Chapter 13

How Are We To Know?

Theories and Beliefs

People have invented a dazzling array of theories, beliefs, stories, and myths to help them explain, understand, and manipulate the world around them. All around us are phenomena that demand explanation. The sun rises in the east and sets in the west. Why? And will it always do so? One theory imagined that the sun was pulled across the heavens by an invisible chariot. Nowadays we explain its apparent "travel" by the earth's rotation. Falling rain was once explained as coming from holes in the roof covering the earth through which water occasionally leaked. Now we use theories of condensing atmospheric moisture and frontal systems to account for rain and other phenomena of weather and climate. We have theories to account for fossils found in rocks, for the sun's almost limitless energy, for earthquakes and volcanoes, for the diversity of life forms, for mental behavior, for the birth and death of stars, and for essentially everything else we can perceive about the universe.

We also construct theories about everyday experience—both social and personal. Why are crime rates falling in New York City? Why did Booth assassinate Lincoln? Why is my child falling behind in school? Why is unemployment so high (or so low)?

What good are theories—besides helping us to understand experience? Mainly, they help us achieve what we want in the world. A theory about the rising and falling of the waters of the Nile helped the ancient Egyptians plan their planting and harvesting of crops. Theories about rocket propulsion and orbital mechanics help us launch and control satellites and spacecraft. Theories about whether a particular type of reader will understand a sentence help us shape that sentence. In order to serve us in these ways, a theory must be capable of making predictions.

Theories come in many forms. Scientific theories often involve mathematical formulas, such as $E = mc^2$ or Maxwell's equations. Stories and myths can be thought of as theories. Computer programs, such as those that are used to help predict the weather, are theories. Even simple statements of belief, such as "John Jones is trustworthy," are theories. All of them help us make predictions.

How are theories invented? The astronomer Johannes Kepler commented: "The roads that lead man to knowledge are as wondrous as that knowledge itself."[1] Sometimes new theories are elaborations or combinations of already existing theories. Our everyday experience provides material out of which to construct theories. Experience with small, hard, round objects provided the basis for early theories about the atom. Seeing wave interference patterns in a pool of water may have helped inspire the wave theory of light. Analogies like these are very important for theory invention (and for creativity in general), even though they can also lead us astray. For example, the idea that waves require a medium for propagation led people to think that electro-magnetic waves had to travel in some invisible, space-filling "luminiferous aether." The concepts out of which we construct our theories sometimes aren't advanced enough for the job at hand, but we persevere in our inventions anyway, using whatever concepts we do have, because mysterious phenomena itch to be explained.

We marvel at the ingenuity of humans in coming up with theories. But ingenuity alone cannot be relied upon to produce theories that always serve us well. Indeed, human creativity has produced many useless and even some harmful theories. How should we decide in which theories to trust? How are we to know? Because we use theories to help us predict, we should

[1]Quoted in Arthur Koestler, *The Sleepwalkers: A History of Man's Changing Vision of the Universe*, p. 332, New York: The Macmillan Company, 1959. (This particular quotation is a paraphrase of a quotation Koestler uses earlier in the book on page 261, namely "The roads by which men arrive at their insights into celestial matters seem to me almost as worthy of wonder as those matters in themselves.")

evaluate them by how well they actually make predictions. *A good theory is a theory that predicts well.* The same thing goes for our everyday beliefs.

Any method for evaluating theories is faced with a trade-off. If our criteria for theory acceptance are too strict, we are likely to exclude some useful theories. Missing out on useful theories is the price of extreme skepticism. On the other hand, if we want to be sure that we do not exclude useful theories, we are likely to accept many questionable, even useless or harmful ones as well. Believing bad theories would be the price of extreme credulity. Science usually operates toward the skeptical end of this spectrum, whereas people differ widely in their willingness to accept beliefs. Either kind of error, missing out on good theories or accepting bad ones, has its consequences. That's why it's important for us to be rigorous in evaluating our everyday beliefs, just as it is for science to be careful in evaluating its theories.

Robot Beliefs

We can gain perspective about the role of our own theories and beliefs by considering how robots create, modify, and use theirs. Does it seem strange to think that robots have beliefs? In everyday usage, we often ascribe beliefs to computers, which after all are the "brains" of robots. It's common to say things like "the word-processing program *thought* I wanted that lower case 'i' capitalized." Or "the chess-playing program *believed* it had me checkmated."

Robots have computational models of their environments stored away in their memory structures. They choose many of their actions based on what those models predict the consequences of their actions to be. It's not too much of a stretch to say that those models are a robot's *beliefs* about its world. Analogously, our own theories, beliefs, and stories can be thought of as *models*. Like robots, some of our models we carry along with us in our brains, and some of them are external to us in writings, maps, computer databases and other formats.

Robots have several different kinds of models. One type is called "declarative" because such models are in the form of "sentences." A computerized database is a declarative model because its tabular information is a set of sentences (of a special form, of course). For example, a factory robot's database might contain information equivalent to the sentence "engine replacement parts are stored in aisle 17." Sometimes the sentences are organized in large hierarchical webs called "semantic networks," which encode relationships and dependencies among the sentences.

Many psychologists and neuroscientists think that humans also have a declarative form of memory, organized in semantic networks analogous to the robotic ones just mentioned. Most of what we call "beliefs" seem to be represented in the cerebral cortex. One theory about the cortex views it as a theory-inventing prediction machine."[2] In any case, however they are represented in our brains, we can think of our theories and beliefs as being in declarative form because we so easily use sentences to mention them.

It is worth mentioning that, in addition to declarative models, robots and humans have "procedural" ones too. Declarative models know "what," whereas procedural ones know "how." Procedural models encode their information directly in the routines that control action. Usually, the kinds of actions that use procedural models are those that require very fast sensor-motor coordination—so fast that there is insufficient time to access and reason about declarative information before deciding on an action. A soccer player—either robot or human—needs to execute actions that depend on predictions about where and when a soccer ball will arrive after being kicked. To make fast predictions and execute appropriate responses, knowledge about "ball physics" needs to be encoded directly in the relevant perception and action routines. There are important connections between declarative models and procedural ones. For example, a golf swing can benefit from declarative verbal instruction followed by practice. But because our primary concern here is with evaluating theories and beliefs, we concentrate on declarative models.

What do robots *do* with their models? Just like humans, mostly they use them to make predictions to better guide their actions. It's important for robots to be able to predict the main effects of actions taken in various situations. For this purpose, a precondition-action-postcondition format is

[2] Jeff Hawkins (with Sandra Blakeslee), *On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*, New York: Times Books, 2004.

used extensively in advanced robotics. For example, a robot might have coded in this format information such as "if the door is closed but not locked (the precondition), and I open it (the action), then it will be open (the postcondition)." We humans use information of this sort also: "if the paint in the can is white, and if I paint the wall with it, then the wall will be white." In order to make plans to achieve goals consisting of chains of actions, both robots and humans need to know the preconditions and the effects of the various actions in the chains.

Where do robot models come from and how are they modified? Programmers install some of them directly during construction. A robot designed to work in a parts warehouse might have a map of the warehouse installed at delivery. Analogously, many animals are born with some environmental models hard-wired in by eons of evolution. Perhaps humans have some innate models also.

There are several ways to update a robot's declarative model by additions, modifications, or deletions of sentences. Human programmers can add software updates directly. A mail delivery robot's database could be re-programmed, for example, with the information that John Jones is now in Room 135. This kind of model modification might be called "robot brain surgery."

There are also methods for the robot itself to make updates during interaction with its environment. Robots have perceptual apparatus such as visual and auditory systems, touch sensors, and sonar systems. It is possible for robots to sense anything that engineers can build sensors for. Information about the environment gained by sensory perception is used to make appropriate model changes. An important example involves learning about the effects of actions. Such learning attempts to discover what aspects of a situation can be used as a precondition to guarantee that the action just taken in that situation will have the effect just observed.

Analogously, humans can see, smell, taste, hear, and feel. They acquire much of their information about their world by being told things, by reading, and by interacting with their environments. Some of what they learn this way, they accept as a new belief, and sometimes they use what they learn to modify existing beliefs. Indeed, the process works both ways—what is perceived is influenced by what is already believed.

These are the *only* ways for robots and humans to acquire models of their world: by having them pre-installed at birth, by being told, by reading, and through their other sensors. These are the only portals to reality!

What about "reasoning?" Can't we (and robots) find out about the world through reason? Some early philosophers, including Plato and his followers, thought that we could. Some people may still believe that reason gives us a window on reality. Reason does allow us to manipulate the information we get through our senses so that it can be cast in more useful forms. And reason can be used to detect inconsistent theories. But, it does not tell us anything new—it only re-arranges what we already know.

Like robots, we humans are machines—albeit very, very complex ones, still little understood. Robot abilities and limitations regarding their models suggest that we humans are similarly limited. We use this point of view to help clarify our thinking about our own beliefs.

Reality and Truth

One might be tempted to say "a good theory is one that is *true*." But then there is the question, "true according to what?" Well, true according to the way reality actually *is*. But there is a fundamental problem with trying to use the concept of truth as a way to evaluate theories. We can never be sure of what reality *is*. In the Middle Ages, for example, many people were certain that the sun (and everything else, for that matter) revolved around the earth. The geo-centric Ptolemaic system was proclaimed to be a *true* model of reality, and people could be, and were, punished for denying that so-called truth. But now, science—in its attempts to construct ever better descriptions of reality—routinely overturns previous theories and adopts modified ones or entirely new ones instead. All scientists acknowledge that scientific theories are subject to change. In fact, even to be a "scientific theory," it must be possible to *imagine* experiments that, in principle, might overturn it.

The problem is that we can't apprehend reality directly, even though the things we perceive *seem* real enough. It might seem strange to claim there "really" is no such thing as a rock after stubbing your toe on one. But we use our perceptual apparatus (even a pain from the toe) to construct mental *models* of the world. The very concept of "a rock" is *our invention*, carefully crafted after much experience. It's how we carve up and describe reality. As a slang expression would have it, "reality doesn't know from rocks—it just *is*." Using our model-building apparatus and invented concepts and informed by our perceptions, we can only say things *about* reality. We can never say what it *is.* And, what we say about it is always subject to revision.

And, some of our descriptions of reality, unlike those that use concepts like "rock," cannot be related easily to common, everyday experiences. In quantum mechanics for example, weird phenomena such as "superposition" and "entanglement" have no satisfying, intuitive explanations—yet they are described mathematically in ways that make consistently confirmed laboratory predictions and may even be used someday in super-fast computers. There are physicists who claim not to be bothered by this lack of a mental picture; they belong to what some call the "shut-up-and-calculate" school. So long as the model makes good predictions, who cares? Some thinkers around the time of Galileo distinguished between what they thought was the *physical reality* of the geo-centric, Ptolemaic universe and the helio-centric system of Copernicus. The former was *real*, whereas the latter was merely *useful* for calculating predictions needed for navigation. Our view is that both are descriptions, and that the ultimate test of a description is its ability to predict.

Solipsists and some philosophical relativists deny the existence of a reality that is independent of who is affected by it. But, we are realists and therefore believe that there is *something out there* that affects our perceptions, and we believe that we can develop useful theories about it—whatever it might be. As we continually expand our abilities to perceive through newly invented sensory apparatus, so also do we continually enlarge our descriptions of the reality opening up to us.

There have been many attempts by philosophers to define the concept of "truth." Some hold that a statement is "true" if (and only if) the statement "corresponds" to the way the world is. But, as argued here, we can't determine what the world is. Because it is impossible to say whether or not a statement corresponds to reality, this very idea of correspondence is meaningless.

Even so, it's convenient to be able to use the word "true" as shorthand to label those beliefs and theories in which we have very high confidence. We might say, for example, that it is "true" that Columbus sailed across the Atlantic in 1492—because we are very confident of it. We are sure (that is we predict with confidence) that any future evidence that might turn up will support the statement and none will refute it. Similarly, we can say that the theory of evolution is "true." Therefore, rather than attempting to use some pre-defined concept of truth to say that this-or-that belief is true, we say that a belief is true only after labeling it as a highly confident belief.

Such a view of truth doesn't acknowledge any "absolute truths." Because belief labeling is an individual activity, "truths" are relative to the individual, even though reality itself is not. Nevertheless, we individually and we as a society may have supremely strong beliefs that are stable and highly resistant to change because they continue to make confirmed predictions.

The Scientific Method

In the last few centuries, a set of cultural practices has taken shape for evaluating theories. These practices are responsible for the development of extremely useful theories. Taken together, they have come to be called "the scientific method." Many aspects of the scientific method are refinements of "common-sense" techniques which humans have used for millennia. Hunting bison by stampeding them over a cliff, for example, was observed to be usually successful, and thus persisted as a "theory" of bison hunting. Children, even small infants, in their constant probing and testing employ parts of the scientific method.[3]

The main idea of the scientific method is that a scientific theory must make predictions that can be tested. To be a "good" theory, the predictions must be repeatedly borne out by experiment, or at least by observation when staged experiments are impractical or impossible.

The requirement that a scientific theory must make testable predictions is augmented by the additional requirement that it must be possible, in

[3]Alison Gopnik, Andrew Meltzoff, and Patricia Kuhl, *The Scientist in the Crib: What Early Learning Tells Us About the Mind*, New York: William Morrow and Company, Inc., 1999.

principle, that such tests might fail. That is, the theory must be "falsifiable." (Adhering to the view that nothing can ever be *completely* certain in science, even disbelief, we take "falsifiable" here to mean susceptible to being weakened to the point of abandonment.) If a theory is not falsifiable, it is not a scientific theory. There are many ways in which a theory might not be falsifiable. One of the most common ways is that the theory has an inexhaustible supply of variable parameters that can be adjusted, *after* an experiment, to save the theory from a failure in the experiment. Extra-sensory perception (ESP), for example, makes testable predictions, but experimental failures are usually explained away by an unending set of excuses. Thus, most scientists think that ESP, as promoted and propped up by its adherents, is not falsifiable.

People have a lot of beliefs that are not falsifiable, and therefore such beliefs cannot be considered to be scientific. Sometimes people even base actions on non-scientific beliefs. These actions might have consequences that are harmless, good, or bad for the people concerned (and for others too). Most theories about immortality—that is, life after death—cannot be tested, so they are not scientific theories even though many people might believe them.

It must be admitted that even our belief in *reality*—that there is something out there that impacts our perceptions—cannot be falsified. We can always cook up reasons to describe whatever so-called "reality" throws at us. This inexhaustible supply of explanations for our perceptions makes the idea of an alleged "cause" of these perceptions, namely "reality," an unfalsifiable belief and therefore an unscientific one. But the very reason that we continue to be successful at generating useful, prediction-generating explanations makes the belief in reality compelling.

Not only is the scientific method used to test theories that describe present reality, but it can also be used to test theories about past reality. There are theories about the birth of the universe in a "Big Bang," for example. One consequence of that cosmological theory is that stars and galaxies should all be receding from the earth (and from each other), and that consequence is consistent with modern astronomical observations. (We can't do any staged experiments involving a repeat of the "Big Bang," so passive observations must suffice.) Much of science, such as geology and paleo-anthropology, is concerned with past events. So, of course, is the subject of human history. We can evaluate theories about these past events by making predictions about what the theory claims we might find by subsequent "digging." (Some people call the predictions made from theories about past events "postdictions.")

Scientists sometimes make a distinction between "facts" and "theories." They use the word "fact" to describe specific results of an experiment or observation. Back in 1895 for example, an experimental situation was set up in which Mrs. Roentgen placed her hand on a certain photographic plate. When the plate was developed, there was a "shadow" of a hand on the plate. Results like that are called "facts." Theories, on the other hand, are more general and attempt to account for a number of such facts and also predict what additional facts might be observed under a number of different situations. A theory about x-rays, for example, explained the hand shadow and could also be used to predict (among other things) that many different objects would cast similar shadows.

This distinction between facts and theories is sometimes misunderstood. It is commonly (but mistakenly) thought that a fact is necessarily more reliable than a theory. But so-called facts can become less credible than a highly credible theory. Thus to say that something is "just a theory" does not cast doubt on it. Quantum mechanics is just a theory, yet it is one to which most scientists affix the label "true."

So, the main way to evaluate a theory is to investigate what it predicts and to test those predictions. Just because the theory survives one test or any number of tests does not guarantee that it will survive all future tests! Someday, a test of an as-yet-unthought-of prediction might destroy the theory—or at least cause it to be substantially modified.

Because interpreting the results of experiments is subject to experimenter bias, and because the experimental setup itself may be inappropriate or faulty, it is extremely important in science that experiments to test predictions be repeated by independent investigators. Experimental results must be repeatable in order for confidence in the theory to be increased (in the case of results matching predictions) or decreased (in the case of results different from predictions).

Besides surviving rigorous experimental tests, scientists look for something called "consistency" in their theories. The simplest way in which two theories might be inconsistent is for their predictions to be different.

For example, if a theory about global warming predicts that Alaskan glaciers will recede and some other theory (about global warming or about anything else, for that matter) predicts that Alaskan glaciers will advance, the two theories are inconsistent. Eventually, what actually happens to Alaskan glaciers might resolve the matter, but in the meantime one (or both) of the theories must go or at least be revised to remove the inconsistency.

Logical and mathematical reasoning also play important roles in checking the consistency of theories. Scientists often look for theories which can be used to deduce other theories. Or, to put it another way, they seek to *reduce* one theory to a more detailed one that implies the first. For example, Newton's laws of motion, together with his theory of gravity, can be used to deduce Kepler's theories about planetary motion. If Newton's theories logically implied that planetary orbits were anything other than Kepler's ellipses, then Newton's and Kepler's theories would be inconsistent.

When one theory can be used to derive another, the first is sometimes cited as an "explanation" or "cause" of the second. Gravity and the laws of motion are said to explain or cause elliptical orbits. Theories about atomic and molecular interactions explain various chemical theories, and these explain various biological theories. In fact, finding an explanation for a theory in terms of a more detailed theory lends support to the theory being explained. (And, not being able to find any such explanation counts against a theory.)

In science (and also in other human affairs), theories and beliefs have "domains of applicability." Thus, Newtonian mechanics and quantum mechanics make different predictions about atomic and sub-atomic phenomena, so they are inconsistent. We accept quantum mechanics and reject Newtonian mechanics as theories of the very small because the predictions of quantum mechanics are borne out (so far) by experiment whereas the predictions of Newtonian mechanics are not. Nevertheless, we accept Newtonian mechanics as a theory of larger-scale phenomena (such as spacecraft trajectories) because its predictions about these phenomena involve practical calculations and are consistent with the accuracies we can obtain in experiments designed to test these predictions. To complicate matters further, Newtonian mechanics and relativity are inconsistent for large-scale phenomena at very high velocities and near very large masses.

So, we restrict the applicability of Newtonian mechanics further to what might be called "middle-scale" phenomena—not super-small, not super-fast, and not super-massive. To the frustration of many scientists, relativity and quantum mechanics are also inconsistent. Physicists are still searching for a satisfactory "theory of everything" that will not exhibit these inconsistencies. Even if such a theory is found, scientists and engineers will still use specialized theories in those areas where they provide useful predictions obtainable through practical calculations.

So, to preserve all-important consistency, scientists have learned to pigeon-hole their theories into pockets in which inconsistencies do not arise—even though there may be global inconsistencies over all the pockets. Such is the state of twenty-first century science.

It is common for people too to have inconsistent beliefs. Ralph Waldo Emerson wrote: "A foolish consistency is the hobgoblin of little minds ... With consistency a great soul has simply nothing to do."[4] But perhaps he should also have said that one of the things great souls could do, like scientists, is to partition their beliefs into pockets where the inconsistencies won't get them into trouble. Others have said that one mark of a wise person is the ability to hold inconsistent beliefs. Perhaps so, but again, it seems important that inconsistent beliefs be insulated from each other in separate places informing different kinds of activities.

If theories are not inconsistent on their face, one should look for inconsistencies among their logical implications. Because these implications can be arbitrarily lengthy, inconsistencies can stay undiscovered for long periods—maybe forever. One way to aid the search for inconsistencies is through joint work and criticism. Different people, with different points of view and with different motivations, should thoroughly analyze, discuss, and argue about theories and beliefs. Criticism can also reveal hidden assumptions and weak points that can be further explored and perhaps modified.

The philosopher Karl Popper thought that a critical attitude is fundamental to our attempts to make models of the world. He wrote:

> There is only one element of rationality in our attempts to know the world: it is the critical examination of our theories. These

[4] Ralph Waldo Emerson, "Self-Reliance," *Essays, First Series*, 1841.

> theories themselves are guesswork. We do not know, we only guess. If you ask me, 'How do you know?' my reply would be, 'I don't; I only propose a guess. If you are interested in my problem, I shall be most happy if you criticize my guess, and if you offer counterproposals, I in turn will try to criticize them.'[5]

Of course, people are not always happy to have their beliefs criticized. Perhaps they should welcome criticism if they seek beliefs that are self-consistent and useful for robust predictions.

At various times in the history of science, different theories have been proposed whose predictions about the same experimental or observational results seem (more-or-less) equally on the mark. Two such theories may well be inconsistent with each other, so we cannot hold both. But how are we to choose between them? The main criterion that scientists have adopted, and that seems to work well, is to choose the simpler of the two. To favor the simplest theory *among those which are equally consistent with observations* is known as the principle of "parsimony." As an example of an application of this principle, Kepler's helio-centric solar system with planets moving around the sun in elliptical orbits is far simpler than Ptolemy's epicycle-laden geo-centric one with the sun and planets going around the earth. Even though for a while, the two systems made passably good predictions about planetary observations, Kepler's was simpler and came to dominate. Now, of course, there are many other reasons to adopt the Keplerian one and to reject completely the Ptolemaic one.

Most people seem to prefer the simplest explanation for everyday experiences and phenomena also (among those explanations that make equally good predictions). There are even statistical arguments for the claim that a simpler theory is more likely to make more accurate predictions about future observations and experiments than a more complex one would—given that both are consistent with previous observations and experiments.

[5]From *Popper Selections*, David Miller (ed.), p. 30, Princeton, NJ: Princeton University Press, 1985.

Uncertainty

Most theories and beliefs have shades of gray about them. It has already been mentioned, for example, that there are theories about which scientists are quite confident (such as quantum mechanics and evolution) and others about which they are less sure (such as string theory). There are various ways to describe our degree of certainty about a theory or belief. For example, we might use phrases such as "it is likely that" or "it is virtually certain that" or "it is doubtful that." Or, we might assign numerical probabilities or odds to theories. We can imagine a spectrum between 100% certainty about a theory and total disbelief in it. "It is likely that" might then correspond to a probability of, say 80% or so. Virtual, but not complete, certainty might then involve a number like 99.99% or so. Because, in science, we are always willing to entertain the idea that a theory might be overturned, we would never assign 100% probability to a theory, for belief in such a theory (according to a technical result in probability theory) could then never be weakened—whatever the evidence.

For theories that are less credible than "virtually certain," we aren't able to use strict logical methods for checking for inter-theory consistency. If we have a theory, A, say, in which we have 85% confidence, a theory, B, say in which we have 70% confidence, and if theory A "supports" theory B (with some numerical way of quantifying that support), then we need to have some alternative way of checking, in this setting, whether our uncertain theories A, B, and the support of one for the other are "consistent."

Because probability numbers are useful as measures of confidence, one turns to probability theory for an appropriate elaboration of the idea of consistency. Given all the relationships among a set of uncertain beliefs, do the probability numbers associated with them and their relationships satisfy the rules of probability? If so, they can be said to be "consistent."

For computers, a useful way to characterize the relationships among uncertain propositions is to represent them in a network called a Bayes net. If propositions A and B jointly support a proposition C, then there will be a link in the network between A and C and between B and C with numbers expressing the degree to which A and B support C. Skipping the technical details of these networks, suffice it to say that a large set of related beliefs can be represented as a large interconnected web. Changing the strength of one of the beliefs in the web would imply that the strengths of several related ones should change also—in order to ensure that the entire set of strengths, expressed as probabilities, satisfies the laws of probability. Computational processes, such as those that might be used by a robot, have been developed to update the probabilities of beliefs in a Bayes net in response to changes in one or more of them.

Although no one knows how humans maintain their webs of beliefs, some of the processes that humans use for reasoning with uncertain information are analogous to those used in Bayes nets. Here are three very common modes:

1. *Evidential reasoning.* If some particular phenomenon is thought to cause a second, then observance of the second provides some evidence for the first. That is, it would increase our strength of belief in the first. Additionally, two such pieces of evidence combine to increase our belief in the first even more. For example, observing a long-term increase in ocean temperatures would lend support to a belief in global warming (one of its possible causes). So would the observance of long-term trends of earlier spring travels pole-ward of migratory birds. Because, in fact, both of these pieces of evidence have been observed, we do have increased confidence that global warming is occurring. A combination of evidence is often sufficient to confirm our belief in something that the individual pieces of evidence alone might not.

   Medical diagnosticians use this sort of reasoning. If a certain disease causes a certain symptom, then observance of the symptom increases confidence in a diagnosis of the disease, and observing several of the symptoms increases it further. On the other hand, lack of some symptom known to be caused by a disease rules out the disease (or at least weakens belief in it). For this reason, this type of reasoning is sometimes called "diagnostic reasoning."

   Science also uses this mode. Confirmations by several independent experiments or observations testing the predictions of a theory combine to increase confidence in the theory. And, failure of repeated experiments to confirm a prediction made by the theory overturns the theory (or at least weakens belief in it).

Trial lawyers appeal to this mode of reasoning, for example, when they claim that "the weight of evidence" leads to the conclusion that a defendant is guilty "beyond a reasonable doubt." Again, lack of evidence has the opposite effect.

2. *Causal reasoning.* Here we have something like the reverse of evidential reasoning. If some particular phenomenon is thought to cause a second, then observance of the first increases confidence in the second. Insurance companies might use this kind of reasoning to set rates. Because smoking is thought to cause lung cancer, believing that a person smokes increases confidence in the belief that he or she will develop lung cancer—justifying, perhaps, an increased insurance premium.

   Science uses this mode when they appeal to a highly believable causal explanation for some derivative theory. To return to a previous example, a confident belief in a theory about the properties and effects of greenhouse gases serves to increase our belief in global warming.

3. *Explaining away.* This type of reasoning uses causal reasoning to defeat competing evidential reasoning. Suppose we have reason to believe that A causes C. Then, observing C would increase, by evidential reasoning, our belief in A. Now, suppose we discover (or learn about or decide to take into consideration) the idea that B also causes C, and, additionally, we have strong belief in B. Then, by causal reasoning, we might confirm or even increase our belief in the observed C, but there is less reason now to think that A was the cause of C (because there is an alternative, highly believable cause, B). So, our degree of belief in A is decreased. A, as a possible cause of C, is *explained away* by B.

   This type of reasoning is quite common and very important. We can use it, for example, to decrease the credibility of the so-called "intelligent design" theory. Even though intelligent design is not a scientific theory (because there are no imaginable experiments that might falsify it), some people believe in it nevertheless because they think it is or was the cause of the complexity of life forms. But the processes of evolution can also cause this complexity, and most scientists think the combination of independent pieces of evidence for evolution is quite compelling—so compelling that it explains away intelligent design, grouping it along with other creation myths. Intelligent design is thus doubly suspect. First, it's not a scientific theory, and second, it's explained away by evolution.

Where do these probability numbers come from? This question is important both for robots and for people. With regard to people, there are some highly credible foundation beliefs arising from our perceptions ("to see is to believe"—well, most of the time). Other beliefs to which we give high strength, at least the ones that don't conflict with each other, arise from what trusted authorities tell us. These beliefs, we might feel, are "virtually certain." We make guesses about the strengths of other, less certain beliefs. In a Bayes net, as used by a robot, such guesses are called "subjective probabilities."

It doesn't seem too difficult for humans to come up with subjective probabilities for their beliefs. Gamblers who quote odds for various events do this all the time. Odds can be converted into probabilities; for example stating that the odds of some event happening are 3 to 2 is the same as saying that the probability of it happening is 60%. Of course, one would only bet on something whose eventual outcome could be determined—a case in which the prediction implied by the bet is either verified or not. Beliefs backed up by bets satisfy the "falsifiability criterion" for scientific beliefs because the odds given in the bet acknowledge that it is possible to lose.

What about infinite odds? Such would imply a probability of 100% or certainty, and nothing is certain. Well, aren't some things certain? Isn't it certain that the sun will rise tomorrow? Almost, but we ought to allow for some very, very small chance, for example, that an asteroid will obliterate the earth first. These small chances preserve our ability (though perhaps rarely used) for us to change some of our virtually certain beliefs.

In addition to bets between individuals, there are internet-based "markets" at which one can buy futures contracts for prices that reflect the odds on various propositions. In principle, such markets could be established to buy contracts on all sorts of beliefs—even beliefs that might never be established or refuted. People's guesses about the probabilities of the beliefs might change over time, so there is the possibility of buying low and selling high. If there were such a market in 1600 C.E., the market price of the Ptolemaic theory would have been falling.

Acknowledging that our beliefs have degrees of uncertainty about them should make us all the more receptive to critical discussion and continued testing. We think such discussion and testing is the basis for developing sound beliefs—that is, beliefs that engender more accurate predictions.

Religion and Faith

During and after the European Renaissance, theories about reality that were evolving in the Western world gradually became less constrained by prior and existing religious and philosophical beliefs. People began to question theories that were protected by various interpretations of "scripture" and by religious authority and sought to apply the developing scientific method instead. Thus began many conflicts between science and religion.

Perhaps a kind of truce can be effected between these two cultures by separating religious beliefs from scientific ones. How might such a separation be made? Some religious beliefs are not falsifiable and for that reason they are already separate from scientific theories. Examples might include re-incarnation, divine right of kings, trans-substantiation, and the existence of the Holy Ghost (and the existence of the other two members of the "Trinity" also for that matter).

As in the example of "intelligent design" mentioned earlier, the supposed explanatory power of some non-scientific religious beliefs may give them some evidential support, and some people may therefore hold them with high confidence. Even though many such religious beliefs can be explained away by highly credible scientific ones, there need be no conflict so long as the religious beliefs are not put forward as scientific theories.

Some non-scientific theories and beliefs held by individuals may contribute to (or may even be necessary for) a satisfying and healthy emotional life. Or, they may be a kind of artistic expression. (The philosopher George Santyana once said, presumably in an artistic vein, "There is no God and Mary is His mother.") We don't argue against holding these sorts of beliefs—unless they cause harm. But considerations of emotional health or aesthetics do not and cannot make such beliefs "true." To be labeled "true" (as that word is used here), a belief has to be consistently good at making non-guaranteed predictions, implying that it is a *falsifiable* belief, which non-scientific beliefs are not. (Incidentally, we can have scientific beliefs about non-scientific beliefs. For example, we might have a theory that belief in angels leads to a lower incidence of hypertension. That's a scientific belief because it makes testable predictions. But here is where care is needed; even if belief in angels were to lower one's blood pressure, that's not evidence for angels. It's evidence that a belief in angels lowers blood pressure!)

These kinds of non-scientific religious beliefs are what their adherents call "faith-based." But simple "faith" can never be the basis for preferring one of these beliefs to another. There are so many different non-scientific beliefs to choose from! How does a "person of faith" decide? Probably mainly by upbringing and by community. It's tragic that the different communities fight over their different faiths and attempt to inflict their faiths on others.

Let's turn now to those religious beliefs that are (or at least under some interpretations might be) falsifiable. Potential conflicts between religion and science remain for these. For example, theories about the earth being four billion years old or so lead to predictions consistent with geological findings whereas theories of religious fundamentalists about the earth being 6,000 years old or so do not.

Another example concerns the efficacy of "intercessory" prayer. Do prayers for an ill person (done without that person's knowledge) facilitate recovery? The nineteenth century English anthropologist, Francis Galton (a first-cousin of Charles Darwin) thought that experiments could be conducted to shed light on this question:

> The efficacy of prayer seems to me a simple, as it is a perfectly appropriate and legitimate subject of scientific inquiry. Whether prayer is efficacious or not, in any given sense, is a matter of fact on which each man must form an opinion for himself. His decision will lie based upon data more or less justly handled, according to his education and habits. An unscientific reasoner will be guided by a confused recollection of crude experience. A scientific reasoner will scrutinise each separate experience before he admits it as evidence, and will compare all the cases he has

selected on a methodical system. [6]

Religious beliefs of this sort, where evidence can be gathered to compare against predictions, should be treated in the same way as other scientific beliefs are treated—subject to evaluation through experimental tests and criticism. It should be *expected* that such religious beliefs would be in conflict with other scientific beliefs, just as competing scientific beliefs are forever in conflict with each other. That's science!

Should One Believe All This?

Even though "all this" does involve some beliefs, it's mainly an "attitude." It's a way of approaching the problem of how to think about beliefs. Similarly, the scientific method is not a belief; it's an approach for coming up with useful descriptions of reality. Each "believer" can decide to adopt this way of thinking about beliefs and theories or not.

Of course, one can (and we do) have beliefs about this attitude. One's beliefs about it might influence a decision to adopt it or not. We believe, for example, that it offers an effective way to evaluate all of our beliefs including this one—effective in the sense that beliefs evaluated in this way are predicted to make better predictions than other approaches that we might adopt.

To follow the advice of the approach itself, we should subject this very belief to all of the evaluation criteria the approach recommends—including comparison with other approaches, criticism, experimental test, parsimony, and so on. As part of doing so, we cite as evidence that employing the scientific method has been most effective in giving people vast powers over their environments. By writing a book about the approach, we invite discussion and criticism.

Of course, there is more to life than making predictions. Joseph Campbell has said that what we all seek is the "experience of being alive."[7]

[6] Francis Galton, "Statistical Inquiries into the Efficacy of Prayer," *The Fortnightly Review*, No. LXVIII, New Series, August 1, 1872. (Quotation taken from: http://www.abelard.org/galton/galton.htm.)

[7] Joseph Campbell (with Bill Moyers), *The Power of Myth*, p. 5, New York: Doubleday, 1988.

Being in a position to enjoy that experience, however, depends on our ability to make good predictions.