so or not at all on the northern slope of a hill or when the south side of the trees is shaded in some other way. Is the hypothesis rendered unacceptable or invalid because, so far as we know, leaves do not 'deliberate' or consciously 'seek,' have not been to school and learned the relevant laws of science or the mathematics to calculate the 'optimum' position, and cannot move from position to position? Clearly, none of these contradictions of the hypothesis is vitally relevant; the phenomena involved are not within the 'class of phenomena the hypothesis is designed to explain'; the hypothesis does not assert that leaves do these things but only that their density is the same *as if* they did. Despite the apparent falsity of the 'assumptions' of the hypothesis, it has great plausibility because of the conformity of its implications with observation. We are inclined to 'explain' its validity on the ground that sunlight contributes to the growth of leaves and that hence leaves will grow denser or more putative leaves survive where there is more sun, so the result achieved by purely passive adaptation to external circumstances is the same as the result that would be achieved by deliberate accommodation to them. This alternative hypothesis is more attractive than the constructed hypothesis not because its 'assumptions' are more 'realistic' but rather because it is part of a more general theory that applies to a wider variety of phenomena, of which the position of leaves around a tree is a special case, has more implications capable of being contradicted, and has failed to be contradicted under a wider variety of circumstances."

Milton Friedman, "The methodology of positive economics"

Analysis Friedman's hypothesis about leaves seeking to maximize the amount of sunlight they receive cannot be used for reasoning by analogy or abstraction. It does not begin by either (a) looking at a real situation and comparing it to the growth of leaves, allowing us to distinguish what are the similarities and what are the differences, or (b) abstracting from experience to state what are the points of similarity that are supposed to hold, ignoring all else.

Rather, what he has posited is not an abstraction, but the addition of properties to a given situation. We are asked to suppose that leaves behave anthropomorphically with the skills of a terrific calculator. And then we are asked to ignore that as well. This doesn't make sense as a method of reasoning: why should we have confidence that predictions made from such a hypothesis will be accurate? That some of the predictions turn out to be accurate cannot be enough, any more than they are in astrology. We need to know why they turn out accurate in order to have confidence in the theory or model.

The alternative hypothesis of passive adaptation that Friedman presents is better, but not for the reasons he gives; rather, it is better for the reason he says is not meaningful. Namely, we have better reason to accept the alternative hypothesis precisely because we can see that in this case it is reasonable to believe it is true. No clearly false assumption incapable of fitting into reasoning by analogy or abstraction has been made.

Examples of Science Arguments

Example 1 "UFO enthusiasts often claim that the flying saucers they 'observe' are held suspended in the air and obtain their propulsion from a self-generated magnetic field. However, it is not possible for a vehicle to hover, speed up, or change direction solely by means of its own magnetic field. The proof of this lies in the fundamental principle of physics that nothing happens except through interactions between pairs of objects. A space vehicle may generate a powerful magnetic field, but in the absence of another magnetic field to push against it, it can neither move nor support itself in midair. The earth possesses a magnetic field, but it is weak—about 1% of that generated by a compass needle. For a UFO to be levitated by reacting against the earth's magnetic field, its own field would have to be so enormously strong that it could be detected by any magnetometer in the world. . . . And, finally, as the magnetic UFO traveled about the earth, it would induce electrical currents in every power line within sight, blowing out circuit breakers and in general wreaking havoc. It would not go unnoticed." Milton A. Rothman, *A Physicist's Guide to Skepticism*

Analysis This is first-rate refutation of a commonly held idea, clear with no fluff. We can easily follow each step of the argument.

Example 2 "Thus it is observed by the easy experiment of opening an artery at any time in living animals that blood is contained in the arteries naturally.

In order that on the other hand we may be more certain that the force of pulsation does not belong to the artery or that the material contained in the arteries is not the producer of the pulsation, for in truth this force depends for its strength upon the heart. Besides, we see that an artery bound by a cord no longer beats under the cord, it will be permitted to undertake an extensive dissection of the artery of the groin or of the thigh, and to take a small tube made of reed of such thickness as is the capacity of the artery and to insert it by cutting in such a way that the upper part of the tube reaches higher into the cavity of the artery than the upper part of the dissection, and in the same manner also that the lower portion of the tube is introduced downward farther than the lower part of the dissection, and thus the ligature of the artery which constricts its calibre above the cannula is passed by a circuit.

To be sure when this is done the blood and likewise the vital spirit run through the artery even as far as the foot; in fact the whole portion of the artery replaced by the canula beats no longer. Moreover, when the ligature has been cut, that part of the artery which is beyond the cannula shows no less pulsation than the portion above."

Andreas Vesalius, *Fabrica*, VII.19, written in 1543,
trans. S. Lambert, in *The Origins and Growth of Biology*

Analysis This is a clear exposition of an important experiment Vesalius made. He shows that the force of the blood going through the arteries is due not to the arteries themselves, though this passage leaves open why he believes the force is due to the heart.

Example 3 "But someone is bound to ask, can you prove that the computer is not conscious? The answer to this question is: Of course not. I cannot prove that computer is not conscious, any more than I can prove that the chair I am sitting on is not conscious. But that is not the point. It is out of the question, for purely neurobiological reasons, to suppose that the chair or computer is conscious. The point for the present discussion is that the computer is not designed to be conscious. It is designed to manipulate symbols in a way that carries out the steps in an algorithm. It is not designed to duplicate the causal powers of the brain to cause consciousness. It is designed to enable us to simulate any process that we can describe precisely."

John Searle, *New York Review of Books*, vol. 46, no. 6, 1999, p. 37

Analysis Searle writes with great style. But all his style can't cover up that his argument rests on an unstated premise that is false: "If something is not designed to do a task, it cannot do that task."

Example 4 "It is claimed that the earth is at rest in the center of the universe . . . Ptolemy feared that the earth and all earthly things if set in rotation would be dissolved by the action of nature, for the functioning of nature is something entirely different from artifice, or from that which could be contrived by the human mind. But why did he not fear the same and indeed in much higher degree, for the universe, whose motion would have to be as much more rapid as the heavens are larger than the earth? Or have the heavens become infinite just because they have been removed from the center by the inexpressible force of the motion; while otherwise, if they were at rest, they would collapse? Certainly if this argument were true the extent of the heavens would become infinite. For the more they were driven aloft by the outward impulse of the motion, the more rapid would the motion become because of the ever increasing circle which it would have to describe in the space of twenty-four hours; and conversely, if the motion increased, the immensity of the heavens would also increase. Thus velocity would augment size into infinity, and size, velocity. But according to the physical law that the infinite can neither be traversed, nor can it for any reason have motion, the heavens would, however, of necessity be at rest."

Copernicus, *Dialogue on the Two Chief World Systems*, trans. Stillman Drake in *The Origins and Growth of Physical Science*

Analysis Copernicus ably refutes Ptolemy's argument that the earth is at rest in the center of the universe first by analogy and then reducing to the absurd. The elegance of his style comes from the clarity and completeness of the presentation.

Example 5 [In the Seventeenth Century it was believed that worms and flies were spontaneously generated from mud and rotting or putrefying material.]

"I began to believe that all worms found in meat were derived directly from the droppings of flies, and not from the putrefaction of meat, and I was still more confirmed in this belief by having observed that, before the meat grew wormy, flies had hovered over it, of the same kind as those that later bred in it. Belief would be in vain without the confirmation of experiment, hence in the middle of July I put a snake, some fish, some eels from the Arno and a slice of milk-fed veal in four large wide-mouthed flasks; having well closed and sealed them, I then filled the same number of flasks in the same way, only leaving these open. It was not long before the meat and fish, in these second vessels, became wormy and flies were seen entering and leaving at will; but in the closed flasks I did not see a worm though many days had passed since the dead flesh had been put in them. Outside on the paper cover there was now and then a deposit, or a maggot that eagerly sought some crevice by which to enter and obtain nourishment. Meanwhile the different things placed in the flasks had become putrid and stinking."

Francisco Redi, *Experiments in the Generation of Insects* (1688), trans. Mab Bigelow, 1909, in *The Origins and Growth of Biology*

Analysis Redi describes his experiment clearly so we can duplicate it, and indeed, replicate the results. He assumes that it is obvious how to fill out the argument that the experiment proves what he set out to prove.

Example 6 "Every species of plant or animal is determined by a pool of germ plasma that has been most carefully selected over a period of hundreds of millions of years.

We can understand now why it is that mutations in these carefully selected organisms almost invariably are detrimental. The situation can be suggested by a statement made by Dr. J.B.S. Haldane: my clock is not keeping perfect time. It is conceivable that it will run better if I shoot a bullet through it; but it is much more probable that it will stop altogether. Professor George Beadle, in this connection, has asked: "What is the chance that a typographical error would improve *Hamlet*?"

Linus Pauling, *No More War*

Analysis Pauling apparently thinks that by just stating an analogy and asking some questions we'll be led to conclude as he does. But the dissimilarities are too great for this to be a good analogy. Clocks and *Hamlet* are man-made and are designed in advance to perform some function or achieve some goal. To change them is to tinker with their design. But the whole point of evolutionary theory is to replace design and goals and purposes as the causes of why plants and animals are like they are now. No matter how clear his writing is, it just covers up confused thinking.

Example 7 "*Sleepwalking and spontaneous parapsychological experiences: a note*

Two studies were conducted in which a questionnaire in Spanish with a true and false response format was used. It included, among other items, five questions about parapsychological experiences (waking ESP, dream ESP, apparitions, out-of-body experiences, and auras) and one question about somnambulism as follows: Some people have told me that I have sometimes walked in my sleep. The studies were conducted at the Centro Caribeno de Estudios Postgraduados, a private institute of graduate psychology studies in San Juan, Puerto Rico. In the first study, 120 questionnaires were collected by masters and doctoral students taking a graduate psychology course offered by the author. The students collected questionnaires from family, friends, and acquaintances outside the institution. In the second study, 52 questionnaires were collected by a colleague in two of his graduate courses. To measure frequency of psi [parapsychological] experiences, an index was formed from the above-mentioned five questions, assigning a score of 1 for true and a score of 0 for false answers.

The composite parapsychological experiences measure produced scores with the following characteristics: Study 1 (N = 120, M = 2.03, Range: 0–5, SD = 1.59); and Study 2 (N = 52, M = 1.48, Range: 0–4, SD = 1.23). The frequency of positive replies to the sleepwalking question was 17% for Study 1 (N = 119) and 24% for Study 2 (N = 51).

In the first study, those participants who replied affirmatively to the sleepwalking question (N = 20) obtained a mean of parapsychological experiences of 2.60, as compared to a mean of 1.94 for those who replied negatively, (N = 99), t(117) = 1.70, p = .045 (one-tailed), r = .16. In the second study, those with sleepwalking experiences (N = 12) obtained a mean of parapsychological experiences of 2.00, as compared to a mean of 1.28 for those without, N = 39, t(49) = 1.80, p = .039 (one-tailed), r = .25. The combined assessment of the p values in both studies produced a Stouffer z of 2.45, p = .01 (one-tailed). The combined effect size, using a Fisher z transformation [reference given] was .21. The difference between the effect sizes of Study 1 (r = .16) and Study 2 (r = .25) was not significant, z = –.52, p = .603 (two-tailed).

The results support the idea that sleepwalking is related to the frequency of parapsychological experiences. This, in turn, provides further evidence of a low-magnitude association between parapsychological experiences and dissociation. Further work should be conducted using better measures of sleepwalking, probing for both the frequency of experiences and for the stage in the experiencer's life in which sleepwalking took place or was most frequent. Habitual sleepwalkers should also be compared to nonsleepwalkers in future studies. . . ."

Carlos S. Alvarado, *Journal of Parapsychology*,
vol. 62, i4, 1998, p. 349

Analysis Did you go glassy-eyed and get wowed by the statistics? They're meaningless. The sample was not chosen randomly and has lots of room for bias. And even if the statistics did show a correlation between the responses, all that would show is that there's a correlation between the responses people give on a questionnaire to whether they sleepwalk and whether they have parapsychological experiences. Or rather to whether someone has *told* them they sleepwalk. The author's conclusion assumes that there is a correlation between people saying they have parapsychological experiences and actually having those experiences. Even in this journal which takes for granted that there are parapsychological experiences, the correlation between reporting that you've had such an experience and actually having one has to be established. The whole study is nonsense.