

---

# 4

---

## ACCEPTABLE REASONS

Critical thinking is reasonable and reflective thinking aimed at deciding what to believe and what to do. It is reasonable in part because it requires us to have reasons for our beliefs and decisions—reasons to think that our beliefs are true or that our decisions are the right ones. Critical thinking is reflective in part because it requires us to think about whether our reasons are good enough, and this means that to think critically we need to think about our reasons as reasons.

When evaluating an argument there are only two questions to ask: Are its premises true? Is it valid?

We saw in Chapter 3 that a person's reasons for believing or doing something can be put in to the form of an argument, with the reasons as premises and the belief or action as the conclusion. And we saw that there are really only two questions to ask when thinking critically about an argument. First, are the premises true? Second, is the argument valid? And we saw that these questions are wholly independent of each other. An argument might be valid even when its premises are false. And an argument's premises might be true even if they do not support its conclusion. So when we think critically about an argument, we always need to ask these two questions. In Chapter 3 we studied the question about validity. In this Chapter we look at the question of truth.

When considering whether an argument's premise is true, we need to consider whether its source is reliable. If it is not, then it would be unreasonable to accept the premise.

You might have noticed, though, that this Chapter is not entitled "Truth." It is entitled "Acceptability." There is a good reason for this. Sometimes, we might already know whether the premises in an argument are true or not. Maybe we are already an expert on the topic, or the argument will concern matters we are quite familiar with. But often we will not already know whether the premises are true. In that case, it is important to consider the source of the information in the premise. Some sources of evidence are very trustworthy, and others are not. If the source of the evidence is not trustworthy, then the evidence will not be acceptable. This Chapter is about how to think critically about when evidence is acceptable.

To start, imagine that you are on a jury and that Jones has been charged with murder. Suppose that Smith testifies that Jones was at the murder scene. We can think of this testimony as a premise in an argument for the conclusion that Jones is the murderer. The prosecutor is offering that testimony as a reason to believe that Jones is guilty. As a member of the jury, you need to assess that argument. As we have seen, there are really only two questions you need to answer. First, does the testimony prove that Jones is guilty? To answer this question you would use the strategies we discussed in Chapter 3. Second, is the testimony true? To answer this second question, you need to consider whether Smith is trustworthy. Should you take him at his word? Is he lying, or misinformed? Was he confused about who he saw that night? These are questions about whether the evidence Smith provided is acceptable.

Observation, measurement, and testimony are among our most reliable sources of evidence.

The aim of this Chapter is to study when evidence is acceptable. Smith is a source of evidence in that trial. In this case it is testimonial evidence. But there are many different sources of evidence including observation and various kinds of measurement. To decide what the weather is or will be, we collect evidence by looking out the window and by measuring the air's temperature and pressure. To decide whether our diet is working, we collect evidence by looking at our waist, and by measuring our weight on a bathroom scales. When trying to predict the future of the local economy, we collect evidence by reading the newspaper, by measuring unemployment rates, by studying business starts and stops. In all these cases, we rely on a source of evidence to supply us with reasons for our beliefs and actions. Observation, measurement, and testimony are among our most valuable sources of evidence. Without them, we probably would not have much if any knowledge at all.

In this chapter we will focus on these three sources of evidence. These sources differ from one another in important ways, and this means that there are special

questions we need to ask and terminology we need to use when thinking critically about each of them. We will study these questions and cover this terminology. But first, it will be helpful to begin at a somewhat more abstract level. For no matter what source of evidence we consider, the fundamental question to ask is the same: Is the source of evidence reliable in this particular case? Let us take a closer look at what this question means, and how we can answer it in a particular case.

#### 4.1 RELIABLE SOURCES

A source of evidence is **reliable** when it provides correct evidence more often than not. Three points are important to note. First, reliability is a matter of degree: some sources are more reliable than others. Second, reliability can depend on the surrounding conditions: a source of evidence may be more reliable in certain conditions than in others. Third, reliability depends on topic: a source may be reliable on evidence for one topic but not for another topic. Let us look at each of these in turn.

We can illustrate the idea of a reliable source of evidence with a somewhat silly example. Suppose that we want to know whether it is currently raining in Washington DC. Here are two sources of evidence. We might use the coin-flip method which tells us the following: flip a coin and if the coin lands heads, believe that it is raining; if the coin lands tails, believe that it is not raining. We might instead use the phone-call method: phone our friends who live in DC and ask them whether it is raining. If they say that it is, then believe that it is raining; if they say that it is not, then believe that it is not raining. It is pretty obvious that the phone-call method is better than the coin-flip method. But why is it better? Why is a friend who lives in DC a better source of evidence than the result of randomly flipping a coin? Part of the answer, surely, is that the phone-call method is far more likely than the coin-flip method to provide us with accurate evidence about the weather in Washington DC.

The reliability of a source of evidence is always a matter of degree, and some sources are more reliable than others.

The coin-flip method might give us the right answer on this particular case. But this would only be by sheer luck. And the chance that it would give us the right answer next time is not very high. In fact, the chance that it would give us the right answer is probably no better than 50%. By contrast, the chance that the phone-call method would give us the right answer is much, much higher, maybe even as high as 95%.

So part of what makes the phone-call method a better source of evidence than the coin-flip method is that the phone-call method is far more likely to give us the right answer. This is part of what makes the phone-call method a better source of evidence than the coin-flip method, at least concerning the weather in Washington DC. This silly example nicely illustrates how the acceptability of evidence depends on whether it comes from a reliable source. Because the phone call method is more likely to give

us the right answer means that it is more **reliable** than the coin-flip method. The more often a source of evidence gives the correct answer, the more reliable it is.

The coin-flipping story illustrates the important fact that the reliability of a source of evidence is a **matter of degree**. Some sources of evidence are more reliable than others. It is hard to imagine a source that could be 100% reliable, since every source we know about can malfunction. So we should not demand that level of reliability. But it is reasonable to think that there is a minimum level of reliability needed for the evidence provided by a source to be acceptable. And it seems right that the coin-flip method will never yield acceptable evidence simply because it is not nearly reliable enough.

The reliability of a source of evidence depends on whether it is operating under optimal conditions.

A second important point is that the reliability of a source of evidence depends on whether it is operating in **optimal conditions**. My friend in DC might be a better judge of the weather in the morning on his way to work than in the afternoon when he is sitting in his office cubicle. This means that as a source of evidence on the DC weather he is more reliable in the morning than in the afternoon. The bathroom scale is reliable only when it is on a flat and level surface. Our eyesight is reliable only when the lighting is right. Optimality conditions can vary from one source of evidence to another. In asking whether a source of evidence is reliable we need to keep in mind whether the conditions are optimal. This means that we cannot judge whether a source of evidence is reliable unless we know whether it is operating in optimal conditions, and this means knowing what those conditions are. The more we understand about how our sources of evidence work to provide us with evidence, the more reflective we can be in our thinking.

A source of evidence may be reliable on some topics but not on others.

The third point to keep in mind is that reliability is **topic relative**. My DC friend might be a reliable source of evidence on the weather in DC, but not on which wines to have with fresh fish. Maybe he can tell when it is raining, but not whether it is better to have a German or a French white wine with grilled ocean salmon. The bathroom scale is a good source of evidence on my weight, but not on my mood or cholesterol levels. This means that when we judge whether to accept some evidence we need to ask whether the source is an appropriate source on this topic. Just because the source is reliable for one topic does not mean that it is reliable for others.

### **SUMMARY: RELIABILITY**

A source of evidence is reliable when it provides accurate information most of the time. The reliability of a source of evidence is always a matter of degree, depends on optimal conditions, and is topic relative.

## 4.2 UNDERMINING AND OVERRIDING EVIDENCE

In Chapter 1 we saw that pieces of evidence can conflict in two ways. First, one piece of evidence **directly** conflicts with another when it supports an opposite conclusion. Second, one piece of evidence **indirectly** conflicts with another, when it indicates that the second piece of evidence is from an unreliable or untrustworthy source. This is a good time to look again at these different kinds of conflicts.

One piece of evidence directly conflicts with another when it supports an opposite conclusion. Sometimes when one piece of evidence directly conflicts with another, one of the pieces of evidence will be stronger than the other. In that case, the stronger evidence **overrides** the weaker evidence. Overridden evidence is not acceptable. This means that it would be unreasonable to rely on overridden evidence when deciding what to believe or do.

When two pieces of evidence directly conflict, if one is stronger than the other, then the stronger evidence overrides the weaker evidence. It is unreasonable to accept overridden evidence.

For example, Smith's testimony that Jones was at the hotel the night of the murder directly conflicts with Sam's testimony that Jones was at home the night of the murder. Those two bits of testimony directly conflict. And suppose that Smith's testimony is stronger than Sam's evidence and so overrides it. Sam's evidence is thus not acceptable. This means that it would be unreasonable for the members of the jury to rely on Sam's evidence.

It is not always easy to tell which piece of evidence is stronger than another. Where we have directly conflicting evidence we have to be extremely cautious. It will not always be obvious which source of evidence is at fault. We can formulate some general rules of thumb. A piece of evidence is overridden if:

- it conflicts with evidence from a known reliable source; or
- it conflicts with expert opinion; or
- it conflicts with what we already have good reason to believe.

In cases of conflict, we need to make a judgment. Do we reject the new evidence; do we reject the evidence it conflicts with; or do we reject both? It will not always be obvious which piece of evidence is most accurate. Maybe the source we thought was reliable has made a mistake. Maybe the expert opinion is wrong in this case. Maybe the proper response to the new evidence is to make a relatively large revision to our standing beliefs. If we want to avoid making a mistake, we should withhold judgment altogether, and collect more evidence before we make a judgment. We need to decide which source of evidence is most reliable. Until we make that decision, the best or at least the most prudent course is to withhold judgment as long as possible.

One piece of evidence indirectly conflicts with another when it indicates that it is from an untrustworthy or unreliable source.

Evidence can conflict in a second way. One piece of evidence **indirectly** conflicts with another when it indicates that it is from a source that is not reliable or trustworthy. If the first piece of evidence is stronger than the second, then the first **undermines** the second. Undermined evidence is not acceptable. This means that it would be unreasonable to rely on undermined evidence when deciding what to believe or do.

When one piece of evidence indirectly conflicts with another, if the first is stronger than the second, then the first piece of evidence **undermines** the second piece of evidence. It is unreasonable to accept undermined evidence.

Suppose, for example, that Ayesha testifies that Smith hates Jones and would lie on the stand to harm him. Ayesha's testimony indirectly conflicts with Smith's, because it suggests that Smith is not trustworthy. Let us suppose that Ayesha's testimony is stronger than Smith's and so undermines it. Smith's testimony, then, is not acceptable. This means that it would be unreasonable for the members of the jury to rely on Smith's testimony when deciding whether Jones is guilty.

Undermined evidence can still be true.

There is an extremely important point to keep in mind. Undermined evidence may still be true. Consider again Ayesha's testimony. Her testimony was stronger than Smith's and so undermined it. This means that Smith's testimony is not acceptable, and that the member of the jury should not rely on it when deciding whether Jones is guilty. But none of this means that Smith's testimony is false. He might have been telling the truth, even if he hated Jones and was willing to lie on the stand to harm him. Even liars sometimes tell the truth. So when the members of the jury decide that Smith's testimony is not acceptable, this does not mean that they can then also decide that Smith's testimony is false. All they can reasonably decide is that Smith is not trustworthy—his testimony is not acceptable. They should remain agnostic, and withhold judgment about where Jones was the night of the murder until more of the evidence is in.

Consider another case. Suppose that my kids have been playing with the bathroom scale again and that I know that the last time they did this, the scale broke and gave crazy readings. I now have reason to think that they have once again broken the scale's internal mechanisms. In this case, I should not trust what it tells me when I step on it. Of course, the reading I get from it might be accurate, but still I should withhold judgment until I can make sure that it is not broken.

### CRITICAL THINKING MISTAKES: APPEAL TO IGNORANCE

It is a mistake to believe something just because you do not have evidence that it is false. This is a mistake because a bit of investigation might show that it is false, and thinking critically requires looking for evidence when one can. One form of this mistake is to accept a piece of evidence just because you do not know of any overriding or undermining evidence. Critical thinkers should look for overriding and undermining evidence, before accepting a piece of evidence.

We have seen that evidence is acceptable if it comes from a reliable source and it is neither (ii) undermined nor (iii) overridden by other evidence we have. We usually do not need to know that a source is reliable in order to be justified in relying on it. But if we have reason to think the source is not reliable or to think that the evidence is inaccurate, then we are not justified in relying on it. These general points about when we can trust evidence from a source apply to every source of evidence. They are general enough that we can keep them in mind whether we are relying on observation, testimony, or measurement. This will help us as we move forward. Still, there are important differences between these three sources of evidence, and seeing them will help us identify some additional questions to ask.

### SUMMARY: UNDERMINING AND OVERRIDING EVIDENCE

When two pieces of evidence directly conflict, if one is stronger than the other, then the stronger evidence overrides the weaker evidence. Overridden evidence is not acceptable.

When one piece of evidence indirectly conflicts with another, if the first is stronger than the second, then the first piece of evidence **undermines** the second piece of evidence. Undermined evidence is not acceptable.

## EXERCISE 1

### A. Comprehension Questions

- a. What does it mean for a source of evidence to be reliable?
- b. Why is reliability a matter of degree?
- c. Explain why reliability is topic relative? Could there be a source of evidence whose reliability is not topic relative? Explain.
- d. What is the difference between overriding and undermining evidence?
- e. Construct an example of a case in which some evidence is overridden but not undermined.

- f. Could a source that is highly reliable nonetheless provide false evidence? Describe an example other than the ones discussed in the text.
- g. Suppose that you had evidence that undermined the evidence provided by some source of evidence, S. Could it still be that S is highly reliable? Explain, and use examples to illustrate.
- h. Suppose that the evidence provided from one source always conflicted with the evidence provided by another source. Should we continue to trust those sources? Which one should we doubt?
- i. Some people think that fortune telling is a good source of evidence. What do you think? Why?
- j. Suppose that we wanted to determine whether perceptual observation is a reliable source of evidence about the colors of medium-sized objects. How could we do this?

### 4.3 OBSERVATION

The English philosopher John Locke (1632–1704) claimed that without perceptual observation we would have no ideas or thoughts and so no knowledge at all. It is very difficult to disagree with this claim. From the moment we wake up in the morning, we rely on our observations of our surroundings to get around—to find out where we are and what we have to do to get our breakfast. We know that—blind people still have beliefs and knowledge, as do deaf people and people like Helen Keller who lack several senses. But it is hard to imagine how a person who had no sense organs at all could possibly have any knowledge of anything at all. Perceptual observation certainly seems essential to knowledge, or at least to human knowledge. At the same time, we know that observation is not infallible—it can and sometimes does make mistakes. Sometimes, things are not quite as they seem to be. So while we have little choice but to rely on our observations, we need to do so reflectively. In this section, we will study the conditions under which we are justified in relying on perceptual observation.

By perceptual observation we can include the ordinary five senses—taste, touch, smell, hearing, and sight. But we can also include our capacity to tell such things as when we are hungry or thirsty and to tell the relative position of our body parts, such as where our arms are in relation to each other—a capacity called “proprioception” by philosophers and psychologists. Each of these sources of evidence about the world and ourselves is reliable, but only under certain conditions. Sight, for example, only works properly when the external conditions are right. The light has to be bright enough, but not too bright. Changing the color of the light can affect the visual appearance of things. There are also internal conditions that must be right. Vision is not reliable after the optometrist has put dilating drops in—everything looks fuzzy and shadowy. The internal and external optimality conditions are already pretty familiar to us, and we do not need to go into a lot of detail about them here.

Perceptual observation is reliable on some topics but not others. We can often tell by looking what colors things are and where a thing is in relation to other things.



We can tell whether the toaster is on the counter or in the cupboard. We can tell by hearing whether the radio is on. We can tell by smell whether the stew is burning. We also know that different senses are reliable on different topics. We can tell by looking, but not by smell, whether the TV is still on. We can tell by smell, but not by hearing, whether the milk has gone sour. We can tell by touch, but not by sight, whether the water is cool enough for swimming. It is obviously important to make sure that we are using the proper sense for a given topic. It is also familiar that there are lots of things we cannot tell with any of our sense organs. We cannot tell by observation whether a person has AIDS—we need to run complex tests for this. Nor can we tell by observation alone whether the economy is improving—we need to make some complex measurements for this. Running the tests and making these observations would be impossible without observation—but the evidence they yield is not observational evidence.

Observation improves as we mature and with training. Anyone who has tried to teach little kids to swing a baseball bat knows that it can be frustrating. It seems to take little kids a long time to learn how to time the bat's swing—something that seems so easy for adults. It is as if the kids cannot even see the ball. Recent studies on the development of the visual system seem to suggest that this is exactly what is going on! The capacity to tell how quickly things are moving requires a relatively advanced level of brain development. On reflection, this is not that surprising. Our perceptual systems are after all just part of our body and we know that our bodies mature and change. In fact, it would be surprising if our perceptual systems did not become more reliable as we grew up. We also know that we can increase the reliability of our perceptual systems with training. Trained musicians can hear rhythms and melodic progressions and patterns in musical performances that others cannot hear. Skilled gardeners can tell by looking whether plants need watering or fertilizer. Doctors learn to identify various skin conditions just by sight. Experts on wine can taste things in wine that most of the rest of us cannot. These improvements are not just the result of maturation—they result from training and practice.

Perceptual capacities are also subject to illusions. This is especially familiar in the case of vision. Some visual illusions are **optical**—that means that their explanation has to do with the way light works. For instance, a straight stick in a glass of water looks bent because the light reflected off the part of the stick in the water is slowed down as it travels through the water causing it to change directions slightly, producing the illusion. But other visual illusions are **cognitive**—they have to do with the way our visual system is structured or the way it works. The illusion that parallel railroad tracks meet has to do with the distance between our eyes. The same is true of the Müller-Lyer illusion, we saw in Chapter 1. Others are harder to classify. For instance, it is a familiar experience that a full moon seen close to the horizon looks a lot bigger than a full moon seen high in the sky. At first, people thought it was an optical illusion, caused by the fact that light reflected off the moon has to travel through much more atmosphere when it is on the horizon than when it is in the high sky. But if this was so, then one would expect the image on the eyeballs to be of different sizes. But the images are the same size whether the moon is on the horizon or in the

high sky. It is now thought that the illusion is produced as the brain “interprets” or processes the information. One possibility is that it has to do with the fact that the moon on the horizon is seen as close to other objects. But this illusion is not yet fully understood.

#### 4.4 MEMORY

We often rely on our memories to ground or sustain our beliefs. But memory’s role in the justification of beliefs is a special one. For memory is not a source of evidence; rather, it is a repository of evidence. Whereas observations are bits of evidence, memories are not. Memory stores evidence. This means that when we rely on our memory, our evidence is no more acceptable than the acceptability of the evidence we remember.

But we also know that memory can be unreliable: it is as if in the storage process the evidence gets modified or changed. The US National Transportation Safety Board (NTSB) provides an especially striking example of this. After a plane crash, the agents of the NTSB collect as much evidence as they can in the hopes of trying to recreate the sequence of events that led to the crash. Among the evidence they collect are reports from eyewitnesses on the ground. But over time, they have found that eyewitness reports are highly variable. In the case of the crash of American Flight 587 in 2001, the NTSB interviewed 394 eyewitnesses. They found that:

... 52 percent said they saw a fire while the plane was in the air. The largest number (22 percent) said the fire was in the fuselage, but a majority cited other locations, including the left engine, the right engine, the left wing, the right wing, or an unspecified engine or wing. Nearly one of five witnesses said they saw the plane make a right turn; an equal number said it was a left turn. Nearly 60 percent said they saw something fall off the plane; of these, 13 percent said it was a wing. (In fact, it was the vertical portion of the tail.)<sup>1</sup>

Why are eyewitnesses so unreliable? It might be that their visual observations of the event are unreliable. But it is hard to see how so many people could have had such different visual experiences, especially when they were all looking at the very same event, and maybe even standing right next to one another. More likely, their visual experiences were somehow distorted as they got put into memory, while they stayed in storage, or while they were being retrieved from storage. Whether the distortion happened before the storage, during the storage, or during the retrieval process, their memories are distorted. The NTSB still collects eyewitness reports, but they do this as for largely public relations reasons. They no longer rely on these reports when trying to figure out what happened. Admittedly, memories of horrific visual scenes such as the crash of an airplane are the exception, and the fact that such experiences

<sup>1</sup>Wald, M. For air crash detectives, seeing isn’t believing. *The New York Times* (June 23, 2002), Section 4, p. 3.

are misremembered does not by itself show that memory is not in general reliable. But it does illustrate once again the reason for the following maxim: trust, but verify.

### **CRITICAL THINKING IN PSYCHOLOGY**

Researchers have found that a person's memory can be manipulated in different ways.<sup>1</sup> In one study, subjects were shown a fake advertisement of Disneyland featuring Bugs Bunny standing next to the Magic Castle. The ad looked just like a real ad for Disneyland. After studying the ad, subjects were asked to describe their own experience as children visiting Disneyland. Sixteen percent of them said that they remembered meeting Bugs Bunny at Disneyland. The greater the number of exposures to the fake ads, the higher the percent who claimed to remember personally meeting Bugs in Florida. Some even claimed to remember specific details, such as hugging him and touching his ears. But since Bugs Bunny is not a Disney character, these supposed memories are all false, somehow implanted or encouraged by the false advertisements. Researchers have even been able to instill false memories of quite unusual and memorable events. In one study, a subject's parents were enlisted to tell the subject that she had poured a slimy substance onto the head of her Grade 1 teacher. The story was the very same for each subject, aside from names and places, and was full of detail. They were even provided with fake photos of the event to show to the subject. Remarkably, 65% of the subjects later reported to researchers that they remembered the event in vivid detail, and expressed shock and surprise when told the entire event was fictional.

### **EXERCISE 2**

- A.** List five conditions under which visual perception is not reliable. Do the same for our sense of touch.
- B.** We learned that our senses are reliable for some topics but not others. List some topics for which vision but not touch is reliable. List some for which hearing and sight are both reliable. List some on which no sense is ever reliable.
- C.** Sometimes, our different senses provide us with conflicting evidence. Describe such a case. Which sense should we trust in a case like that? If you can, think of a general rule or principle that can be used to always decide which sense to trust when senses conflict.

### **4.5 TESTIMONY**

It is difficult to exaggerate the importance that testimony plays as a source of evidence for us. Without it we would have almost no knowledge. But it is also a very complex

source of evidence, one that critical thinkers need to be very cautious about. In this section we will study why.

Imagine that you only relied on your own, personal observations. How much knowledge would you be able to acquire? Not very much, probably. Just think of how little you can actually see and feel at any one time. Even if our perceptual faculties are as highly reliable as we hope, they are also extremely limited. But we all know (or at least think that we know) a lot about things we have never seen or touched. We know about the history of the US constitution, about the battles of the two world wars. We may have seen the 9/11 attacks with our own eyes, but we need more than our own eyes to figure out the causes of the attack or what structural causes made the towers fall down. We know about distant places and times, beyond our observation. Most of our knowledge, in fact, would be out of reach if we did not rely for information on other people, whether it be our parents, friends, teachers, authors we read in the news or see on TV shows, or even just people we overhear in the local coffee shop or bar. (Hopefully you have learned a thing or two from me in reading this book.)

Evidence that consists in what other people tell us is **testimonial evidence**. It may sound a bit fancy to call the information we get from newspapers, teachers, and parents, “testimony.” It sounds a bit odd to say that our best friend is testifying when he tells us that the fridge is full of beer. But the analogy between these ordinary cases of believing what people tell us and the role of witnesses in a trial is very strong. In all of these cases we are treating other people as sources of evidence. We are taking them at their word; trusting their say-so. This raises the question: when is testimonial evidence acceptable?

#### **SUMMARY: TESTIMONIAL EVIDENCE**

Testimonial evidence is acceptable only if

- (i) it is on an appropriate topic; and
- (ii) the witness is properly trained; and
- (iii) the witness is properly informed; and
- (iv) the witness is unbiased.

The first three conditions are already familiar, though we will see that there are some special factors to keep in mind. But the fourth condition is a new one. It is needed because whenever we are deciding whether to trust what someone is telling us, we need to think about whether that person is biased or prejudiced. Let us look at each of these conditions in more detail.

#### **4.5.1 Appropriate Testimony**

First, testimonial evidence is acceptable only if the topic is **appropriate**. As we saw above, a source of evidence might be reliable on some topics but not others. The same is true for people. When a person is a reliable source of evidence on some topic, we can think of her as an expert on it. Some people are experts on sports, while others are

experts on the chemistry of cells. A person can be an expert on several very different topics. But there are some topics where there simply are no experts.

A familiar example is any topic where there are no real objective facts, but just matters of taste. For example, I doubt whether there are facts about whether one popular musician is better than another. There are, of course, people who know a lot about pop music, about the different performers and their histories and musical capacities. There are experts about who can carry a tune and play the guitar. But is there really such a thing as being an expert on whether Madonna's music is better than Bruce Springsteen's? I have always liked Bob Dylan's singing, though many people find it (and my musical preferences) horrifying. But is there anything more to this disagreement than just a difference in taste? I doubt it. I am inclined to think that it is inappropriate to appeal to experts to settle disputes about which pop musician is better.

But it is not just in matters of taste that there are no experts. Sometimes, when a new field of study is very young and just getting established, there will not yet be experts either. This is the case at cutting edge fields in natural science. When the scientists working at that cutting edge disagree among themselves about the field, and especially when they disagree about which methods are best for measuring or describing the phenomena, then there are probably no real experts yet. In cases where the best-placed people in the field cannot agree, then we as nonexperts should probably withhold judgment too. This was the case at some time in almost every branch of science. It was the case for the science of global warming until about 20 years ago. But now there is no doubt that there are experts on global warming, and that it is perfectly appropriate to rely on what they say when we decide what to believe about global warming.

#### 4.5.2 Trained Testimony

Second, testimonial evidence is acceptable only if the witness is properly trained. This is just a way of asking whether the witness is reliable, whether there is a high likelihood that his testimony will be true.

What it takes to be properly trained depends of course on the topic. Trials provide lots of good examples. Only in special kinds of cases will an eye witness to a crime have to show that her eyes were working properly the afternoon of the crime, though if she ordinarily wears glasses that will be relevant to whether she really did see what happened. But it is much more common for a witness on a specialized scientific topic to have to demonstrate to the court that she has the proper training. Expert witnesses on DNA testing or finger printing need to show they have the training and certification needed for their testimony to be acceptable. Usually, the fact that a witness has been certified by the relevant organization is good reason to think she has the proper training to count as an expert witness on that topic.

These examples are from court trials. But the same issues arise in more mundane cases too. I would trust my highly trained electrician over my 9-year-old daughter to tell me whether the wiring in the panel is adequate. I would turn to the pharmacist and not the grocery clerk for advice on which antihistamine to buy, though I might trust the grocery clerk over the pharmacist on which streetcar will get me downtown fastest. In this last case, it is not the training but the experience that matters.

### 4.5.3 Informed Testimony

The difference between being properly trained and being properly informed is an important one. A person might be an expert at fingerprint identification, but if she has not actually studied the fingerprints given during the trial, then she is not properly informed. The following analogy might help: a thermometer is really good at telling the temperature of the water in a glass. But it will not get the reading right unless it is in the water. This is like the difference between being trained and being informed.

Sometimes, people who are considered experts find it difficult to admit that they do not know an answer to some question. This is understandable. But it is also an obstacle to critical thinking. We should prefer for them to keep quiet or admit to ignorance than say something that is ill informed. In general, we should not accept what a witness says if we have reason to think she is not sufficiently informed on the issue at hand, even if we think she is an expert on the general topic.

### 4.5.4 Unbiased Testimony

Finally, testimonial evidence is acceptable only if the witness is **unbiased**. The reason for this condition is pretty straightforward: sometimes witnesses are motivated in different ways to lie about the facts.

The example of the murder trial illustrates one possible source of bias. The defense attorneys claimed that the witness was biased because he was being paid by the prosecution to give his testimony. The implication was that the witness might not have given the same testimony had he been paid by the defense attorneys, or by no one. Of course, the fact that the witness was being paid for his testimony does not prove that he was lying or overstating or understating anything. But it might, and for some of the jury it did, raise the possibility that he was biased against the defendant and so was not to be trusted. Desire for financial gain is one source of bias, but it is not the only one. Just about anything can be a source of bias. People are moved to lie by jealousy and by love, by pride and by humility, out of loyalty and out of revenge, by a desire for fame and by a desire for anonymity. If we know that someone is biased, then obviously we should not accept what they say. Their testimony is acceptable only if there is no reason to suspect that they are biased.

#### **CRITICAL THINKING MISTAKES: UNACCEPTABLE TESTIMONY**

It is a mistake to accept testimony from a witness if the topic is inappropriate, the witness is not properly trained, or not properly informed, or if the witness is biased. It is a mistake because such evidence is not acceptable. Testimony is appropriate only on topics for which there are recognized experts. An expert must be properly trained and properly informed. And a witness must not be motivated to lie about or exaggerate the facts.

Judgments about witness bias can be tricky and can require balancing different facts about the witness. Suppose that the lead scientist for a well-known environmental group testifies before Congress that the water levels in the Great Lakes are dangerously low and that expensive conservation steps must be taken to reduce water usage in the cities and farms that depend on the water from the Lakes. What are her motivations? We know that she is paid by the environmental group to champion its environmental policies. If there were no environmental problems to report on, she would be out of a job. She probably also wants to keep her high-profile position, and might enjoy being in the spotlight before Congress. She might think that advocating an extreme position might, given the political realities involved in passing complex regulatory legislation, be the best strategic move. All of these considerations suggest that she might be motivated to exaggerate or even lie about the real findings.

### **CRITICAL THINKING MISTAKES: AD HOMINEM**

It is a mistake to believe that a piece of testimony is false just because the witness is unreliable or biased. It is a mistake because it confuses undermining and overriding evidence. Testimony can be true even if it is from an unreliable or biased source.

This mistake is traditionally called “ad hominem” because it involves criticizing testimony by criticizing the witness (the “hominem”). But we need to be a bit careful here in identifying this mistake. For it is not always a mistake to conclude that a witness is unreliable or biased. There can be very good reason to believe this. But it is always a mistake to conclude that *a witness’s testimony is false* just because they are unreliable or biased.

On the other hand, it is often hard to get away with a lie, and the reputation of her organization will suffer if it becomes public that its lead scientist has been lying. She also has a professional reputation as a scientist that she probably wants to maintain and even enhance. Lying or exaggerating will surely hurt her image among other scientists. And she probably has a personal sense of honor that forbids lying or exaggerating, except perhaps in extreme cases, which this surely is not. These considerations suggest that she is highly motivated to speak the truth, at least as she sees it. What should we conclude about her testimony? Since we are not experts, we are in a difficult spot.

### **CRITICAL THINKING AND LEGAL HISTORY**

In a famous American trial from the 1990s, the defendant was accused of having viciously murdered his wife and her friend. The prosecution’s case was built on forensic evidence that seemed to connect the defendant to blood samples found at the scene of the crime, in the defendant’s home and on a bloody glove of a kind once owned by the defendant. To many outside observers, the case seemed

quite strong. But the defense attorneys did a remarkable job of undermining the prosecution's star witness, a forensic scientist who testified that the blood samples matched the defendant's blood type. First, the defense lawyers argued that this testimony was inappropriate because the science of blood sampling was too young and not yet fully established. Second, they argued that the scientist was not properly trained to use the equipment involved in the sophisticated analysis of the blood. Third, they argued that the scientist was not properly informed about the facts at hand, because the police had mixed up the blood samples they had collected and there was no sure way to tell which samples were found in which place. Finally, they argued that the scientist was biased because he was being paid by the prosecution to give his testimony. Point by point, the defense attorneys had done a masterful job of undermining the witness. In the end, the jury decided that the scientist's evidence was not acceptable.

#### 4.6 ADVERTISING

Advertisements are a special case of testimonial evidence. Advertisements can serve many purposes, but generating sales of the advertised product or service is surely one of the most important. Advertisements can try to achieve this in different ways. Some appeal to emotions, as in the wonderful ads for Apple's iPod music player that involve nothing but a person dancing to the music they're listening to on their iPod. That ad works not by informing us of the product, but by trying to establish an emotional connection between the product and a desirable lifestyle. But many ads do aim to produce sales by informing potential customers about the product or service. Such ads can be thought of as involving testimonial evidence. (Indeed, some even involve "testimonials" by famous people describing their experience with the product or service.) We should evaluate claims made in such ads in the very same ways that we evaluate any other case of testimonial evidence.

We should first ask whether testimonial evidence is appropriate on that topic. Advertisements for food sometimes include claims about great taste. But are there really experts on taste? Recently, drug companies have been permitted to advertise their products. These ads are highly regulated and the drug companies are required to provide quite detailed information about potential side effects. But there is an underlying concern about the appropriateness of these claims, given that for many medicines so little is known about potential long term effects, both positive and negative ones. This is especially true for claims about the health benefits of diet supplements, since many of those products are neither tested nor regulated by the government, though advertisements of them are subject to the regulations that govern all advertisements against being misleading. In general, the acceptability of claims in advertisements about the health effects or benefits of products is questionable simply on the grounds that such claims are inappropriate.

When famous people promote things in advertisements we should ask ourselves whether they are properly trained and properly informed. Is a famous movie star



really an expert on which phone plan is best for me? (Set aside for now the fact that the actor is being paid to say that it is.) Is the CEO of an automotive company really an expert on the performance of his company's cars relative to the competition? Should we trust what students in college ads say about the benefits of their college compared to the competition? Unless the ad involves a recognized expert, there is good reason to not accept the claims made in the advertisement.

Finally, though, given that the advertisement is aimed at producing sales, the risk of bias is inevitable and serious. Advertisers know this and sometimes include favorable evaluations from independent organizations. An ad for a car might refer to the results of crash tests performed by an independent safety group, or might cite awards the car received in independent performance tests. (Of course, claims by outside organizations are themselves just more testimonial evidence and need to be evaluated on their own.)

There are also governmental rules regulating commercial speech, designed to prevent false or misleading ads. But these regulations are difficult to enforce. One recent study, reported in *The New Yorker*, suggests that more than 50% of advertisements for nutritional supplements involved false or misleading claims.<sup>2</sup> The history of advertising also offers little reason to trust claims made in advertisements. It is perhaps best to approach claims made in advertisements with an initial and healthy skepticism: given the high risk of bias, best not to accept the advertised claims.

We have been discussing the conditions under which testimonial evidence is acceptable. We have seen that it is acceptable if, but only if, (i) the testimony is on an appropriate topic; (ii) the witness is properly trained; (iii) the witness is properly informed; and (iii) the witness is not biased in any way. If any one of these conditions is not met, then the testimony is not acceptable. Of course, testimonial evidence that is not acceptable might still be true. A nonexpert might be right about the facts. A biased person might still be speaking the truth. Deciding that testimonial evidence is not acceptable is not itself reason to believe that it is false. In cases where our only evidence is testimonial evidence and we have determined that the testimony is not acceptable, the reasonable thing to do is to withhold belief.

## 4.7 NEWS REPORTS

The news media is a further special case of testimonial evidence. News reports, whether in newspapers, magazines, on TV or on the Internet, all involve a reporter making claims about some topic or other. The reporter might be writing about what happened yesterday on Capitol Hill or in a refugee camp in the Middle East or Africa. Or maybe the report is on economic conditions in Asia or in our local region. Whatever the topic, we should treat what we read or hear on the news in the way we treat other forms of testimonial evidence and we should be ready to ask the same questions. Is the report on an appropriate topic (one on which there are experts)? Is the reporter an expert? Is the reporter biased in any way? Let us look at each question in turn.

<sup>2</sup>Specter, M. (2004). Miracle in a bottle. *The New Yorker* (February 2, 2004), pp. 64–75.

Testimonial evidence is **inappropriate** if it is on a topic where there are no experts. This can happen where the facts are so complex that no one counts as an expert. Sometimes, news reports will make claims about the nature of causes of events where it is questionable whether anyone really knows what is going on. It is, of course, not always obvious to us when a topic is that complex. But in some kinds of cases, it is perhaps better to err on the side of caution. Here are two examples. The first concerns reports on the stock market. Reporters who cover the stock market not only report on changes in the value of various stocks as the day goes on, they sometimes offer explanations of why the markets as a whole are moving in one direction or another. The sharp drop was caused by fear that interest rates will go up; or the rise in stocks was a reaction to the morning news that the unemployment rate has once again crept up. These kinds of claims are almost never trustworthy. No doubt there is some explanation of the change in stocks values. But surely that explanation is enormously complex. The same is true, I think, when reporters offer explanations of complex international events, and this is especially clear in the case of wars. During the war in Iraq, it was regularly reported that the violence in Baghdad was getting worse in the middle of 2007, even as the Americans sent in more troops. These reports were difficult to assess. It is hard to know how to evaluate violence in terms of worse and better—if there are fewer attacks but each attack is more deadly, does that mean the violence is worsening? And because there were so many attacks every day, and because Baghdad is a huge city, it is hard to know everything that is going on. The complexity involved in measuring and defining the level of violence is so high that it might be best to treat claims like that made by reporters as inappropriate.

In general, testimonial evidence is acceptable only if the source is an **expert** on the topic. But this is almost never the case with news reports. Many reporters are trained in journalism schools. This means that they are trained in how to collect information and how to present it in various media. But it does not mean that they are trained in or well informed on the topics on which they report. Indeed, most reporters are not experts on what they are reporting on. This is why they rely on experts in the field when preparing their reports. They present the expert's testimony. In a way, this makes our task as critical thinkers even harder. For not only do we need to assess whether we should accept the reporter's account of the expert testimony, we have also to assess whether that expert testimony is itself acceptable. Cases when reporters rely on expert testimony are like double testimony! Just to make matters worse, reporters sometimes rely on witnesses who insist on remaining anonymous. The witnesses might have legitimate reasons to insist on this—perhaps their career or health depends on it. But this makes our task next to impossible: how can we assess whether the witness is properly trained, informed, and unbiased if we do not even know who it is? To some extent, perhaps, we can trust the reporter to tell whether the expert she has interviewed is competent and unbiased. But this is less than ideal.

If you think that some piece of testimony is not acceptable, then you need to say which of the four conditions was violated.

A news report is acceptable only if the reporter is well informed. It is not always easy to tell how hard the reporter worked to collect the information she used in her report. **Passive reporting** occurs when a reporter merely accepts testimonial or other evidence without doing background checks and without questioning the acceptability of the evidence. We know that most news reports are produced very quickly, and have to be quite short. Articles in the newspaper and stories on the evening news are unlikely to be as well informed as full-length documentaries. The time pressures are too great. But this means that we, as critical thinkers, have to decide whether the report is based on acceptable information. There are some things to look for in deciding whether a report is well informed: (i) Number and variety of sources. Did the reporter rely on one source or on many? In general, it is better if the reporter asked for many expert opinions. Did the reporter rely on a variety of experts? If all the experts are from a single organization, then there might not be enough variety in their testimony. (ii) Background and fact checking. Did the reporter do any background investigation of her own, or did she solely rely on sources? Without some background work, it can be difficult to know what questions to ask the experts or how to follow up on their answers. Did the reporter check whether factual claims made by her sources are correct? Or is she simply uncritically reporting what the source told her? The more reason there is to think the reporter relied on a large number and variety of sources and did some background and fact checking, the more acceptable the report will be.

Like any testimonial evidence, a news report is acceptable only if it is **unbiased**. In the case of news reports, bias is a potential factor at several different levels, starting with **reporter bias**. Reporters are under pressure to produce, just like anyone else. And we all know that salacious and juicy stories are more fun to read than dry and factual stories, even if the factual ones are intrinsically more important or newsworthy. "If it bleeds, it leads." Reporters are under pressure not just to report the news, but also to report what they think their audience wants to read or hear about. Very rarely, reporters even react to this pressure by making up the news, or by focusing their report on what they think will be most interesting or catchy.

**Corporate bias** is also possible. Newspapers and news stations are businesses and this means that they are in the business of making money. They do this by reporting on what their audience is interested in. Inevitably, this means that they leave some stories completely unreported. The latest celebrity arrest for drunken driving is covered in more detail than the thousands of children who die every day in refugee camps around the world. The fact that news organizations are in the business of making money also means that they are reluctant to report stories that would make their audience feel uncomfortable. A newspaper that published every day the names of all the children who die of hunger would not be very successful. News organizations are also careful not to offend the patriotism of their audience. News reports that focus on war crimes sometimes face criticism.

Finally, news reports are subject to **cultural biases**. This affects not just what stories are told but how they are told. We find stories about local events more interesting than stories about events in distant parts of the world. It is not easy to detect bias. Respectable news organizations work hard to draw a sharp line between news

reporting and editorializing, and some have ombudsmen whose job it is to keep an eye on and even report on the extent to which the organization is succeeding at being unbiased. Finally, it is always good advice not to rely on only one source for news. Reading different reports on the same event is the best way to avoid falling prey to biased reporting.

### EXERCISE 3

#### A. Comprehension Questions

- a. Under what conditions is testimonial evidence acceptable?
- b. What is the difference between being trained and being informed? Illustrate your answer with an example.
- c. If a witness is biased, does this make their testimony false? Give an example to illustrate.
- d. If a witness is testifying about his own personal observations, what are the critical thinking questions that we should ask before accepting his evidence?
- e. Under what conditions is testimonial evidence overridden and undermined? Use examples to illustrate your answer.
- f. What are some sources of media bias?
- g. What is passive reporting? Is it a form of bias? Why or why not?
- h. When are claims made in advertisements acceptable?

#### B. The following passages involve appeals to testimonial evidence. Determine whether the evidence is acceptable. If not, then identify which of the conditions is violated. Be as detailed as you can.

- a. One of Thomas Jefferson's most trusted advisors said that the United States should not trade with tyrants. We should take that advice and cut off all economic relations with tyrannical regimes around the world.
- b. There is milk in the fridge. I just called home and Joan told me that there is.
- c. I am failing this class. My teacher just told me so.
- d. The cookbook says to boil the eggs for 12 minutes to make them hard, so this is what I am doing.
- e. The newspaper just reported that the stock market will drop tomorrow, so I am selling all of my stocks now.
- f. The bank president says that there is no risk that his bank will go out of business, so I have decided to keep my money in it for now.
- g. The man at the garden supply store told me that this plant will thrive best in a shady spot, so I am going to put it underneath that tree.
- h. The man at the garden supply store told me that this hose will not leach lead poisoning into the soil. That is why I bought it.
- i. I think that you will really like this band. The girl at the music store said that they are the best all boy band since 'N Sync.

- C. The following passages involve attempts to undermine a witness' testimony. Using the concepts we have discussed in this chapter, explain whether the attempt is successful. If not, explain why it fails. Be as detailed as you can.
- a. The man at the garden supply store said that the plant thrives in shade, but he just stocks the shelves there. I do not trust him.
  - b. Our local politician says that we should revise campaign finance legislation to make it harder for politicians to be influenced. But he is just trying to raise funds from the "little guy." Do not trust him.
  - c. The regional coordinator of that environmental group gets paid to recruit new members, so we cannot trust what he says about the effects of global warming. He is just trying to get us to contribute.
  - d. The newspaper article says that the new highway is very dangerous. But I do not think the reporter has any background in that field. He cites a lot of expert reports, but he has no training for himself. We should not trust him.
  - e. The witness claims that Jones robbed the bank, and that she saw him leaving with the bag of money. But there were so many people coming and going that day that there is no way the witness could be reliable on this. No one could remember every face they ever saw.
  - f. The cookbook says that it is best to add fresh anchovies to the dressing, rather than canned ones. But the cookbook author owns a chain of fresh food stores, so she is probably just trying to increase sales.
- D. Look at a newspaper article. Using the concepts discussed in this chapter, assess whether the reporter is guilty of passive reporting. Be as specific as you can in your criticisms, and make sure to support them carefully.
- E. Look in the newspaper or on TV for five examples of testimonials. Assess whether the testimonial evidence is acceptable. Be as specific as you can, and be sure to support your conclusions with reasons.
- F. Look at today's newspaper and find a report on some international incident. Then go to the library or go online and find news articles from different continents on the very same incident. Compare the reports looking for significant differences.

## 4.8 MEASUREMENT

When we try to decide what to believe or what to do, we often rely on evidence from measurement. After all, one way to collect information on some subject or phenomena is to measure it. We have a huge number and variety of measuring instruments and tools at our disposal. We measure our own mass using a bathroom scale and that of objects in deep space using highly sophisticated instruments; we measure public opinion using surveys and questionnaires; we measure the intelligence of our children using IQ tests; meteorologists measure the speed and direction of traveling storms; we measure student performance using final exams. (What, in your view, do final

grades actually measure?) Businesses use personality tests to find out about their employees and to build better teams. It would not be much of an exaggeration to say that measurement plays as important and prevalent a role as observation or testimony in our reasoning about what to believe and do. Thinking critically about the acceptability of measurements involves asking some of the very same questions we asked about observation and testimony.

We need to use an appropriate scale when we measure something.

Here is an initial question: what is it, exactly, to **measure** something? Measuring usually involves assigning a number. But not always: Doctors sometimes ask children with sore throats to indicate the level of soreness using a chart of face drawings, ranging from a happy smiling one to one that is crying. The kids are using this chart to measure their pain, even though no number are involved. And assigning a number to something is not always measuring it. Counting, for instance, involves assigning a number. We count and find that there are five donuts on the table. Have we measured the donuts? We have not measured their size or mass. Have we measured their number? This sounds odd. We do, though, measure the size of crowds.

Calculating averages also illustrates the difference between measuring a phenomena and merely assigning it a number. As you may know, there are many different kinds of averages. Suppose that 100 students wrote a final exam and that each was assigned a grade from 0 to 100. The **mean** grade is the result of adding the 100 grades and dividing by 100. The **median** grade is that grade such that half of the grades are above it. The **mode** grade is the most common grade. These three averages vary independently of one another—changing any one of them might not change the other two. Which kind of grade measures average student performance depends on what question we are asking. More students might be in the B-range, even if the mean is C- and the median is a C+. The moral is clear: simply assigning a number to a phenomenon is not the same as measuring it.

When we measure something we assign a number and relate that number to a standardized scale of some kind. We say, not just that my weight is now 175, but that it is 175 **pounds**. The concert lasted not just 30, but 30 minutes. The restaurant review gives the new Indian restaurant not just five, but five stars. Understanding a measurement requires knowing which scale is being used. The acceptability of a measurement depends on the scale too. Some scales are appropriate for measuring some aspects of a phenomenon but not for measuring others. We can measure a liquid's temperature but not its mass using degrees Celsius. We can measure a car's velocity, but not its acceleration, using meters per second. In some cases, there is more than one appropriate scale, as in the case of temperature, which we can measure using either the Celsius, the Kelvin or the Fahrenheit scale. So, one question we need to ask in deciding whether some measurement is acceptable is whether the measurement employs an **appropriate scale**.

We measure something using a measuring instrument of some kind. We measure our mass using a bathroom scale; temperature using a thermometer; student performance using a final exam; voltage using a voltmeter; public opinion using a

questionnaire. Whether a measurement is acceptable depends on whether the instrument used in the measurement tends to yield accurate measurements. A measuring instrument is **reliable** only if it tends to accurately measure what it is supposed to measure. If a test is designed to measure a person's latent hostility, then it is a reliable instrument only if it does in fact provide accurate measurement of the latent hostility of people who take the test. A procedure for measuring the size of crowds at rallies and demonstrations is reliable only if it provides an accurate (or accurate enough for the purposes at hand) measure of the crowd size.

Measurements are acceptable only if they are made using a reliable instrument.

It is sometimes difficult to know for sure whether an instrument is reliable. Just because something is called an intelligence test does not guarantee that it really is a reliable instrument for measuring intelligence. Likewise, just because something is called a public opinion survey, this does not mean that it reliably measures public opinion. There is a deep and difficult methodological problem involved in determining whether a measuring instrument is reliable.

The only way to know for sure whether a measuring instrument is reliable is to compare its readings with those of a device that is known to be reliable. This is called "calibrating" the instrument. I can check to make sure that my meat thermometer is reliable by comparing its readings to one I know is reliable. But what if we do not know whether that second one is reliable? Do we need to compare its reading with those of a third instrument? Where would this regress end? Or what if we do not have another instrument for measuring the phenomena? If we have independent access to the phenomena being measured, then this is not a serious problem. Measuring the length of wooden boards is such a case, since we can more or less confirm by looking whether the tape measure is giving us the right reading. But with tests used to measure human intelligence or student performance, this is a serious problem, since we often have no independent way to measure the phenomena. We have no independent way to check to make sure that it really is the students' intelligence, as opposed to some other trait, that is being measured.

Thankfully, we do not need to sort out this difficult problem here. It is enough for us to be aware of the methodological problems facing measurement, and to be armed with the concepts needed to think critically about them. But in deciding whether a measuring instrument is reliable, there are two very important points to keep in mind here.

#### 4.8.1 Measurement Consistency

First, consistency is not a guarantee of reliability. An instrument is consistent if it gives the same readings on repeated uses. Whether a reliable measuring instrument must be consistent depends on the nature of the phenomena being measured. If it is a phenomenon that can change quickly between measurements, such as levels of sugar in a person's blood or the electricity usage of a building, then a reliable instrument might not need to be consistent, since its readings would have to change to keep track

of the changes in the phenomena. An instrument that always gave the same reading of a person's blood sugar levels throughout the day would be very consistent, but that might actually be a reason to think that it is not reliable, since a reliable instrument for measuring blood sugar should give different readings at different times of the day. On the other hand, if the phenomenon does not change rapidly between measurements, then a reliable instrument would have to be consistent. If my bathroom scale gives five different readings within a few minutes then this is good reason to think it is no longer reliable. If a drugstore pregnancy test gave different readings every 5 minutes, then we would have good reason to question its reliability. So, whether consistency is a virtue in measurement depends on what it is that we are measuring. An inconsistent instrument might be reliable, and a consistent instrument might be unreliable.

A measurement instrument is consistent if it gives the same measurement on repeated uses. Consistency is not a guarantee of reliability.

If we know that the phenomenon we wish to measure does not change very quickly, then a measuring instrument would have to be consistent in order to be reliable. The easiest way to tell whether an instrument is consistent would be to use it to measure the same case several times. I can test the consistency of my new food scale by repeatedly putting the same bag of potatoes on it to see whether it always gives the same reading, which it would if it was consistent. But just to make matters a bit more complicated, in certain kinds of cases, an instrument designed to measure a very stable phenomenon might nonetheless not be consistent. This is the case for instruments designed to measure human intelligence or knowledge, like the SAT test or standardized tests in grade school. A subject who has written a standardized test once knows all the questions and so it is likely that she would do better the second time around. It is not that taking the test has by itself made her smarter or more knowledgeable (although she does know more about the SAT test); if we did repeat the test, we would expect different outcomes. But this need not be a reason to think that the test is not reliable. The relations between instrument reliability and instrument consistency are obviously quite complex.

#### **4.8.2 Measurement Precision**

Second, precision is not a guarantee of reliability. The precision of an instrument is a matter of how finely graded its readings are. A scale that measures only in pounds is less precise than one that measures in ounces. But precision is no guarantee of reliability. If the more precise scale is off by more than 2 pounds, then the less precise one may be more reliable. Still, precision is alluring. Sometimes, the results of public opinion polls involve decimal points, such as that 56.3% of the population is opposed to some policy. This precision seems to suggest that the instrument they used to measure public opinion must be very accurate. But this is a mistake. Precision is neither necessary nor sufficient for reliability.



### **DECIDING WHAT TO DO: MEASURING COSTS AND BENEFITS**

Deciding what to do requires comparing the anticipated costs and benefits of competing proposals. But this comparison is often very difficult. Here are two of the main reasons for this.

*Incomplete Information.* Sometimes, it is just not possible to know what the anticipated costs and benefits of a given proposal are. Unfortunately, global warming is a good example. The nature of the Earth's climate is so complex and its dynamics so subtle that it is very difficult to know which of the many proposals for counteracting its effects will be most efficacious.

*Incommensurability.* Sometimes, the costs and benefits of a proposal are difficult to measure, and so difficult to compare and balance. Decisions that affect quality of life are like that. There is no easy way to measure, compare, and contrast pains and pleasures.

Inevitably, we sometimes must decide on a course of action in the absence of complete information or in the presence of incommensurability. In such cases, it is good to make the ignorance or incommensurability as clear as possible. If one must act in the face of ignorance and incommensurability, it is better to do so knowingly than blindly.

### **CONFUSING THE LIKELIHOOD OF SOMETHING WITH ITS VALUE**

We need to be careful not to confuse how good or bad some consequence would be with how likely it is. Winning the lottery would be terrifically good, but it is extremely unlikely. So when comparing the costs and benefits of alternative proposals, you need to factor in the likelihood of those costs and benefits, as well as their value. A course of action that promises very high value but at very low probability (e.g., spending your money on the lottery), might not be as good as one that promises a high probability of moderate value (investing cautiously for the long run). This distinction between the likelihood and the value of a cost or benefit is especially important to keep in mind when assessing risk. Being killed in a plane crash would be far worse than getting a flat tire on the highway, but the odds of the flat are much higher than those of the crash. Flat-tire insurance may be a better investment than crash insurance.

### **CRITICAL THINKING MISTAKES: APPEAL TO IGNORANCE**

It is a mistake to believe something just because you do not have evidence that it is false. This is a mistake because a bit of investigation might show that it is false,

and thinking critically requires looking for evidence when one can. One form of this mistake is to discount or ignore potential costs or benefits of a proposal just because you do not know how to measure or compare them. It is important for critical thinkers to do what they can to discover these costs or benefits.

## 4.9 SURVEYS

Opinion surveys are a very familiar form of measurement. Pollsters and researchers ask people for their opinions on everything from politics, to sexuality, to sports, to food, to history.

### **SUMMARY: ACCEPTABLE SURVEY EVIDENCE**

Evidence from a survey is acceptable just in case:

- (i) The survey questions are not ambiguous, biased, loaded, or otherwise bad;  
and
- (ii) Those surveyed are properly trained and informed on the topic of the survey;  
and
- (iii) There is no researcher or subject bias.

Researchers even use questionnaires to learn more about the nature of happiness, as we will see in a moment. Often, the researchers are interested in measuring the opinion of a large population—perhaps all Americans—and administer their survey to a sample of the general population. In Chapter 6, we will consider when reasoning using samples is valid, but here we will stay focused on the question of when the results of a survey are themselves acceptable.

It pretty much goes without saying that happiness is a prime motivator in our life. Everybody wants to be happy and most of us want to help make others happy too. Companies and governments want people to be happy with the products and services they provide. Economic theory assumes that we are all “happiness maximizers,” that we can be expected to act in ways that we think will make us the happiest. If we choose the burger over the green salad, then that shows that we believed that we would derive more happiness from the burger. In moral theory, Utilitarians hold that the moral value of an action, practice, or policy is a matter of how much overall happiness that action, practice, or policy would produce as compared to alternatives. Public discussions about whether to implement one kind of government program or another often turn on questions about which program would have the best results, where this is usually a matter of how much happiness it would produce. Building a new bridge will make truckers and commuters happier, but it will make those living near the waterfront less happy. Deciding whether to build the bridge requires thinking hard about happiness. Given the central role that happiness plays in our thinking about

our own lives and in our thinking about public policy, it might come as a bit of a surprise that we actually have little idea how to measure it.

Suppose that we were trying to measure happiness by asking people a simple question like this: On a range of 1–10, how happy would you say that you are? The question is simple enough that most people would be able to answer it. It includes a numerical range so we can get a number to deal with. If we asked enough people, we could even calculate an average of some kind (a mean or a mode or a median?), and figure out how happy the average person is. We could do the same for people under 30 and college kids and grandparents, and thereby learn quite a lot about how happiness varies from one group to another. But in fact there are good reasons for doubting whether this simple instrument is really a valid instrument for measuring human happiness.

One difficulty is that the question is a **bad question**. There are (at least) two problems with the way it is worded. First, it is **ambiguous**. Opinions about what it is to be happy vary quite widely. It is not just that people find that different things in their life make them happy. Some people find that a career-centered life makes them happy, while others derive happiness from their hobbies or from a vibrant community of family and friends. Variation from one person to another—and from one culture or time period to another—in views on the essential elements of a “good life” is not at all surprising. But people who agree on the elements of a happy life might still disagree about what it means to be happy. Is happiness, as John Stuart Mill thought, a kind of feeling, like an emotion or a sensation? If being happy is the same as feeling pleasure, then maximizing happiness is a matter of maximizing the amount of pleasure in the world. Or is happiness, as Aristotle thought, a matter of achieving and sustaining a kind of balance in one’s life, independently of one’s actual feelings? In that case, maximizing happiness would require organizing the elements of one’s life—work, family, friendship, love, sport, pleasure, etc.—in a special way. Of course, people who agree on this view of the nature of happiness might also differ over just what elements are required and how to properly balance them. Disagreements over the nature of happiness are, as we know from Chapter 2, disagreements over the meaning of the word “happiness.” So, the question is bad because it is ambiguous.

The other reason to think the question is badly framed concerns the point we saw earlier that measuring something is not the same as simply assigning it a number. Measuring requires both a number and a scale. But what is the scale in our proposed question? What are the units? Happiness units? But what are they? Moreover, what is the relation between the numbers on the scale? Is it like the relation between numbers on a scale for measuring mass, where something that weighs 6 lbs is twice as heavy as something that weighs only 3 lbs? Is a person who is 3 on the happiness scale half as happy as one who is a 6? Or is the relation between the numbers on the happiness scale like the relations between numbers on a temperature scale, where 20°C is not twice as hot as 10°C? Because we do not know the answers to these kinds of questions, we do not know what the question is asking us. This is a second reason to think that our proposed survey question is a **bad question**.

Survey evidence is acceptable only if the survey is reliable at measuring opinion.

There are other ways that a survey question can be bad. A question is bad if it uses **charged** or **slanted** words. If we want to know a person's opinions about the morality of abortion, it will not do to ask whether they are in favor of the deliberate killing of unborn humans. A question is bad if it presupposes something such that no matter how one answers it, one will be saying or implying something controversial. The classic case of "Have you stopped cheating on your exams?" nicely illustrates the problems with a loaded question. Answering "Yes" implies that you once did cheat; answering "No" implies that you still are cheating. Lawyers questioning a witness in a trial have to be very careful that their questions do not presuppose anything controversial. The goal is always to get as good a handle as possible on the truth, and badly framed questions are an obstacle to the pursuit of truth.

But even if we could agree up front on the nature of happiness and on the scale for measuring it, there are other reasons for doubting the validity of that simple survey question as a measure of human happiness. Because the survey asks people to report on their own level of happiness, it is a measuring instrument that relies on testimonial evidence. As we have seen, testimonial evidence is not acceptable if we have reason to think the witness is **incompetent** or **biased**. So the instrument is reliable only if people are pretty good at telling whether they are happy and only if they are not motivated to exaggerate or lie about their levels of happiness. As it turns out, we have some reason to think that people are not very good at detecting their own levels of happiness. We are all familiar with the way that people live in denial. People sometimes lie to themselves about their true feelings, fabricating stories about themselves and their lives to paper over the problems that lie just beneath the surface. And if happiness is a matter of having a well-balanced life, it is likely that some people will not be good at telling when their life is well balanced. Researchers have found that people's reports on their happiness are not very consistent and can be influenced by irrelevant factors. In one study, researchers found that subjects tended to report a higher degree of happiness if the sun was shining or if they had just found a bit of money (left on purpose by the researcher.) Subjects also focus on the best events and ignore or fail to remember the negative events in their daily life. All of this suggests that some amount of error and ignorance about one's own happiness is inevitable. If so, then we have some reason to doubt whether people are good at telling whether they are happy, and this is some reason to doubt the validity of the proposed measuring instrument.

Evidence from an opinion survey is acceptable only if those surveyed are properly trained, properly informed, and unbiased.

Surveys of popular opinion on policy matters also raise questions about whether the people being surveyed are properly trained and informed. It is one thing to have an opinion, but quite another to have an educated or justified opinion. Consider a

poll asking for opinions on a certain plan for the changing the health-care system. Most people pay very little attention to the details of government proposals. And most of those proposals are so complex, and most of the problems they are nominally intended to remedy are so multidimensional that it is not very likely that most people will have an educated opinion on the matter. Even if the survey accurately reports people's opinions, these opinions are themselves of only little value since they are probably not educated opinions.

### CRITICAL THINKING ABOUT HUMAN HAPPINESS

Researchers on human happiness have proposed a different series of questions to measure a person's happiness. Instead of asking something like our proposed question, some have used what is called a "day reconstruction method", where the subject is asked to describe the sequence of events in their day and report on the emotions they felt during each event. These researchers hope that this method will lead to a more accurate report from the person about just how happy they are with their day-to-day life. So there may be ways to formulate surveys that will enable more accurate reports from subjects. (For a discussion, see "The not so dismal science: how economists measure happiness," Tim Harford, *Slate*, <http://www.slate.com>.)

### CRITICAL THINKING MISTAKES: BAD QUESTION

It is a mistake to ask a question that is ambiguous, contains charged or slanted words, or that hides a controversial presupposition. It is a mistake because it makes it harder to know what the person answering the question really believes.

But suppose that people were in fact quite good at telling whether they are happy. Their testimonial reports would still be unacceptable if we had good reason to think that they are **biased**. The possibility that they are biased would be a case of what in critical thinking we call **subject bias**. It is probably not right to suspect that most people would lie about how happy they are. We do know that people tend to exaggerate what economic class they are in, so that people well in the working class tend to report that they are in fact in the middle class. So it is probably right that we should expect that some people will be motivated to overstate their happiness, either out of shame or in order to impress. There is also in the case of this kind of survey, the possibility of **researcher bias**. The researchers asking the questions might themselves introduce a bias into the reports, either in the way they ask the questions or in other subtle features of their interactions with the subjects.

The best way to eliminate the risk of subject or researcher bias is to make the test **double blind**. A study is double blind if both the subjects that are being studied and the researchers doing the studying are ignorant of key facts about the study. In

the case of drug trials, researchers ensure that they study is double blind by ensuring that neither the research subject nor the researcher knows who got the placebo and who got the trial drug. One way to reduce the risk of bias in a survey is to add a lot of questions on irrelevant topics. That way, neither the subject answering the questions nor the person hired to ask the questions knows what the researchers are really trying to find out about. This ignorance reduces the risk that bias will distort the results.

#### EXERCISE 4

##### A. Comprehension Questions

- a. What is the difference between mode, median, and mean? Use examples to illustrate when each would be valuable.
  - b. What is the difference between measurement validity, measurement consistency, and measurement precision? Use examples to illustrate your answer.
  - c. Suppose that you had a watch that was designed only to tell the hour and another designed to tell the time to the closest second. Which is most consistent? Which is most valid? Which is more precise?
- B.** Make a list of 10 measuring instruments you use every day. Order them from the most valid to the least valid. Order them from the most precise to the least precise. Order them from the most consistent to the least consistent.
- C.** Make a list of 5 phenomena that you do not currently know how to measure, but which you think should be measureable. Pick one of them and think of a way to measure it.
- D.** Look through the newspaper and find 5 articles that make measurement claims. Using the concepts we discussed in this section, identify some questions that need to be considered in deciding the measurements are acceptable.
- E.** In the following passages, identify the premises and conclusion(s). Then identify whether the argument relies on evidence from observation, testimony, or measurement. Finally, using the concepts we have discussed in this chapter, identify some questions that need to be considered in deciding whether the argument's premises are acceptable. Be as specific as you can.
- a. The roast is probably done. I just took out the thermometer and it read 150 degrees Fahrenheit, and the cookbooks says that a roast is done when it reaches 145 degrees Fahrenheit. It also smells like it is starting to burn.
  - b. It is safe to go into the water. The city tested the water yesterday and the level of potentially infectious chemicals was very low.
  - c. The new standardized math test is really accurate. We used it on several students from grade three and they all got the same score.
  - d. You have heard the witness testify that she saw the defendant enter the bank on the time of the crime. And you have been presented with ballistics evidence

proving that the defendant's gun was used in the shooting. You have no choice but to find the defendant guilty.

- e. Our new American Motors sedan is the safest car in America. It scored a record 98.79 overall safety rating in our crash tests, the highest of any car we have ever manufactured. Just watch what happens in this video of a head-on car crash. See how the airbags inflate in time to prevent serious injury. This shows that the car is as safe as can be.
  - f. Beauty Derm skin lotion removes 87.95% of wrinkles. Our clinical tests show it.
  - g. South Park High School is the best school in the city. Its students recently scored higher than every other student in the city on the new standardized tests.
- F. For each of the arguments in the previous question, decide whether the premises provide sufficient support for the conclusions. If not, suggest ways the argument might be strengthened.
- G. Look in the newspaper for articles reporting on surveys. Find the questions that were asked, and assess whether they are ambiguous, loaded, biased, charged, or otherwise bad. Be as specific as you can, and do not charge a survey with asking a bad question unless you can support your claim.

### CHAPTER SUMMARY

Reasons to believe or do something have to be acceptable. They are acceptable if they are from a **reliable** source and there is no **undermining** or **overriding** evidence. A source of evidence is reliable if it tends to provide true or accurate information most of the time. Observation, testimony, and measurement can be reliable sources of reasons.

## 4.10 CRITICAL THINKING IN PRACTICE

### 4.10.1 Critical Thinking Mistakes

*Appeal to Ignorance.* It is a mistake to believe something just because you have no evidence that it is false. This is a mistake because a bit of investigation might show that it is false, and thinking critically requires looking for evidence when one can.

One form of this mistake is to accept some evidence just because you do not know of any overriding or undermining evidence. Critical thinkers should look for overriding and undermining evidence, before relying on some evidence.

Another form of this mistake is to discount or ignore potential costs or benefits of a proposal just because you do not know how to measure or compare them. It is important for critical thinkers to do what they can to discover these costs or benefits.

*Unacceptable Testimony.* It is a mistake to accept testimony from a witness if the topic is inappropriate, the witness is not properly trained, or not properly informed, or if the witness is biased. It is a mistake because such evidence is not acceptable. Testimony is appropriate only on topics for which there are recognized experts. An expert must be properly trained and properly informed. And a witness must not be motivated to lie about or exaggerate the facts.

*Bad Question.* It is a mistake to ask a question that is ambiguous, contains charged or slanted words, or that hides a controversial presupposition. It is a mistake because it makes it harder to know what the person answering the question really believes.

#### 4.10.2 Critical Thinking Strategies

*Trust, But (Be Prepared to) Verify.* Most critical thinking theorists agree that evidence can be acceptable even if we have not proven that its source is reliable. Instead, they recommend the following: evidence from some source is acceptable *unless one has strong reason to think the source is not reliable*. Trusting our sources is a default right, as it were. But we should not let ourselves get carried away. For we know that some apparent sources are not reliable at all, and others even ones that are reliable can still yield mistaken evidence. To borrow Ronald Reagan's remark about the proper attitude to take to enemy superpowers: trust, but be prepared to verify.

*Measure Twice, Decide Once.* The goal of critical thinking is knowledge, and this means that it places a premium on getting the right answer. To this end, it is better (to paraphrase a familiar wood-working lesson) to measure twice and decide once. This applies just as much to observation and testimony as it does to measurements. It is, all things considered, more prudent to collect more evidence than less evidence, and evidence from different sources is best of all. Finally, it is just as important to consider possible undermining evidence as it is to consider possible overriding evidence. For part of what makes critical thinking reflective is that it requires us to think about what makes a source of evidence reliable.

#### 4.10.3 From Theory to Practice: Applying What We Have Learned

**4.10.3.1 Thinking Critically about Ourselves** This exercise is designed to help you reflect on your strengths as a witness. In Chapter 1, you compiled a list of five or six character traits that you think are essential to being a morally good person. Being a trustworthy person might not have been on that list, but I think that most of us would agree that we strive to be someone others can trust. We have seen that testimonial evidence is acceptable only when (i) it is appropriate; (ii) the witness is properly trained and informed; and (iii) the witness is not biased in any way. This exercise is designed to have you reflect on the extent to which you meet these criteria. As always, the more sincere effort you put into it, the more you will get out.

For 2 days, observe yourself as you answer people's questions or give them information, or tell them your beliefs. As you do this, be willing to think critically about whether you are meeting the standards for being an acceptable witness. Do you ever offer an opinion as if it were the truth on a subject where there may not be



experts? Do you ever offer a firm opinion on a subject where you are not really fully trained or fully informed? Do you ever let biases creep into your responses?

One way to do this exercise is to keep a journal for 2 days, pausing at noon and before bed, to reflect and describe a few events from the day. Use the Testimonial Rubric we discussed in the text.

**4.10.3.2 *Thinking Critically in the Classroom*** This exercise is designed to help you identify the different sources of evidence that you rely on in studying or engaging in your chosen discipline. Some disciplines rely on one source of evidence much more than others. (Philosophy involves virtually no measurement at all, and little direct observation.) In Chapter 1, you compiled a list of five or six of the tasks that you are required to do in your chosen discipline that require critical thinking. That list might have included such things as performing measurements, collecting observational data, and doing factual research. Make a list of the kinds of evidence that you rely on in studying or that someone actively engaged in your chosen discipline would regularly rely on.

- (i) Direct observation
- (ii) Reliance on testimony
- (iii) Measurement

**4.10.3.3 *Thinking Critically at Work*** This exercise is designed to help you think critically about the sources of information you or the organization you work for rely on for success. All organizations rely on testimonial evidence and measurement in order both to achieve their organizational goals and to make adjustments to their internal operations. In Chapter 1, you compiled a list of tasks that you regularly perform at work that require critical thinking. Look over that list and identify those tasks that require you to collect or rely on testimonial evidence or measurement. Pick one of each kind, and do the following:

1. Assess how reliable those sources typically are.
2. Think of some ways to improve their reliability.
3. What other testimonial sources or measuring instruments would help you with your task?