Chapter 1

A Conversation About Beliefs

Last year, Sam spent the summer in Cambridge, Massachusetts—staying with his uncle Nick, a philosophy professor at M.I.T. On warm summer evenings, they talked about many things—global warming, evolution, history, medicine, religion. Frustrated with all of Nick's "on-the-one-hand"s and "on-the-other-hand"s, Sam kept asking, "But, how are we to *know*?"

"I'm afraid we have to settle for beliefs," Nick replied. "Even if you told me you *knew* something, I would have to say that it's just another one of your beliefs—albeit a strongly held one."

"Are you saying that all the things I know are really just beliefs?" Sam asked.

"That's all we have," Nick answered, "just beliefs."

"But when I say I believe something," Sam said, "I usually mean I'm less than certain about it. And when someone else says he or she believes something, I take that to mean it's just his or her opinion."

"Personally, I think it makes sense to use the word 'belief' to cover the whole range," Nick said. "At the certainty end, there are beliefs in which we have very, very high confidence. As shorthand, we could even say we *know* those things or that they are *true*. At the other end, there are beliefs that we are quite unsure about. In the middle, there are beliefs in which we have various degrees of confidence. The advantage is that we can use one word, with appropriate qualifiers like believe strongly, believe somewhat, and so on, for the whole range."

"But I use the word 'believe' in other ways too," Sam said. "For example, I might say that 'I believe in my country,' or 'I believe in freedom'."

"The verb, 'believe', is sometimes used that way," Nick agreed, "but the noun, 'belief', usually refers to a statement that we are willing to bet on, given the right odds of course. Let me rephrase your question from 'how are we to know?' to 'how are we to decide what to believe and with what conviction?"'

"Well, ok," Sam agreed, "let's suppose that's my question."

"I learned a lot about how to think about beliefs by working with Mia, a colleague of mine who designs robots," Nick said. "She claims some of her robots have beliefs. It's instructive to think about the problem of what to believe from a robot's point of view. In fact, Mia and I are working on a book about beliefs. Would you like to join us in some of our discussions? Come to think of it, your participation might even help us as we think about the next draft."

"Sure," Sam said, "but what do robot beliefs have to do with human beliefs?"

"An important question," Nick answered. "Building robots that have beliefs might contribute to a better understanding of human beliefs. Comparative studies help illuminate the subjects being compared. Take planetary geology for example. Studying geologic processes on Mars and other planets leads to generalizations that give us a better understanding of Earth's geology. Studying different languages contributes to useful generalizations about linguistics."

So Nick brought Sam along to his next meeting with Mia.

"Mia, I'd like you to meet my nephew, Sam," Nick said. "He's a bright college student who's interested in how we humans come to believe things, and I thought he might like to join us as we talk about our book."

"I don't know about the 'bright' part," Sam said, "but Nick says you design robots that have beliefs. Is that right? It sounds amazing!"

"Nice to meet you," Mia said. "Yes, I think it's useful to talk as if robots had beliefs—doing so helps us predict how they might act in various

circumstances.[1] One of my most advanced robots, Gio, not only has beliefs, but he even believes that he has beliefs—and he can introspect about them too. Gio is my 'exhibit A' whenever I talk to people about beliefs. Let's have him join us. He can record our conversation too if we'd like."

"Ok," Sam said, "but why do you call him a 'he'? Shouldn't 'he' be an 'it'?"

"People often impute a definite gender to objects," Mia said. "After all, many people speak of boats and ships as being feminine. And many languages use gendered nouns. We've just gotten used to speaking of Gio as a 'he'."

"Ok, bring him on," Sam said.

Mia invited Gio out of the lab and introduced him to Sam exactly as if he were a human.

It took a little time for Sam to get used to talking to a robot, but he finally asked, "Gio, do you agree that you have beliefs?"

"Of course," Gio answered. "I have a lot of beliefs about my world and how it works."

"How do you get beliefs and what do you do with them?" Sam asked.

"I use my beliefs to make predictions." Gio answered, "The predictions are useful for deciding on actions."

> [Belief] ... that upon which a man is prepared to act.
> —Alexander Bain, Psychologist[2]

"What just flashed blue on your screen, Gio?" Sam asked.

"I can access a lot of information through the Internet," Gio replied. "That's something I found that I thought might be interesting."

"Gio often comes up with a quote or a picture on topics we talk about," Mia explained. "Sometimes he reads out loud what he finds, and sometimes he just puts up something on his screen."

[1] Such ascription has been called the *intentional stance* by Daniel Dennett in: Daniel C. Dennett, *The Intentional Stance*, Cambridge MA: MIT Press, 1987.

[2] Quotation taken from Louis Menand, *The Metaphysical Club: A Story of Ideas in America*, p. 225, New York: Farrar, Straus and Giroux, 2001.

"How interesting," Sam said. "Does he always find something useful and important?"

"There is a lot of information available on any subject you might mention," Gio answered. "I try not to overwhelm you with too much, so I make a random selection from among what I think is relevant. I have some favorite authors though. Quoting from them helps me narrow down what I tell you."

"What Gio finds is generally germane to our conversation," Mia said. "But he's just a robot, so he might turn up with most anything. And, of course, there is a lot of nonsense on the Internet."

"So, tell me, Gio," Sam asked, "do you *know* anything, or do you agree with Nick that all we really have are beliefs?"

Gio replied, "Occasionally I do say I *know* something, but I just mean that I *believe* it very, very strongly. All I really have are beliefs. All of them have degrees of credibility—some have a lot, some have much less."

"Gio, like many robots, can be overly precise about how he uses words, Sam," Mia said. "Actually, he seldom uses 'know' because when we programmed him, we weren't sure exactly where to draw the line between having him know something and having him believe it with less strength."

"If I believe something that's *really* true, can't I then say I *know* it?" Sam asked.

"You can *say* anything you want," Gio said, "but I wouldn't understand what you meant by 'really true'. I use the word 'true' as a label I attach to my strongest beliefs."

"I think we are getting too philosophical too early in our discussion," Nick interrupted. "Let's postpone talking about reality and truth until a bit later."

"Ok," Sam agreed as he turned to Gio and asked "Well then, how do you decide whether your beliefs are any good?"

"Good beliefs are ones that help me make good predictions," Gio replied. "Checking predictions involves my perceptual system—vision, hearing, touch, and so on. If my perceptions match what my beliefs predict I would perceive, the prediction is verified. But, if I'm surprised, I need to change my beliefs."

"Gio uses his beliefs to predict what the effects of his actions will be so that he can take those actions that will achieve his goals," Mia said. "Because being able to predict is so important, Gio assigns high credibilities to those beliefs that consistently help him make good predictions."

"That's right," Gio said, "For example, I put high credibility on my belief that the mail room is in the basement because every time I use that belief to predict that I will find the mailroom in the basement, I find it there."

"Nick and I think we humans are in pretty much the same boat as Gio," Mia said. "We live and act in the world, and we use our beliefs about the world to help us get what we want. To do that, our beliefs need to help us make accurate predictions."

"My beliefs are stored in a *knowledge base* or *KB*," Gio volunteered.

"Wait!" Sam exclaimed. "Even though you all say Gio has only *beliefs*, he says he has a *knowledge* base."

"It's just what it's called," Mia replied. "They're still beliefs. Even Nick and I use the word 'knowledge' in this informal way—to mean a body of beliefs we have about a subject, even beliefs that we are not completely sure about."

"Is Gio's KB anything like my computer database of names and addresses?" wondered Sam.

"It's something like that," answered Mia.

"My KB has a huge collection of beliefs about many, many things—all in the form of statements," Gio said. "There are statements like 'robots are machines,' 'a Ford is a brand of automobile,' and 'General Robotics manufactures robots'."

"Gio has millions of these kinds of statements," Mia said. "They are arranged throughout several parts of his KB—all very well organized and indexed so that he can usually get at the ones he needs quickly. Gio's KB is much bigger, more flexible, and better organized than most run-of-the-mill computer databases."

"Do you really have sentences like those in your KB, Gio?" Sam asked.

"Gio's KB doesn't actually contain English sentences," Mia replied. "His KB uses a special kind of computer language that his reasoning and

action systems can work with."

"Well, however they are represented," Gio said, "they constitute my beliefs about the world."

"And you say you have more confidence in some of your beliefs than you do in others?" asked Sam.

"Yes," answered Gio, "and the credibility of many of my beliefs is so high that I label them 'true'."

"Well, what about the ones you are less certain about?" Sam asked. "How do you measure credibility anyway?"

"I do it with probabilities," answered Gio. "Here is an example. Someone told me yesterday that the bridge I usually use to go across the Madison River would be closed today, but I'm not sure I trust that information. I've decided to believe it at around the 40% probability level."

"How did you decide that?" asked Sam. "Why not 30% or 50%?"

"I think Mia should answer that one," Gio said.

"Gio's beliefs are interrelated in a kind of web," Mia said.[3] "For example, if Gio had a fairly strong belief that no bridge repairs are undertaken on Saturdays, and also believed that today is certainly Saturday, then he would have a fairly strong belief that the Madison River Bridge would be open today. Now, if someone whom he trusts says that bridge is actually closed today, all of these beliefs would have to 'fight it out' amongst themselves computationally to arrive at a conclusion about the bridge being closed."

"The result is 40%," Gio said.

"The web of all Gio's beliefs must be reasonably consistent," Mia went on. "They should all hang together. Gio has numerous computational mechanisms that check a new statement against other beliefs—also taking into account the reliability of the source. The end result is that the new statement gets assigned a probability number that harmonizes with the probabilities of his other beliefs. The same kind of thing goes on for all of his perceptions."

[3]The web that Mia has in mind might be similar to the informal notion discussed in: Willard Van Orman Quine and Joseph S. Ullian, *The Web of Belief*, Random House, New York, second edition, 1978.

"I do those computations very rapidly," Gio said. "Usually, I'm not aware that I'm doing them."

"Well, Nick," Sam said, "you claim that we humans believe some things quite strongly and others not so strongly. Do you think we use probabilities also?"

"We don't usually use numbers like Gio does," Nick replied. "And most of the time we aren't consciously aware that we are assigning strengths to our beliefs—it's mainly sub-conscious. When we do think about it, we consider the reliability of the source of information, whether it fits in with other beliefs, whether it is verified by later experience, and how surprising it is, among other things. Different people have different standards for deciding what to believe and with what conviction."

"But some people use a more disciplined approach," Mia said.

"Mainly the more disciplined approaches are used by people in various professions," Nick said, "like science, or law, or the study of history. Most scientists use what has come to be called the scientific method for evaluating scientific theories. As another example, consider courts of law. They have long-standing traditions to guide them in admitting and considering evidence and reaching conclusions. And historians have devised methods for helping them produce accurate descriptions of past events."

"We think the scientific method is especially useful," Mia said.

"But just consider all the things that people believe," Nick suggested. "How varied and interesting they are. There are historical beliefs like 'Hannibal crossed the Alps with elephants,' and 'Hitler invaded Poland in 1939,' and 'The Irish visited North America in the ninth century'."

"I've never heard the one about the Irish," Sam said.

"And, of course, we have many scientific beliefs," Nick said.

"You aren't going to call the stuff scientists know just beliefs too are you?" Sam asked.

"Many scientific theories are highly credible," Nick agreed. "But there are also scientific theories that are much more speculative—in fact scientific beliefs are spread out along the entire spectrum."

"Some people might mention their religious beliefs," Nick continued with his reverie on human beliefs. " 'Moses received the Ten

Commandments on the top of Mt. Sinai,' 'People are reincarnated after death,' 'Mother Earth and Father Sky created the first people,' 'The world was created in 4004 B.C.E.,' 'God cares about each and everyone of us and directs our lives'."

"How do we decide about those?" Sam asked.

"It's pretty much a personal matter," Nick replied, "but we'll probably talk about religion later if we continue these discussions."

"There are economic beliefs too," Nick went on. 'The stock market crash of 1929 and the ensuing depression were caused by over speculation in the late 1920s,' 'We can spend our way out of a depression,' 'Bond prices have an effect on stock prices'."

Continuing, Nick gave examples of beliefs with legal consequences as well as ordinary, everyday beliefs of individuals: " 'John Smith owns lot 27 in tract 4,' 'Don Jones is guilty of aggravated assault.' 'Mary Adams has gone to Florida,' 'Bill will meet me at 4 p.m. today at his office'."

"People must have pretty large KB's to keep all those beliefs in," Gio guessed.

"There are many more human beliefs than Mia was able to put in your KB, Gio!" Nick said.

"Do you think that people's beliefs are stored in KB's?" Sam asked.

"Maybe they are stored in neural structures that are something like Gio's KB," Nick speculated. "Even though people use a 'natural language' like English, Chinese, or French, when they *mention* their beliefs to you, many psychologists think that our actual beliefs are encoded in a form called *mentalese*, which our thinking systems use. If people actually have things like KB's, I'm sure they are much more sophisticated than even Gio's."

"I'm glad you brought up all those examples, Nick," Mia said. "If we are going to talk about how we come to believe things, we ought at least to remind ourselves about the variety of different things that humans believe."

"We base actions on beliefs like those," Nick said. "That's why it's important to evaluate our beliefs."

"Although a lot of our actions *are* based on beliefs," Mia added, "it must be admitted that some are not."

"I agree," Gio said, "although I usually use my beliefs to decide how to act, when I'm pressed for time I often find myself automatically executing some kind of reflex-type action."

"I guess I sometimes act impetuously too," Sam said.

"Reflexes produce the action more quickly than thinking about beliefs does," Nick said. "As you know, if you touch a hot object, your arm withdraws without requiring any conscious thought about it. Many human actions are evoked automatically by deep emotions, desires, and fears. And, people often talk about how they acted based on a hunch or a 'gut' feeling. There are even psychologists and neuro-physiologists who think that people cook up rational explanations later for why they acted. They say that people often confuse cause and effect by imagining that a particular belief *caused* an action rather than being a rationalization *resulting* from the action. The neuro-physiologist Benjamin Libet and his colleagues have done experiments implying that humans sometimes initiate an action *before* they could possibly know they are going to do it! And then they mistakenly claim that they knew they were going to act before they acted."[4]

"Here's something about unconscious acting by the biologist Edward O. Wilson," Gio said. "I'll read it to you."

> The self, an actor in a perpetually changing drama, lacks full command of its own actions. It does not make decisions solely by conscious, purely rational choice. Much of the computation in decision-making is unconscious—strings dancing the puppet ego. Circuits and determining molecular processes exist outside conscious thought. They consolidate certain memories and delete others, bias connections and analogies, and reinforce the neurohormonal loops that regulate subsequent emotional response. Before the curtain is drawn and the play unfolds, the stage has already been partly set and much of the script written.[5]

[4]Benjamin Libet, Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl, "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-potential)," *Brain* 106, pp. 623-642, 1983.

[5]Edward O. Wilson, *Consilience: The Unity of Knowledge*, p. 119, New York: Alfred A. Knopf, 1998.

"Nevertheless," said Nick, "we do base many of our actions on our beliefs, both as individuals and as societies."

"Sure," Mia agreed, "we decide about diets, health-care, finances, education, child-rearing, friendships and many other things based on our beliefs about those subjects. And if our beliefs are ill considered, our decisions about those things could lead to harm."

"Take health care, for example," Nick said. "The medieval belief that possession by demons caused mental illness led to ineffective, but often cruel, 'treatments'. For a long time, medical practitioners believed that blood-letting was an effective treatment for certain diseases. In fact, it seriously weakened many patients. Even today, many people have erroneous beliefs about the curative powers of certain herbs. Some herbs, taken in combination with other medicines, can be harmful."

"Nick and I have been talking about global warming," Sam said. "If a society believed that the earth's temperature wasn't really increasing, it wouldn't take the steps needed to combat it."

"We could list several examples where bad beliefs cause harm," Nick said.

"Such as, casino gamblers believe their 'good luck' can beat the bad odds," Mia volunteered.

"And guardian angels will protect you from harm," Sam said.

"As the folk saying goes," Nick said, "it's not so much what you don't know that hurts you, but what you know that ain't so. That's why it's important to evaluate beliefs carefully and continuously."

"Ok, so how do we do that?" Sam asked.

"We'll be talking about a number of ways," Nick replied, "but first we should mention some influences that aren't very helpful in getting us useful beliefs."

"Like what?" Sam inquired.

"Well, there's wishful thinking, for example," Nick said. "Sometimes people believe things just because it feels good to believe them. That's called *credo consolans*—I believe because it's consoling."

"If I thought that the only reason I believed something was because it

was consoling, I might begin to question that belief," Sam said. "Then the consolation would disappear."

"Consolation has nothing to do with whether or not *I* believe something," Gio said. "But, maybe it's different for humans."

"People are different from you, Gio," Mia said. "You're a purely rational agent, but people are an interesting mixture of the rational and the irrational."

"I've noticed that," Gio said. "Here are some song lyrics I found."

> truth is hard and tough as nails,
> that's why we need fairy tales.
> I'm all through with logical conclusions,
> why should I deny myself illusions?[6]

"I'm glad you didn't try to sing it for us, Gio," Sam said.

"Profound emotional experiences also affect our beliefs," Nick said. "And then there is the influence of one's culture and upbringing. People believe some things because they were brought up with those beliefs and they've never critically examined them."

"Questioning the beliefs we were brought up with can't hurt," Mia said. "It might reinforce some worthwhile ones and jettison others."

"The economist Robin Hanson has an interesting analogy about beliefs," Nick said. "He says that the beliefs people hold are like the clothes they wear.[7] People wear clothes for a variety of reasons—for the strictly utilitarian reason to keep warm, of course. But they also wear them for fashion, for modesty, for comfort, and because they are the clothes that they just happen to have. Analogously, people hold some beliefs because they are effective for guiding actions. But they also hold beliefs that make them feel good or that their peers believe. It's as if people were all actors in their own semi-private play of life. Their beliefs are like the costumes they wear for the play. 'The play's the thing'."

[6] From the lyrics of "Munchhausen" by Friedrich Hollaender.
[English lyrics by Jeremy Lawrence from a translation by Alan Lareau; from the printed matter of the London CD "Ute Lemper, Berlin Cabaret Songs" (452 849-2).]

[7] Robin Hanson, private communication, 1999.

"Sometimes the very act of believing something can *make* it true—at least for humans," Mia said. "There's a saying that goes 'Whether you believe you can or believe you cannot, you're probably right.' Self-confidence, even when unjustified, can be a powerful force for success."

"Like in the book, *The Power of Positive Thinking*, by Norman Vincent Peale?"[8] guessed Gio.

"Yes," said Nick, "and conversely a fatalistic, *qué-será-será* attitude inhibits the 'I-can-do-it' beliefs that encourage actions that might improve on 'whatever-will-be'."

"You guys seem to have a lot of beliefs about beliefs," Sam said.

"Yes," Nick replied, "beliefs about beliefs are what we are talking about."

"When do we start talking about evaluating beliefs then?" Sam asked.

"We'll get to that," Nick replied, "but how about quitting for today and continuing tomorrow?"

"We have a few other topics to discuss before we talk about how to evaluate beliefs," Mia added.

"Like what?" asked Sam.

"First, we have to tell you about *models*, Sam," Mia answered. "For example, Gio's beliefs are a kind of model of his world. He has other models too. We call his beliefs a model, because they help him predict things."

"That's what models are for," Nick added. "They are used to make predictions."

"Do you think of human beliefs as models too?" Sam asked.

"Exactly," Nick said, "And, like Gio, we have other kinds of models also. Ours take many forms—mental pictures, stories, analogies and metaphors, poems, myths, maps, architectural and engineering drawings, theories, equations, articles, books, and even computer programs. People make models of everything—of the everyday world, of the earth, the stars, and galaxies, of the lives and characteristics of other people, of plants and animals, of robots, of all the things that were, are, and might be. Besides

[8]Norman Vincent Peale, *The Power Of Positive Thinking*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1952.

using models for predictions, we use them for many other purposes also. They help us explain and understand things, and they can entertain and even persuade."

"Ok, let's hear about models then," Sam agreed.