How to Reason
A Practical Guide
+
Reasoning in the Sciences

Richard L. Epstein

Illustrations by Alex Raffi

Advanced Reasoning Forum

How to Reason
and Reasoning in the Sciences

Preface

Claims

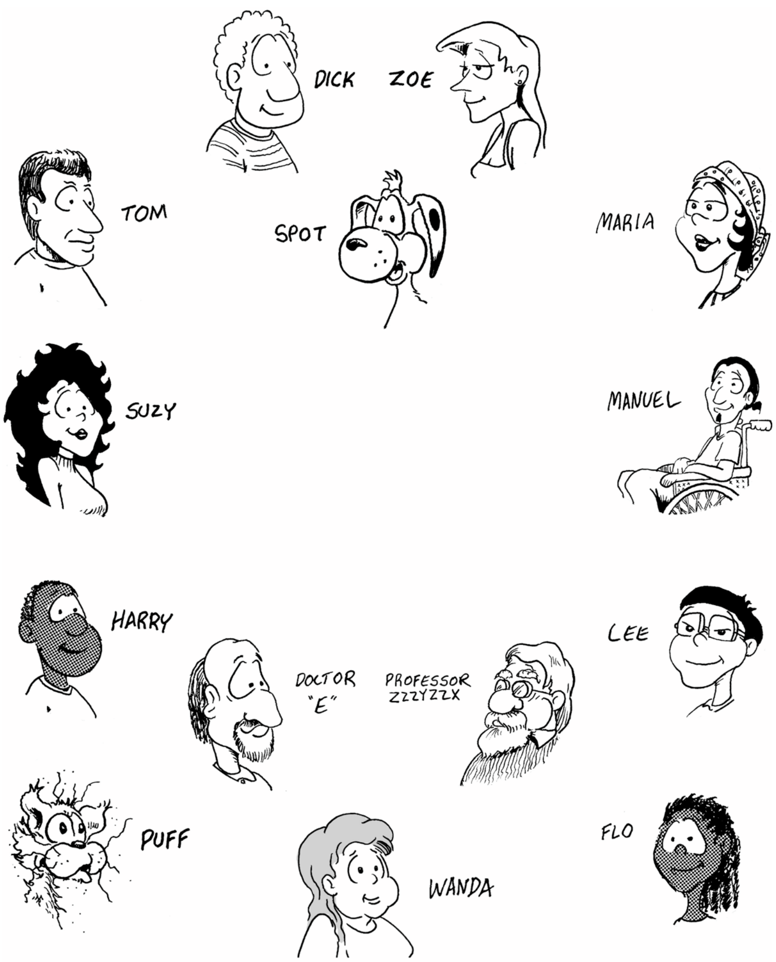
Arguments

The Form of an Argument

Numbers and Graphs

Reasoning from Experience

Cast of Characters



Claims

To reason well, to search for what is true, we need to know how to recognize what in our speech can be true or false—what we call "claims"—and what is so vague that it's just nonsense. Definitions can help us make clear what we're talking about.

Whether a sentence is too vague to be a claim depends in part on whether it's meant as a description of the world outside us or whether it's about thoughts, beliefs, or feelings. What counts as too vague depends also on whether a sentence is meant to say what is or what should be.

We'll see, too, how people can mislead us into believing a claim by a clever choice of words.

1 Claims

Claims A *claim* is a declarative sentence used in such a way that it is either true or false, but not both.

To understand this or any definition we need to see examples of what fits the definition, of what doesn't fit, and what's on the border line. Only then can we begin to use the idea.

EXAMPLES

• *Dogs are mammals.*
This is a claim.

• *2 + 2 = 5*
This is a claim, a false one.

• *Dick is a student.*
This is a claim, even if we don't know if it's true.

• *How can anyone be so dumb to think cats can reason?*
This is not a claim. Questions are not claims.

• *Never use gasoline to clean a hot stove.*
Instructions and commands are not claims.

• *I wish I could get a job.*
Whether this is a claim depends on how it's used. If Maria who's been trying to get a job for three weeks says this to herself, it's not a claim—we don't say that a wish is true or false. But if Dick's parents are berating him for not getting a job, he might say, "It's not that I'm not trying. I wish I could get a job." Since he could be lying, in that context it's a claim.

• *There are more bacteria alive now than there were 50 years ago.*
This is a claim, though there doesn't seem to be any way we could know whether it's true or whether it's false.

We don't have to make a judgment about whether a sentence is true or whether it's false in order to classify it as a claim. We need only judge that it is one or the other. A claim need not be an *assertion*: a sentence put forward as true by someone.

Vague sentences

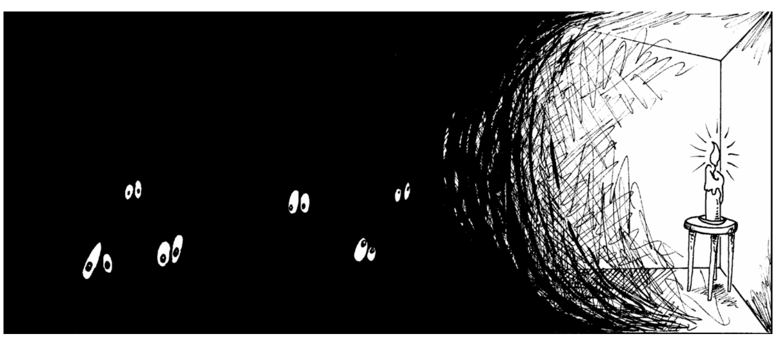
Often what people say is too vague to take as a claim. There's no single obvious way to understand the words.

EXAMPLES

• *People who are disabled are just as good as people who aren't.*
Lots of people take this to be true and important. But what does it mean? A deaf person is not as good as a hearing person at letting people know a smoke alarm is going off. This is too vague for us to agree that it's true or false.

• *Susan Shank, J.D., has joined Zia Trust Inc. as Senior Trust Officer. Shank has 20 years' experience in the financial services industry including 13 years' experience as a trust officer and seven years' experience as a wealth strategist.* —*Albuquerque Journal* April 29, 2010 and the Zia Trust website
"Wealth strategist" looks very impressive. But when I called and asked Ms. Shank what it meant, she said, "It can have many meanings, whatever the person wants it to mean." This is vagueness used to convince you she's doing something important.

Still, everything we say is somewhat vague. After all, no two people have identical perceptions, and since the way we understand words depends on our experience, we all understand words a little differently. So it isn't whether a sentence is vague but whether it's too vague, given the context, for us to take it as a claim. In a large auditorium lit by a single candle at one end, there's no place where we can say it stops being light and starts being dark. But that doesn't mean there's no difference between light and dark.



Drawing the line fallacy It's bad reasoning to argue that if you can't make the difference precise, then there's no difference.

Throughout this text we'll pick out common mistakes in reasoning and label them as a **fallacies**.

EXAMPLES

• *If a suspect who is totally uncooperative is hit once by a policeman, that's not unnecessary force. Nor twice, if he's resisting. Possibly three times. If he's still resisting, shouldn't the policeman have the right to hit him again? It would be dangerous not to allow that. So, you can't say exactly how many times a policeman has to hit a suspect before it's unnecessary force. So the policeman did not use unnecessary force.*

This argument convinced a jury to acquit the policemen who beat up Rodney King in Los Angeles in the 1990s. But it's just an example of the drawing the line fallacy.

• *Tom: My English composition professor showed up late for class today.*
Zoe: What do you mean by late? How do you determine when she showed up? When she walked through the door? When her nose crossed the threshold?

Zoe is asking for more precision than is needed. In ordinary talk, what Tom said is clear enough to be a claim.

• *Zoe: Those psychiatrists can't agree whether Wanda is crazy or not. One says she's clinically obsessive, and the other says she just likes to eat a lot. This psychiatry business is bunk.*

Just because there are borderline cases doesn't mean there isn't a clear difference between people who are really insane and those who aren't.

A sentence is **ambiguous** if it can be understood in two or a very few obviously different ways.

EXAMPLES

• *Zoe saw the waiter with the glasses.*

Did the waiter have drinking glasses or eyeglasses, or did Zoe use eyeglasses? If we don't know which is meant, it's not a claim.

• *There is a reason I haven't talked to Robert [my ex-lover] in seventeen years (beyond the fact that I've been married to a very sexy man whom I've loved for two-thirds of that time).*

—Laura Berman, *Ladies' Home Journal*, June 1996

The rest of the time she just put up with him?

• *Tom: Saying that having a gun in the home is an accident waiting to happen is like saying that people who buy life insurance are waiting to die. We should be allowed to protect ourselves.*

Tom, perhaps without even realizing it, is using two ways to understand "protect": physically protect vs. emotionally or financially protect.

• *Dr. E's dogs eat over 10 pounds of meat every week.*

Is this true or false? It depends on whether it means "Each of

> Dr. E's dogs eats over 10 pounds of meat every week" (big dogs!) or "Dr. E's dogs together eat over 10 pounds of meat every week." It's an **individual versus group ambiguity**.

We can tolerate some vagueness, but we should never tolerate ambiguity in our reasoning, because then we really don't know what we're talking about.

Now you should know what these mean:

- Claim.
- Too vague to be a claim.
- Drawing the line fallacy.
- Ambiguous sentence.
- Individual vs. group ambiguity.

You should be ready to use these, perhaps uncertainly, but as you see them put to use in more examples and with other ideas, you'll soon be able to make them your own.

Try your hand at these!

Which of the following is a claim?

1. College is really expensive now.
2. Pass the salt, please.
3. Bill Gates founded Apple.
4. Your best friend believes that Bill Gates founded Apple.
5. A friend in need is a friend indeed.
6. The sky is blue.
7. The sky is blue?
8. Whenever Spot barks, Zoe gets mad.
9. The Dodgers aren't going to win a World Series for at least another 10 years.
10. If you don't pay your taxes on time, you'll have to pay more to the government.
11. Suzy: I feel cold today.
12. Public education is not very good in this state.
13. Men are stronger than women.
14. Americans bicycle thousands of miles every year.
15. He gave her cat food.

Answers

1. Not a claim. Too vague.
2. Not a claim. A command.
3. A claim (false).
4. A claim, but not the same as the last one.
5. What the heck does this mean?

6. A claim.
7. Not a claim. A question.
8. A claim.
9. A claim. We just don't know whether it's true or false and won't know for another 10 years.
10. A claim.
11. A claim. Sure it's vague, but what do you expect when talking about feelings?
12. Not a claim. Too vague.
13. Not a claim,. Too vague. Strong in what way? Can lift more? Can lift more for their body weight? Can survive trauma better?
14. Not a claim. Individual vs. group ambiguity.
15. Not a claim, ambiguous.

30 What Can Go Wrong with an Experiment?

There are lots of ways experiments and interpretations of experiments can go wrong besides bad sampling or bad causal reasoning.

Experiments Start with a Question

• *Professor Shibokbok asks each of the 103 students in her psychology class to interview at least two people, asking what they've thought during the last two days. Then she and her students examine the answers, going over them again and again, classifying the responses. Eventually they find a correlation—between the weight of the person and his or her view of public transportation.*

There's always some correlation you can find if you look long enough at some mass of observations. But that's going backwards. Unless you start with a question, and have some criteria for what counts as an observation, and follow the usual rules for generalizing and causal reasoning, it's nonsense.

Observations are meant to answer a question, to test an hypothesis, to give some evidence to believe or disbelieve a claim. Otherwise, there would be no justification for the conclusions you'd draw.

> ***The fallacy of mining the data*** It is a mistake in reasoning to comb through a mass of observations looking for some correlation and then claim that the correlation is significant—when the data was not collected for the purpose of studying that correlation

Lying

• *In 2004 the highly respected journal* Science *published an article by Hwang Woo-suk detailing how he and his team had for the first time ever cloned human embryonic stem cells. That had been thought impossible. Hwang, a veterinarian and researcher, was already well-known for having cloned a cow and, for the first time, a dog, Snuppy. In 2006 Hwang admitted that the research papers had faked data. He was convicted by a court in Korea of embezzlement of millions of research dollars and of bioethical violations (taking human eggs from his research assistants).*

• *In 1988* The Lancet, *one of the most respected medical journals, published an article by Dr. Andrew Wakefield and twelve other authors. It said that they had found a link between children receiving the triple vaccine for measles, mumps, and rubella (MMR) and developing autism. The paper did not claim that there was cause and effect. But before it was published, Wakefield held a press conference in which he said that vaccinations using MMR should be suspended until further research was done. A lot of people stopped vaccinating their children. Outbreaks of measles became more common.*

But no one else was able to reproduce his results. More and more cause-in-population studies showed the safety and efficacy of the MMR vaccinations.

In 2011 the medical journal The BMJ *published an article showing that Wakefield's study misrepresented or altered the medical histories of the twelve children studied in it. As Brian Deer, a reporter for the* London Sunday Times *said, Wakefield was responsible for "falsifying medical histories of children and essentially concocting a picture, which was the picture he was contracted to find by lawyers hoping to sue vaccine manufacturers." According to* The BMJ, *Wakefield received $674,000 from the lawyers. Wakefield lost his license to practice medicine in the United Kingdom, and* The Lancet *retracted the article.*

Self-Deception

• *In the late 1800s a German mathematics teacher, Wilhelm von Osten, was convinced that animals have reasoning skills and intelligence. He tried to test his hypothesis with cats, but they were indifferent. His horse, Hans, however learned how to answer simple addition and multiplication problems by tapping his hoof the correct number of times. He could even do problems with fractions, which are beyond many high school students. After much skepticism was voiced, Germany's board of education organized a committee of a psychologist, a horse trainer, several school teachers, and a circus manager to investigate. They concluded there was no trickery. Then Oskar Pfungst, a psychologist, was called to study Hans. He found that it didn't matter whether von Osten or others questioned Hans: the answers were always correct. Unless, that is, the person didn't know the answer or was not visible to the horse. Eventually, Pfungst was able to show that each questioner, even when advised not to, showed a slight relaxation of tension when the horse tapped the correct number, which was what cued "Clever Hans" to stop. Clever Hans wasn't doing mathematics; he was clever only in reading people's expressions.*

Von Osten wasn't deceptive. He continued to believe that Clever Hans could do mathematics. But he was overlooking possibilities. That an animal's behavior can be influenced by subtle and unintentional cues from a questioner is called the **Clever Hans effect**. It also holds for questioning humans, who might not be aware that they're picking up on unintentional cues.

Sometimes, though, the problem is that there is less that is important.

• *Lee is one of 200 students in an experiment on psychic effects done by Professor Shibokbok. She asks each student to predict whether a coin flipped by a machine will land heads or tails. Each student does this 20 times. Most are right about 50% of the time, but 20 of them are right 15 or more times, including Lee. Those, she reckons, are the ones most likely to have psychic abilities. So she tests them again, and 4 of them, including Lee, are right more than 15 times out of 20. She tests those 4 again. Lee gets 16 right, and the others get less than 10. Now Professor Shibokbok wants to know whether Lee has precognition (can predict the future) or whether he's influencing the flip of the coin by telekinesis. The next day she tests Lee again, and he predicts only 6 flips out of 20. Then again, and he predicts only 4 out of 20. Somehow, Professor Shibokbok says, his psychic abilities have deteriorated overnight.*

Lee hasn't lost his psychic abilities because he never had any. With 200 people, it's likely that someone will predict correctly more often than 50% of the time, especially with so few flips. Professor Shibokbok didn't test those who weren't predicting correctly who might have suddenly "gained" psychic powers. In a long enough run of tests, Lee, as the other students, will have a run of correct predictions. But eventually the sum of all his predictions will tend toward 50%, the average (mean) of what is predicted as occurring by chance. Though it's often difficult to rule out chance as an explanation in an experiment, in this case it's not.

Professor Shibokbok couldn't get her work published in a reputable journal because it was obvious to the **referees**, other scientists who judge articles submitted to journals, that she didn't understand probability. She was denied tenure and has now moved to a position in administration.

Regression to the mean When experimental results are found that are far from the average predicted to occur by chance, and in subsequent trials the results tend to that average, it's a case of *regression to the mean*.

The power of suggestion

• *A young psychologist from Hamburg, Germany, named Krüger had been to a market and a gypsy woman gave him a description of his character. He was amazed how well it suited him. He showed it to his colleague, Zietz, without telling him where he'd gotten it and who gave it to him. Zietz thought it applied to himself. Both being of the conviction that they were very different people personality-wise, they decided to investigate the matter more closely. Krüger told the students in one of his psychology classes that he wanted to do a graphological experiment. All the students had to hand in a handwriting sample. To make the experiment look serious, Krüger waited 4 weeks before he gave the students their answers. They were called into a separate room where they were handed a typed version of the character description given to Krüger by the gypsy woman. Not one of the 39 students was dissatisfied with the evaluation of his or her character. Some thought that minor points were a little off, but most were astounded by the accuracy of the description. Afterwards, the students were told that they had all received the same description. The experiment has been repeated many times, with the same result.*

Taking fortune-tellers' predictions seriously or following the daily astrology predictions may seem like intellectual peccadilloes, not serious problems of the power of suggestion. But the power of suggestion can ruin scientific experiments involving humans, especially in testing medications. Special precautions have to be taken to avoid having the expectations of both the subjects and the experimenters skew the results.

• *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,* L. White, B. Trusky, and G. Schwartz, eds.

But the studies Evans cites don't show that the people had those symptoms. The authors show that the subjects reported that they had the symptoms. There's a big difference. And given the looseness of what we count as a headache, people may report what they wouldn't normally call a headache simply in order to have something to report. Wanting to please the researcher can distort responses, too. Further, even a sugar pill can have some effect, perhaps enough to make one a little happier and so have something to report.

It's extraordinarily hard to define "placebo effect" well, but roughly this is what we use.

Placebo A *placebo* is any substance or treatment given to some participants in a controlled cause-to-effect cause-in-population medical study that is assumed to have no effect.

Placebo effect A placebo effect is a reported or experienced change that follows from a person being given a placebo.

Double-blind trial A double-blind trial is a controlled cause-to-effect cause-in-population study in which neither the subjects nor those who administer the drug or placebo are told which is a placebo and which a drug.

With the slipperiness of what really is a placebo, all we can do is use double-blind studies with large enough populations. This allows us to compare the responses of those given the drug or treatment with those in the control group to see if there is a statistically significant difference. If there is, that's taken to be the effect of the drug or treatment.

With uncontrolled cause-to-effect experiments involving people, the expectations of the people involved can result in an unrepresentative sample. No matter how carefully studies are made on the effectiveness of different female contraceptives, they will be only marginally useful in helping women choose which method to use. That's because women who most want to avoid pregnancy choose the contraceptive they think will be most effective. So the women using the pill, which they are told is a highly effective way to avoid pregnancy, will be more motivated to follow the instructions for its use and always use it, while those who use contraceptive foam are likely to be more lax in following the method. According to the people who devise these studies, there doesn't seem to be any way to correct for this bias in the analysis of the data.

> ***Self-selection bias*** Self-selection bias occurs when those in the sample for a survey or experiment select themselves to participate.

Self-selection bias isn't only with people. When an ethologist studies only the chimpanzees that come to a feeding station, the sample of chimpanzees she's studying is self-selected.

Positive publication bias

• *Professor Fergamitz and Professor Lyle, at different universities and unknown to each other, have each done an experiment to the highest standards. They've examined their observations using the best statistical analyses.*

Prof. Fergamitz: I just found out that there's no correlation between being fat and being left-handed.

Prof. Lyle: I just found out that there's a 12% greater chance of being fat if you're left-handed.

There's no hope that Prof. Fergamitz can get his work published. Who would think that there would be a correlation? However, Prof. Lyle got his work accepted by a respected medical journal because it's such a surprising result.

Journals will publish new positive results but rarely new negative ones—unless, that is, to refute someone's experiment that is famous, like Andrew Wakefield's (p. 197). If Professor Lyle's paper starts lots of research on left-handed people, Prof. Fergamitz will likely be able to get his work published.

Journals—and researchers—tend to take positive results more seriously than negative ones. So when there are many people doing research, and only the positive results are published, we get a skewed picture of what's proved.

• [Dr. Joseph Banks Rhine was a famous ESP researcher from Duke University. He devised an experiment using cards with five easily distinguishable symbols: a square, circle, cross, star, and wavy lines. Used in decks of 25 cards, 5 for each symbol, the person being tested was asked to predict which symbol will be drawn.]

Let us imagine that one hundred professors of psychology throughout the country read of Rhine's work and decide to test a subject. The fifty who fail to find ESP in their first preliminary test are likely to be discouraged and quit, but the other fifty will be encouraged to continue. Of this fifty, more will stop work after the second test, while some will continue, more will stop work after the second test, while some will continue because they obtained good results. Eventually, one experimenter remains whose subject has made high scores for six or seven consecutive sessions. Neither experimenter nor subject is aware of the other ninety-nine projects, and so both have a strong delusion that ESP is operating. *The odds are, in fact, much against the run. But in the total (and unknown) context, the run is quite probable. (The odds against winning the Irish sweepstakes are even higher. But someone does win it.) So the experimenter writes an enthusiastic paper, sends it to Rhine who publishes it in his magazine, and the readers are greatly impressed.*

— Martin Gardner, *Fads and Fallacies in the Name of Science*

There's no deception here and no self-deception. The one experimenter did his work well. Still, a reputable scientific journal wouldn't publish his results unless they were reproduced by many experimenters, since they contradict so much else we know. Good referees are alert to the possibility of chance being at work.

Still, not every negative result is significant.

• *The famous Michelson-Morley experiment in the 19th century showed that the speed of light did not vary according to whether it was measured in the same path as the Earth's movement or across that path. This was the observation that led Albert Einstein to formulate his theory of relativity based on the assumption that the speed of light is the same for every observer. In the 1920s, 40 years after Michelson and Morley, a reputable physicist named Dayton C. Miller repeated the experiment and found slight variations in the speed of light, sufficient to question the theory of relativity. He repeated the experiment many times, always with the same results. He published articles about his work in scientific journals. But no one accepts that his experiment refutes the claim that the speed of light is constant.*

The Michelson-Morley experiment has been duplicated many times in virtually all technologically advanced countries, at differing altitudes,

> with different kinds of equipment. Except for Miller's work, the results always show that the speed of light, at least to the accuracy of the instruments involved, is the same. Even now no one knows why Miller's observations were different. It isn't bias toward positive results to conjecture that there was something odd in his equipment or that he, perhaps unconsciously, made errors recording the observations. Yes, it's possible that in that one place, at those particular times, there were differences in the speed of light. But no reputable scientist would accept that, not because he or she would be intimidated by the scientific community but because all the weight of evidence is against it.

Should we ever believe scientists?

One experiment is published; another contradicts its results. Should we just suspend judgment? Is there no standard we can use?

Some say that because scientists differ in evaluating some experiments and because there is no clear line between good work and bad work, between science and pseudoscience, we should suspend judgment on all scientific theories: the theory of evolution is no more to be trusted —or discounted—than astrology. That's a drawing-the-line fallacy. In the extremes, and those extremes are not far from the middle muddle, we can clearly distinguish between a theory that is very likely to be accurate and one that is very likely to be wrong.

Work in science is distinguished by constant testing. No observation, no result, no theory will long be accepted as both correct and important until it has been tested again and again. Yes, scientists make mistakes. Like all of us, they are sometimes careless, sometimes self-deceived, sometimes misled by mistaking chance for significance, and sometimes seduced by hope of money and power. But the community as a whole acts as a skeptical audience. Some call the testing, checking, and revising in science the "scientific method." But that's just a fancy name for all that scientists do, from their first training as students in how to reason to their laboratories and theoretical analyses. Testing, checking, and revising is what all of us would do for all the important decisions in our lives if we had the time and money—and energy.

In this chapter you've learned how to better evaluate experiments by looking for what can go wrong, using these ideas:

- The fallacy of mining the data.
- Clever Hans effect.

- Referee.
- Regression to the mean.
- Placebo.
- Double-blind trial.
- Self-selection bias.
- Positive publication bias.

Yet you also saw that we normally do have good reason to accept claims that scientists make if those claims have been tested by other scientists.

Try your hand at these!

1. List all the mistakes about data mining in the story "The Exaggerated Promise of So-Called Unbiased Data Mining" at <https://www.wired.com/story/the-exaggerated-promise-of-data-mining/>.
2. Explain why it is just as safe to jump from an airplane without a parachute as with one (see https://www.kanw.com/post/researchers-show-parachutes-dont-work-theres-catch).

Evaluate the following experiments and arguments about experiments.

3. 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.
4. 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