Chapter 1

What is thinking?

Beginning to reason is like stepping onto an escalator that leads upward and out of sight. Once we take the first step, the distance to be traveled is independent of our will and we cannot know in advance where we shall end.

Peter Singer (1982)

Thinking is important to all of us in our daily lives. The way we think affects the way we plan our lives, the personal goals we choose, and the decisions we make. Good thinking is therefore not something that is forced upon us in school: It is something that we all want to do, and want others to do, to achieve our goals and theirs.

This approach gives a special meaning to the term “rational.” Rational does not mean, here, a kind of thinking that denies emotions and desires: It means, the kind of thinking we would all want to do, if we were aware of our own best interests, in order to achieve our goals. People want to think “rationally,” in this sense. It does not make much sense to say that you do not want to do something that will help you achieve your goals: Your goals are, by definition, what you want to achieve. They are the criteria by which you evaluate everything about your life.

The main theme of this book is the comparison of what people do with what they should do, that is, with what it would be rational for them to do. By finding out where the differences are, we can help people — including ourselves — to think more rationally, in ways that help us achieve our own goals more effectively.

This chapter discusses three basic types of thinking that we have to do in order to achieve our goals: thinking about decisions, thinking about beliefs, and thinking about our goals themselves. It also describes what I call the search-inference framework, a way of identifying the basic elements in all of these thinking processes.
Types of thinking

We think when we are in doubt about how to act, what to believe, or what to desire. In these situations, thinking helps us to resolve our doubts: It is purposive. We have to think when we make decisions, when we form beliefs, and when we choose our personal goals, and we will be better off later if we think well in these situations.

A decision is a choice of action — of what to do or not do. Decisions are made to achieve goals, and they are based on beliefs about what actions will achieve the goals. For example, if I believe it is going to rain, and if my goal is to keep dry, I will carry an umbrella. Decisions may attempt to satisfy the goals of others as well as the selfish goals of the decision maker. I may carry an extra umbrella for a friend. Decisions may concern small matters, such as whether to carry an umbrella, or matters of enormous importance, such as how one government should respond to a provocation by another. Decisions may be simple, involving only a single goal, two options, and strong beliefs about which option will best achieve the goal, or they may be complex, with many goals and options and with uncertain beliefs.

Decisions depend on beliefs and goals, but we can think about beliefs and goals separately, without even knowing what decisions they will affect. When we think about belief, we think to decide how strongly to believe something, or which of several competing beliefs is true. When we believe a proposition, we tend to act as if it were true. If I believe it will rain, I will carry my umbrella. We may express beliefs in language, even without acting on them ourselves. (Others may act on the beliefs we express.) Many school problems, such as those in mathematics, involve thinking about beliefs that we express in language only, not in actions. Beliefs may vary in strength, and they may be quantified as probabilities. A decision to go out of my way to buy an umbrella requires a stronger belief that it will rain (a higher probability) than a decision to carry an umbrella I already own.

When we decide on a personal goal, we make a decision that affects future decisions. If a person decides to pursue a certain career, the pursuit of that career becomes a goal that many future decisions will seek to achieve. When we choose personal goals by thinking, we also try to bind our future behavior. Personal goals of this sort require self-control.

Actions, beliefs, and personal goals can be the results of thinking, but they can also come about in other ways. For example, we are born with the personal goal of satisfying physical needs. It may also make sense to say that we are born holding the belief that space has three dimensions. The action of laughing at a joke does not result from a decision. If it did, it would not be a real laugh.

The search-inference framework

Thinking about actions, beliefs, and personal goals can all be described in terms of a common framework, which asserts that thinking consists of search and inference. We search for certain objects and then we make inferences from and about them.
Let us take a simple example of a decision. Suppose you are a college student trying to decide which courses you will take next term. Most of the courses you have scheduled are required for your major, but you have room for one elective. The question that starts your thinking is simply this: Which course should I take?

You begin by saying to a friend, “I have a free course. Any ideas?” She says that she enjoyed Professor Smith’s course in Soviet-American relations. You think that the subject sounds interesting, and you want to know more about modern history. You ask her about the work, and she says that there is a lot of reading and a twenty-page paper. You think about all the computer-science assignments you are going to have this term, and, realizing that you were hoping for an easier course, you resolve to look elsewhere. You then recall hearing about a course in American history since World War II. That has the same advantages as the first course — it sounds interesting and it is about modern history — but you think the work might not be so hard. You try to find someone who has taken the course.

Clearly, we could go on with this example, but it already shows the main characteristics of thinking. It begins with doubt. It involves a search directed at removing the doubt. Thinking is, in a way, like exploration. In the course of the search, you discovered two possible courses, some good features of both courses, some bad features of one course, and some goals you are trying to achieve. You also made an inference: You rejected the first course because the work was too hard.

We search for three kinds of objects: possibilities, evidence, and goals.

*Possibilities* are possible answers to the original question, possible resolutions of the original doubt. (In the example, they are possible courses.) Notice that possibilities can come from inside yourself or from outside. (This is also true of evidence and goals.) The first possibility in this example came from outside: It was suggested by someone else. The second came from inside: It came from your memory.

*Goals* are the criteria by which you evaluate the possibilities. Three goals have been mentioned in our example: your desire for an interesting course; your feeling that you ought to know something about recent history; and your desire to keep your work load manageable. Some goals are usually present at the time when thinking begins. In this case, only the goal of finding a course is present, and it is an insufficient goal, because it does not help you to distinguish among the possibilities, the various courses you could take. Additional goals must be sought.

I use the term “goal” throughout this book, but it is not entirely satisfactory. It evokes images of games like soccer and basketball, in which each team tries to get the ball into the “goal.” Such goals are all-or-none. You either get one or you don’t. Some of the goals I discuss here are of that type, but others are more like the rating scales used for scoring divers or gymnasts. This is, in a way, closer to the fundamental meaning, which is that the goals are criteria or standards of evaluation. Other words for the same idea are criteria, objectives, and values (in the sense of evaluation, not the more limited sense referring to morality). Because all these terms are misleading in different ways, I will stick with goals. At least this term conveys the sense that, for most of us, goals have motivational force. We try to achieve them. But we also apply them when we make judgments.
Evidence consists of any belief or potential belief that helps you determine the extent to which a possibility achieves some goal. In this case, the evidence consists of your friend’s report that the course was interesting and her report that the work load was heavy. The example ended with your resolution to search for more evidence about the work load of the second possibility, the American history course. Such a search for evidence might initiate a whole other episode of thinking, the goal of which would be to determine where that evidence can be found.

In addition to these search processes, there is a process of inference, in which each possibility is strengthened or weakened as a choice on the basis of the evidence, in light of the goals. Goals determine the way in which evidence is used. For example, the evidence about work load would be irrelevant if having a manageable work load were not a goal. The importance of that goal, which seems to be high, affects the importance of that evidence, which seems to be great.

The objects of thinking are represented in our minds. We are conscious of them. If they are not in our immediate consciousness, we can recall them when they are relevant, even after an episode of thinking resumes following an interruption. The processes of thinking — the search for possibilities, evidence, and goals and the inference from the evidence to evaluate the possibilities — do not occur in any fixed order. They overlap. The thinker alternates from one to another.

Why just these phases: the search for possibilities, evidence, and goals, and inference? Thinking is, in its most general sense, a method of finding and choosing among potential possibilities, that is, possible actions, beliefs, or personal goals. For any choice, there must be purposes or goals, and goals can be added to or removed from the list. I can search for (or be open to) new goals; therefore, search for goals is always possible. There must also be objects that can be brought to bear on the choice among possibilities. Hence, there must be evidence, and it can always be sought. Finally, the evidence must be used, or it might as well not have been gathered. These phases are “necessary” in this sense.

The term judgment will be important in this book. By judgment, I mean the evaluation of one or more possibilities with respect to a specific set of evidence and goals. In decision making, we can judge whether to take an option or not, or we can judge its desirability relative to other options. In belief formation, we can judge whether to accept a belief as a basis of action, or we can judge the probability that the belief is true. In thinking about personal goals, we can judge whether or not to adopt a goal, or we can judge how strong it should be relative to other goals. The term “judgment,” therefore, refers to the process of inference.

Let us review the main elements of thinking, using another example of decision making, the practical matter of looking for an apartment. “Possibilities” are possible answers to the question that inspired the thinking: Here, they are possible apartments. Possibilities (like goals and evidence) can be in mind before thinking begins. You may already have seen one apartment you like before you even think about moving. Or possibilities can be added, as a result of active search (through the newspaper) or suggestions from outside (tips from friends).
**Goals** are criteria used for evaluating possibilities. In the apartment-hunting example, goals include factors such as rent, distance from work or school, safety, and design quality. The goals determine what evidence is sought and how it is used. It is not until you think that safety might be relevant that you begin to inquire about building security or the safety of the neighborhood. When we search for goals, we ask, “What should I be trying to do?” or “What are my purposes in doing this?” Can you think of other criteria for apartments aside from those listed? In doing so, you are searching for goals. We also often have a subgoal, a goal whose achievement will help us achieve some other goal. In this example, “good locks” would be a subgoal for “safety.” Each possibility has what I shall call its strength, which represents the extent to which it is judged by the thinker to satisfy the goals. In decision making, the strength of a possibility corresponds to its overall desirability as an act, taking into account all the goals that the decision maker has in mind.

Evidence is sought — or makes itself available. Evidence can consist of simple propositions such as “The rent is $300 a month,” or it can consist of arguments, imagined scenarios, or examples. One possibility can serve as evidence against another, as when we challenge a scientific hypothesis by giving an alternative and incompatible explanation of the data. Briggs and Krantz (1992) found that subjects can judge the weight of each piece of evidence independently of other pieces.

Each piece of evidence has what I shall call a weight with respect to a given possibility and set of goals. The weight of a given piece of evidence determines how much it should strengthen or weaken the possibility as a means of achieving the goals. The weight of the evidence by itself does not determine how much the strength of a possibility is revised as the possibility is evaluated: the thinker controls this revision. Therefore a thinker can err by revising the strength of a possibility too much or too little.

The use of the evidence to revise (or not revise) strengths of possibilities is the end result of all of these search processes. This phase is also called inference. It is apparent that inference is not all of thinking, although it is a crucial part.

The relationship among the elements of thinking is illustrated in the following diagram:

![Diagram](https://via.placeholder.com/150)

The evidence \(E\) affects the strengths of the possibilities \(P\), but the weight of the evidence is affected by the goals \(G\). Different goals can even reverse the weight
of a piece of evidence. For example, if I want to buy a car and am trying to decide between two different ones (possibilities), and one of the cars is big and heavy (evidence), my concern with safety (a goal) might make the size a virtue (positive weight), but my concern with mileage (another goal) might make the size a detriment (negative weight).

The following story describes the situation of a person who has to make an important decision. As you read it, try to discover the goals, possibilities, evidence, and inferences:

A corporate executive is caught in a dilemma. Her colleagues in the Eastern District Sales Department of the National Widget Corporation have decided to increase the amount they are permitted to charge to their expense accounts without informing the central office (which is unlikely to notice). When she hears about the idea, at first she wants to go along, imagining the nice restaurants to which she could take her clients, but then she has an uneasy feeling about whether it is right to do this. She thinks that not telling the central office is a little like lying.

When she voices her doubts to her colleagues, they point out that other departments in the corporation are allowed higher expense accounts than theirs and that increased entertainment and travel opportunities will benefit the corporation in various ways. Nearly persuaded to go along at this point, she still has doubts. She thinks of the argument that any other department could do the same, cooking up other flimsy excuses, and that if all departments did so, the corporation would suffer considerably. (She makes use here of a type of moral argument that she recognizes as one she has used before, namely, “What if everyone did that?”) She also wonders why, if the idea is really so harmless, her colleagues are not willing to tell the central office.

Now in a real quandary, because her colleagues had determined to go ahead, she wonders what she can do on her own. She considers reporting the decision to the central office, but she imagines what would happen then. Her colleagues might all get fired, but if not, they would surely do their best to make her life miserable. And does she really want them all fired? Ten years with the company have given her some feelings of personal attachment to her co-workers, as well as loyalty to the company. But she cannot go along with the plan herself either, for she thinks it is wrong, and, besides, if the central office does catch them, they could all get fired. (She recalls a rumor that this happened once before.) She finally decides not to go above the company’s stated limit for her department’s expense accounts herself and to keep careful records of her own actual use of her own expense account, so that she can prove her innocence if the need arises.
In this case, the goals were entertaining clients in style; following moral rules; serving the interests of the corporation; being loyal to colleagues; and avoiding punishment. The possibilities were going along, turning everyone in, not going along, and not going along plus keeping records. The evidence consisted of feelings and arguments — sometimes arguments of others, sometimes arguments that our executive thought of herself.

Initially the executive saw only a single possibility — to go along — but some evidence against that possibility presented itself, specifically, an intuition or uneasy feeling. Such intuitions are usually a sign that more evidence will be found. Here, the executive realized that withholding evidence was a form of lying, so a moral rule was being violated. With this piece of evidence came a new goal that was not initially present in the executive’s mind, the goal of being moral or doing the right thing. She sought more evidence by talking to her colleagues, and she thought of more evidence after she heard their arguments. Finally, another possibility was considered: turning everyone in. Evidence against this possibility also involved the discovery of other relevant goals — in particular, loyalty to colleagues and self-protection.

The final possibility was a compromise, serving no goals perfectly. It was not as “moral” as turning her colleagues in or trying to persuade them to stop. It might not have turned out to be as self-protective either, if the whole plot had been discovered, and it was not as loyal to colleagues as going along. This kind of result is typical of many difficult decisions.

This example clarifies the distinction between personal goals and goals for thinking. The goals for thinking were drawn from our executive’s personal goals. She had adopted these personal goals sometime in the past. When she searched for goals for her thinking, she searched among her own personal goals. Many of her personal goals were not found in her search for goals, in most cases because they were irrelevant to the decision. Each person has a large set of personal goals, only a few of which become goals for thinking in any particular decision.

The examples presented so far are all readily recognizable as decisions, yet there are other types of thinking — not usually considered to be decision making — that can be analyzed as decision making when they are examined closely. For instance, any sort of inventive or creative thinking can be analyzed this way. When we create music, poetry, paintings, stories, designs for buildings, scientific theories, essays, or computer programs, we make decisions at several levels. We decide on the overall plan of the work, the main parts of the plan, and the details. Often, thinking at these different levels goes on simultaneously. We sometimes revise the overall plan when problems with details come up. At each level, we consider possibilities for that level, we search for goals, and we look for evidence about how well the possibilities achieve the goals.

Planning is decision making, except that it does not result in immediate action. Some plans — such as plans for a Saturday evening — are simply decisions about specific actions to be carried out at a later time. Other, long-term plans produce personal goals, which then become the goals for later episodes of thinking. For example, a personal career goal will affect decisions about education. Thinking about
plans may extend over the period during which the plans are in effect. We may revise our plans on the basis of experience. Experience provides new evidence. The goals involved in planning — the criteria by which we evaluate possible plans — are the personal goals we already have. We therefore create new goals on the basis of old ones. We may also decide to give up (or temporarily put aside) some personal goals.

We may have short-term plans as well as long-term plans. When we are trying to solve a math problem, we often make a plan about how to proceed, which we may revise as we work on the problem.

Thinking about beliefs

The search-inference framework applies to thinking about beliefs as well as thinking about decisions. When we think about beliefs, we make decisions to strengthen or weaken possible beliefs. One goal is to bring our beliefs into line with the evidence. (Sometimes we have other goals as well — for example, the goal of believing certain things, regardless of their fit with the evidence.) Roughly, beliefs that are most in line with the evidence are beliefs that correspond best with the world as it is. They are beliefs that are most likely to be true. If a belief is true, and if we hold it because we have found the right evidence and made the right inferences, we can be said to know something. Hence, thinking about beliefs can lead to knowledge.

Examination of a few types of thinking about belief will show how the search-inference framework applies. (Each of these types is described in more detail in later chapters.)

Diagnosis. In diagnosis, the goal is to discover what the trouble is — what is wrong with a patient, an automobile engine, a leaky toilet, or a piece of writing. The search for evidence is only partially under the thinker’s control, both because some of the evidence is provided without being requested and because there is some limitation on the kinds of requests that can be obeyed. In particular, the import of the evidence cannot usually be specified as part of the request (for example, a physician cannot say, “Give me any evidence supporting a diagnosis of ulcers,” unless the patient knows what this evidence would be). In the purest form of diagnosis, the goal is essentially never changed, although there may be subepisodes of thinking directed toward subgoals, such as obtaining a certain kind of evidence.

Scientific thinking. A great deal of science involves testing hypotheses about the nature of some phenomenon. What is the cause of a certain disease? What causes the tides? The “possibilities” are the hypotheses that the scientist considers: germs, a poison, the sun, the moon. Evidence consists of experiments and observations. Pasteur, for example, inferred that several diseases were caused by bacteria, after finding that boiling contaminated liquid prevented the spread of disease — an experiment. He also observed bacteria under a microscope — an observation.

1For a more complete introduction to these concepts, see Scheffler, 1965. We shall also return to them throughout this book.
Science differs from diagnosis in that the search for goals is largely under the thinker’s control and the goals are frequently changed. Scientists frequently “discover” the “real question” they were trying to answer in the course of trying to answer some other question. There is, in experimental science, the same limitation on control over the evidence-search phase: The scientist cannot pose a question of the form “Give me a result that supports my hypothesis.” This limitation does not apply when evidence is sought from books or from one’s own memory.

Reflection. Reflection includes the essential work of philosophers, linguists, mathematicians, and others who try to arrive at general principles or rules on the basis of evidence gathered largely from their own memories rather than from the outside world. Do all words ending in “-ation” have the main stress on the syllable “a”? Does immoral action always involve a kind of thoughtlessness? In reflection, the search for evidence is more under the control of the thinker than in diagnosis and experimental science; in particular, thinkers can direct their memories to provide evidence either for or against a given possibility (in this case, a generalization). One can try to think of words ending in “-ation” that follow the proposed rule or words that violate it. One can try to recall, or imagine, immoral actions that do or do not involve thoughtlessness. In reflection (and elsewhere), new possibilities may be modifications of old ones. For example, after thinking of evidence, a philosopher might revise the rule about immorality: “All immorality involves thoughtlessness, except ______.” Reflection lies at the heart of scholarship, not just in philosophy but also in the social sciences and humanities.

Insight problems. Much of the psychology of thinking concerns thinking of a very limited sort, the solution of puzzle problems. For example, why is any number of the form ABC,ABC (such as 143,143 or 856,856) divisible by 13? These are problems whose solution usually comes suddenly and with some certainty, after a period of apparently futile effort. Many are used on intelligence tests. Essentially, the only phase under the thinker’s control at all is the search for possibilities. Often, it is difficult to come up with any possibilities at all (as in the 13 problem). In other cases, such as crossword puzzles, possibilities present themselves readily and are rejected even more readily. In either case, search for evidence and inference (acceptance or rejection) are essentially immediate, and the goal is fixed by the problem statement. It is this immediate, effortless occurrence of the other phases that gives insight problems their unique quality of sudden realization of the solution.

Prediction. Who will be the next president of the United States? Will the stock market go up or down? Will student X succeed if we admit her to graduate school? Prediction of likely future events is like reflection, in form, although the goal is fixed. The evidence often consists of memories of other situations the thinker knows about, which are used as the basis of analogies — for example, student Y, who did succeed, and who was a lot like X.

Behavioral learning. In every realm of our lives — in our social relationships with friends, families, colleagues, and strangers, and in our work — we learn how...
our behavior affects ourselves and others. Such learning may occur without thinking, but thinking can also be brought to bear. When it is, each action is a search for evidence, an experiment designed to find out what will happen. The evidence is the outcome of this experiment. Each possibility we consider is a type of action to take.

This kind of learning can have much in common with science. Whereas science is a “pure” activity, with a single goal, behavioral learning has two goals: learning about the situation and obtaining immediate success or reward in the task at hand. These goals frequently compete (Schwartz, 1982). We are often faced with a choice of repeating some action that has served us reasonably well in the past or taking some new action, hoping either that it might yield an even better outcome or that we can obtain evidence that will help us decide what to do in the future. Some people choose the former course too often and, as a result, achieve adaptations less satisfactory to them than they might achieve if they experimented more.

An example of behavioral learning with enormous importance for education is the learning of ways of proceeding in thinking tasks themselves — for example, the important strategy of looking for reasons why you might be wrong before concluding that you are right. The effectiveness of thinking may depend largely on the number and quality of these thinking strategies. This, in turn, may (or may not) depend on the quality of the thinking that went into the learning of these heuristics.

The results of behavioral learning are beliefs about what works best at achieving what goal in what situation. Such beliefs serve as evidence for the making of plans, which, in turn, provide personal goals for later decisions. For example, people who learn that they are admired for a particular skill, such as telling jokes, can form the goal of developing that skill and seeking opportunities to display it.

Learning from observation. This includes all cases in which we learn about our environment from observation alone, without intentional experimentation. As such, it can include behavioral learning without experimentation — namely, learning in which we simply observe that certain actions (done for reasons other than to get evidence) are followed by certain events. It also includes a large part of the learning of syntax, word meanings, and other culturally transmitted bodies of knowledge.

The distinctive property of learning by observation is that the evidence is not under the thinker’s control, except for the choice of whether we attend to it or not. By contrast, Horton (1967, pp. 172–173) has suggested that one of the fundamental properties of scientific thinking is active experimentation: “The essence of the experiment is that the holder of a pet theory does not just wait for events to come along and show whether or not [the theory] has a good predictive performance. He bombards it with artificially produced events in such a way that its merits or defects will show up as immediately and as clearly as possible.”

How do search processes work?

All of these types of thinking involve search. Search for possibilities is nearly always present, and search for evidence or goals is often included as well. The critical aspect
of a search process is that the thinker has the goal of finding some sort of mental representation of a possibility, a piece of evidence, or a goal.

Search is directed by the goals, possibilities, and evidence already at hand. Goals provide the most essential direction. If my goal is to protect the walls in my house from my child’s scribbling, I seek different possibilities than if my goal is to teach my child not to scribble on walls. Possibilities direct our search for evidence for them or against them, and evidence against one possibility might direct our search for new ones.

There are two general ways of finding any object: recall from our own memory, and the use of external aids, such as other people, written sources (including our own notes), and computers. External aids can help us overcome the limitations of our own memories, including the time and effort required to get information into them. As I write this book, for example, I rely extensively on a file cabinet full of reprints of articles, my own library, the University of Pennsylvania library, and my colleagues and students. I rely on my memory as well, including my memory of how to use these tools and of who is likely to be able to help with what.

Thinking is not limited to what we do in our heads. The analogy between thinking and exploration is therefore not just an analogy. When an explorer climbs up a hill to see what lies beyond, he is actually seeking evidence. Moreover, libraries, computers, and file cabinets make us truly more effective thinkers. When we try to test people’s thinking by removing them from their natural environment, which may include their tools and other people they depend on, we get a distorted picture (however useful this picture may be for some purposes).

Because thinking involves search, there must be something for the search to find, if thinking is to succeed. Without knowledge, or beliefs that correspond to reality, thinking is an empty shell. This does not mean, however, that thinking cannot occur until one is an expert. One way to become an expert is to think about certain kinds of problems from the outset. Thinking helps us to learn, especially when our thinking leads us to consult outside sources or experts. As we learn more, our thinking becomes more effective. If you try to figure out what is wrong with your car (or your computer, or your body) every time something goes wrong with it, you will find yourself looking up things in books and asking experts (repair people, physicians) as part of your search for possibilities and evidence. You will then come to know more and to participate more fully in thinking about similar problems in the future. It is often thought that there is a conflict between “learning to think” and “acquiring knowledge”; in fact, these two kinds of learning normally reinforce each other.

What we recall (or get from an external aid) may be either an item itself or a rule for producing what we seek. For example, the “What if everybody did that?” rule is not by itself evidence for or against any particular action, but it tells us how to obtain such evidence. When we solve a problem in physics, we recall formulas that tell us how to calculate the quantities we seek. Rules can be learned directly, or we can invent them ourselves through a thinking process of hypothesis testing or reflection. (The use of rules in thinking can be distinguished from the use of rules to guide behavior. We may follow a rule through habit without representing it consciously.)
Recall or external aids may not give us exactly what we want, but sometimes an item suggests something else more useful. We may transform what we get in a variety of ways to make it applicable to our situation. This is the important mechanism of analogy. To see the role of analogies in thinking, try thinking about a question such as “Can a goose quack?” or “How does evaporation work?” (Collins and Gentner, 1986; Collins and Michalski, 1989). To answer the first question, you might think of ducks. To answer the second, some people try to understand evaporation in terms of analogies to things they already know. The escape of a molecule of a liquid might be analogous to the escape of a rocket from the earth. The conclusion drawn from this analogy is that a certain speed is required to overcome whatever force holds the molecules in the liquid. Some people conclude that the force to be overcome is gravity itself. (In fact, gravity plays a role, but other forces are usually more important.)

Notice that analogies, as evidence for possibilities, need different amounts of modification depending on their similarity to the possibility in question. In the 1980s, an analogy with the U.S. military experience in Vietnam was used to argue against military intervention against communists in Nicaragua. Later the same analogy was used (unsuccessfully) to argue against military intervention in Somalia. The analogy was more distant in the latter case, because communists were no longer the enemy. The appeasement of Hitler at Munich has been used repeatedly to support all sorts of military interventions, some closely related, some not so close.

When an analogy requires modification, the person may need to think about how to make the necessary modification. For example, the lesson of Munich may be that fascists should not be appeased, or it could be that one’s enemies should not be appeased. Likewise, if you know how to find the area of a rectangle, how should you apply this knowledge to finding the area of a parallelogram? Do you multiply the base by the length of the sides next to it, or do you multiply the base by the height? For rectangles, both yield the same result. Evidence can be brought to bear about which of these possibilities serves the goal.

Standards for the use of analogies as evidence have changed over the centuries in Western science (Gentner and Jeziorski, 1993). Modern analogies — such as Rutherford’s analogy between the structure of the atom and the structure of the solar system — are based on common relations among elements of two domains: the sun (nucleus) is more massive than the planets (electrons) and attracts them, so they revolve around it. Relations between an element of one domain and an element of the other — such as the fact that the sun gives off electrons — are irrelevant to the goodness of the analogy. By contrast, alchemists made analogies with shifting bases, according to superficial appearance rather than relations among elements. Celestial bodies were matched with colors on the basis of appearance (the sun with gold; the moon with white) but also on the basis of other relations (Jupiter with blue because Jupiter was the god of the sky). For metals, the sun was matched with silver on the basis of color, but Saturn was matched with lead on the basis of speed (Saturn being the slowest known planet, lead being the heaviest, hence “slowest,” metal). Alchemists also thought of some analogies as decisive arguments, while modern
scientists think of them as suggestions for hypotheses to be tested in other ways, or as means of exposition.

Young children’s analogies are more like alchemists’ than like modern scientists’. When asked how a cloud is like a sponge, a preschool child answered, “Both are round and fluffy,” while older children and adults are more likely to point out that both hold water and give it back. This is one of many areas in which standards of reasoning may be acquired through schooling.

Knowledge, thinking, and understanding

Thinking leads to knowledge. This section reviews some ideas about knowledge from cognitive psychology. These ideas are important as background to what follows.

Naive theories

Naive theories are systems of beliefs that result from incomplete thinking. They are analogous to scientific theories. What makes them “naïve” is that they are now superceded by better theories. Many scientific theories today will turn out to be naïve in light of theories yet to be devised. Theories develop within individuals in ways that are analogous to their development in history.

For example, certain children seem to hold a view of astronomy much like that of some of the ancients (Vosniadou and Brewer, 1987). They say, if asked, that the earth is flat, the sun rises and passes through the sky, perhaps pushed by the wind, and so on. Unless the children have been specifically instructed otherwise, these are natural views to hold. They correspond to the way things appear, and this is the reason the ancients held them as well.

When the wonders of modern astronomy are first revealed to them, these children will at first modify their structure as little as possible, to accommodate the new information. For example, one child (according to an anecdote I heard) learned dutifully in school that the earth goes around the sun. When asked later where the earth was, he pointed upward, answering, “Up there, going around the sun.” This earth he had learned about could not be the same earth he already knew, which, after all, was obviously flat and stationary. Another child (described by Piaget, 1929, p. 236) had been taught about the cycle of night and day and the rotation of the earth. She had been told that when it was night in Europe (where she lived), it was day in America. Not wanting to give up her idea of a flat earth, she now reported, when asked, that there was a flat-earth America underneath the flat-earth Europe, and that at night the sun dropped below the European layer to shine on the American layer.

Vosniadou and Brewer (1987) point out that when the modern view is finally adopted, the change is truly radical. First, the concepts themselves are replaced with
Chapter 9

Actively open-minded thinking

The human understanding when it has once adopted an opinion draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusion may remain inviolate.

Francis Bacon

The search-inference framework implies that thinking can go wrong for three reasons.

1. Our search misses something that it should have discovered, or we act with high confidence after little search.

2. We seek evidence and make inferences in ways that prevent us from choosing the best possibility.

3. We think too much.

The second of these problems seems to be the most serious. People tend to seek evidence, seek goals, and make inferences in a way that favors possibilities that already appeal to them. For example, we often ignore evidence that goes against a possibility we like.

The same favoritism for a particular possibility may cause us to prematurely cut off our search for alternative possibilities or for reasons against the one we have in mind. This favoritism therefore leads to insufficient thinking or to overconfidence in hasty conclusions — the first reason for poor thinking. This problem is especially great when something is worth thinking about, such as the choice of our personal goals or moral beliefs.

Poor thinking, therefore, tends to be characterized by too little search, by overconfidence in hasty conclusions, and — most importantly — by biases in favor of
the possibilities that are favored initially. In contrast, good thinking consists of (1) search that is thorough in proportion to the importance of the question, (2) confidence that is appropriate to the amount and quality of thinking done, and (3) fairness to other possibilities than the one we initially favor.

These three principles are also the standards we apply when we criticize each other in academic settings. When I read a student’s paper, or a colleague’s, or sometimes even my own, the things I look for are omissions of relevant evidence, omissions of statements about goals or purposes, omissions of alternative possibilities, other answers to the question at issue, and unqualified assertions not supported with evidence. I also look for partiality to the thesis of the paper, partiality that may itself cause the omissions just mentioned. When students take these kinds of criticisms to heart and try to become more thorough and more impartial, they are becoming more intelligent thinkers. They are acquiring abilities — in the form of habits and values — that will increase their effectiveness just as surely as would an improvement in their memory or their mental speed.

Thinking that follows these principles can be called *actively open-minded thinking*. It is “open-minded” because it allows consideration of new possibilities, new goals, and evidence against possibilities that already seem strong. It is “active” because it does not just wait for these things but seeks them out. These are the features of what I regard as “good thinking.”

A good prescriptive model of thinking, I shall argue, is one that advises and helps people to become more actively open-minded. It counteracts the major biases in thinking, and it serves as a reminder of the normative theory. Advice of this type is an important part of a good prescriptive model. The rest of a prescriptive model consists of many detailed heuristics, or rules of thumb, some of which will be discussed in this book.

**Examples of actively open-minded thinking**

Here is an example of actively open-minded thinking. Students in a class on thinking and decision making were asked to think about the following problem and transcribe their thoughts. The problem concerned the “best” way to allocate, among the nations of the world, the mining rights to minerals on the ocean floor, a resource not yet developed. One student’s transcribed verbal protocol read as follows:

_Wealth must be divided among nations fairly. What does ‘fairly’ mean? Should allocation be based on the size of the country? Some nations are significantly larger than others. But some countries have more people per unit area. Should allocation be based on overall population size? It would be very difficult to get all nations concerned to agree their shares were fair. Wait, the United Nations has a certain number of representatives from each country. They would be the ideal group to handle this. Total wealth should be divided by overall number of representatives, then allocated according to number of representatives per country._
But some nations would be better able to use the mineral wealth. These would be nations with greater technology. Therefore, underdeveloped nations would be unable to benefit as well as nations that are financially more secure. That would be unfair. [The protocol continues for several more pages.]

This student (who earned an A+ in the course) searched for goals, considered several possibilities. After each possibility, she looked for counterevidence and then tried to modify or replace the possibility to meet the objection she had thought of.

By contrast, the following protocol, handed in by another student, shows no evidence of actively open-minded thinking (although it is possible that some such thinking went on beforehand):

I believe that the most logical way of allocating the mineral wealth beneath the ocean is to allocate the ocean floors by extending national borders outward along the ocean floors. In effect, this plan would treat the ocean floor in the same way as exposed land surfaces. The water above the floor should still remain international territory, except where it is already considered national property. . . . Establishing boundaries in this manner is fairly simple, but it will favor nations with long coastlines along large bodies of water, but is no less fair than the rules for establishing national air space. [This protocol also went on for a page or two.]

It must be remembered that this kind of prescriptive model advises moderation, a middle course. More thinking, as we have already noted, is not always better. The search for counterevidence and alternative possibilities reaches a point of diminishing returns — often quickly. We can compare active open-mindedness to a virtue such as thriftiness. Thriftiness, too, must be practiced in moderation. We think of thriftiness as a virtue because most people usually do not have enough of it, but some people are too thrifty, penny-pinchers unable to enjoy the fruits of their labor. Similarly, active open-mindedness is a virtue because most people do not have enough of it, but too much of it can lead to intellectual paralysis.

Here is another example of two thinkers, C and S, sixth-grade students in Alberta, Canada, who were asked to report their thoughts after reading each segment of the following story (Norris and Phillips, 1987):

1. The stillness of the morning air was broken. The men headed down the bay.
2. The net was hard to pull. The heavy sea and strong tide made it even more difficult for the girdie. The meshed catch encouraged us to try harder.
3. With four quintels aboard we were now ready to leave. The skipper saw mares’ tails in the north.
4. We tied up to the wharf. We hastily grabbed our prongs and set to work. The catch was left in the stage while we had breakfast.
5. The splitting was done by the skipper. The boys did the cutting and gutting.

6. Catching fish is filled with risk.

C’s response to segment 1 was, “The men are going shopping. [Why do you say that?] They’re going to buy clothes at The Bay [a Canadian department store]. [Any questions?] No.” Notice the lack of qualification, the high confidence in this statement. C continues the interpretation of a shopping trip through the story, forcing the subsequent details to fit; at the end, C concluded, “Like I said before, they went shopping and got some fish, met some horses and then went to a show.”

S, on the other hand, said, “I think they might be going sailing.” S states a possibility, but qualified, so that it can be easily revised with subsequent evidence. S’s response to segment 3 illustrates well the potential of actively open-minded thinking: “I wonder what quintels are? I think maybe it’s a sea term, a word that means perhaps the weight aboard. Yes, maybe it’s how much fish they had aboard. [So you think it was fish?] I think fish or maybe something they had found in the water but I think fish more because of the word ‘catch.’ [Why were they worried about the mares’ tails?] I’m not sure. Mares’ tails, let me see, mares are horses but horses are not going to be in the water. They are out in a boat on the bay so there’s nothing around except sky and water. The mares’ tails are in the north. Here farmers watch the north for bad weather, so maybe the fishermen do the same thing. Yeah, I think that’s it, it’s a cloud formation, which could mean strong winds and hail or something which I think could be dangerous if you were in a boat and a lot of weight aboard. [Any questions?] I am curious now about whether I’m right about the mares’ tails.”

S considers and rejects possibilities (horses), actively searches for evidence, finds the analogy with local farmers, but remains unsure, awaiting confirmation.

The last example suggests that active open-mindedness can be measured by asking people to think aloud and by subjectively scoring the result for various signs of active open-mindedness, such as considering reasons against a tentative view. Indeed, several measures of this sort have been developed, often on the basis of theoretical schemes very similar to the one I have just presented, although the terms employed are usually different. Ordinarily, some training is required in order to learn to score the results. After the training, the reliability of the scoring can be checked by looking at the correlation between the responses of different judges. One scoring technique is the measure of “integrative complexity” described in Chapter 2. Another scoring system, developed by Perkins, Bushey, and Faraday (1986), scores subjects’ written responses to questions such as whether a state should have a compulsory returnable-bottle law. Responses are scored for whether or not they consider both sides of the question.

One general problem of these sorts of measures is that certain issues, for some people, do not lend themselves to two-sided thinking. If most people today were asked to comment on the question “Is slavery wrong?” or “What do you think of the character of Adolph Hitler?” they would probably not have much to say on the “other side.” In other cases, people have already thought through some position even though others remain on the opposite side. A devout Christian is unlikely to produce
a two-sided discussion of the existence of God. Yet such people are not necessarily poor thinkers in general, or even about these issues.

Even the presence of two-sided discussions does not necessarily indicate actively open-minded thinking, however, when the topic is truly a difficult one. Most people initially consider both sides of an issue when making a decision that they find hard. True active open-mindedness shows itself only after some tentative commitment to one side has been made. Those who still seek out the other side before jumping to a conclusion are the real actively open-minded thinkers. In sum, tests that look at whether people spontaneously consider both sides of an issue are imperfect indicators of the propensity to think well. Perhaps better measures would assess the understanding of the need to consider the other side, in general, before reaching a conclusion. This can be done by asking people to evaluate examples of the thinking of others.

Myside bias and irrational belief persistence

An important part of actively open-minded thinking is fairness to possibilities, regardless of their initial strength. People tend to favor possibilities that are already strong, both in search (particularly search for evidence) and in making inferences from evidence at hand. People tend not to look for evidence against what they favor, and, when they find it anyway, they tend to ignore it. David Perkins has named these two characteristics “myside bias” (Perkins et al., 1986). The term refers to “my side” of the issue at hand rather than one side of an argument with another person. (Of course, it happens there, too.)

Irrational belief persistence may result from myside bias in the search for evidence and in inference. As a result of these biases, incorrect beliefs are slow to change, and they can even become stronger when they ought to become weaker. Hence, their persistence is irrational.

The irrational persistence of belief is one of the major sources of human folly, as many have noted (Bacon, 1620/1960; Janis and Mann, 1977; Kruglanski and Ajzen, 1983; Nisbett and Ross, 1980). We tend to hold to our beliefs without sufficient regard to the evidence against them or the lack of evidence in their favor.

Irrational belief persistence affects our lives in many forms in addition to its effect on the judgment of probability and correlation. Good students, for example (like good scholars), must remain open to counterevidence and criticism, willing to be persuaded of alternative views and willing to criticize their own efforts so as to improve them; poor students often seem to be rigid in defending their mistaken beliefs about what they are learning in school (Baron, Badgio, and Gaskins, 1986).

Certain forms of psychopathology, such as delusions, are essentially defined by irrational persistence of belief. A delusional patient is not just someone who believes (wrongly) that her sneezing and coughing mean that she is dying of an incurable disease; she is someone who continues to believe this even after five reputable physicians tell her that her symptoms are caused by a simple allergy to ragweed.
Depressives, Beck has argued, maintain their depression by ignoring evidence that could cheer them up and by taking seriously only the evidence consistent with their gloomy outlook (Beck, 1976). Beck’s cognitive therapy of depression tries to teach the patient to treat the relevant evidence more impartially.

Irrational belief persistence has also been implicated in faulty decision making by individuals and governments alike. In any war in which one side clearly loses, the loss is apparent before it occurs, but both the government and people of the losing side continue to believe that they can see victory just around the corner. Moral beliefs that underlie political controversies, such as the controversies concerning abortion, sex, or racial inequality, also seem particularly resistant to arguments or evidence.

Irrational belief persistence can cause serious difficulties in personal matters such as relationships and business ventures, too. In romance and in business, confidence is usually a good thing, because it inspires us to undertake difficult ventures, to initiate relationships, to pursue lofty goals, but when the evidence tells us that a particular endeavor — whether a love affair or a new company — is not getting off the ground after several attempts, we need, for our own sakes, to be responsive to the evidence, and sometimes we are not.

Irrational belief persistence involves two types of biases that we have frequently encountered earlier in this book:

1. The **overweighing** of evidence consistent with a favored belief or the **underweighing** of evidence against it: for example, a general’s attending to reports of enemy casualties and ignoring reports of their troop strength.

2. The **failure to search impartially** for evidence: For example, when supporters of U.S. intervention in Nicaragua searched for historical analogies, they were likely to discover the Munich agreement with Hitler, and opponents, when they searched for analogies, were likely to discover U.S. support of Batista in Cuba or the unsuccessful Bay of Pigs invasion.

Some belief persistence is not irrational. Often the evidence against a belief is not strong enough to make a convincing case for giving it up. Michael Faraday persisted in believing that electrical currents could be induced with magnets despite several failures to produce such currents experimentally and finally, of course, succeeded (Tweney, 1985). If we all gave up beliefs as soon as there was evidence against them, we would hold very few beliefs with any certainty, and we would give up many beliefs that were true.

The amount of belief persistence that is **rational** is whatever best serves the goals of the thinker. In most cases, the goal is to adopt beliefs that provide the best basis for decisions that achieve those goals. We cannot judge the rationality of persisting in a belief by knowing the truth in hindsight. Although rationally formed beliefs are, on the whole, more likely to be true than irrationally formed beliefs, we cannot assume that false beliefs are always irrationally formed. For example, many efforts have been made to show that it was irrational of President Roosevelt and his advisers not to suspect that the Japanese would attack Pearl Harbor. Such critics claim that
the president and his policy makers possessed much evidence that favored such an attack but explained it away or ignored it. Some evidence must always be “explained away” or considered less important, however, even when beliefs are true. Evidence is not always consistent. The question is, how much explaining away is too much? For every failure to predict something that happened, Jervis asserts that, “we could look at an event that did not take place, and, acting as though it had occurred, find a large number of clues indicating the event’s probable occurrence which the officials had ignored or explained away” (1976, p. 175).

If we want to find out if, when, and why irrational persistence of belief occurs, we need some normative standards for belief persistence itself. We cannot typically use probability theory as a normative model. Some of the beliefs include strongly held personal commitments that shape a person’s moral, social, and political attitudes. Many of these things are really goals (in the sense of ch. 1). A person who says that she believes strongly that abortion is wrong (or not wrong) may be willing to spend some time every day campaigning for her views. It is worthwhile to bear in mind, though, that people think about their goals in much the same way in which they think about their beliefs. For example, we think about goals by asking whether they are consistent with other goals, and we think about beliefs by asking whether they are consistent with other beliefs. Probabilities apply to propositions that are either true or false, but goals are not propositions of this sort, and the beliefs of interest here are difficult to study in this way.

The order principle, the primacy effect, and total discrediting

How can we tell whether people are weighing evidence in a way that is normatively correct? We have discussed the difficulties in using probability theory as our normative standard. A good alternative might be to focus on the process of weighing evidence itself and to look for general normative principles that a rational response to evidence would have to obey. Once we find such constraints, we can ask whether people violate them.

One such principle is the order principle: When the order in which we encounter two pieces of evidence is not itself informative, the order of the two pieces of evidence should have no effect on our final strength of belief. Or, put more crudely, “When the order doesn’t matter, the order shouldn’t matter.” Suppose I want to find out which of two political candidates is better, and I start out with no opinion at all. I run into a trusted friend, who tells me that Candidate X is better. Later, another equally trusted friend tells me that Candidate Y is better, but by this time I have already formed an inclination to vote for X, so I question the second friend very carefully. It was simply chance that made me run into X first, so the order of the evidence does not matter, yet I might end up favoring X because that evidence came first.

Such an effect is called a primacy effect, because the first piece of evidence is weighed more heavily than it should be. One explanation of the primacy effect is that the initial evidence leads to an opinion, which then biases the search for subsequent evidence, as well as the interpretation of that evidence when it is found. This would
be normatively irrational. If this kind of response to evidence can easily be avoided, we can say that it is prescriptively irrational as well.

Many psychological studies have reported primacy effects in judgment (see, for example, Anderson, 1981, ch. 3; Hovland, 1957), as well as recency effects, in which the judgment is more affected by evidence presented later. Recency effects might be taken to indicate the opposite bias of oversensitivity to evidence (relative to prior belief). Many of these studies have used a method devised by Asch (1946) in which the subjects were given a list of adjectives describing a person, such as “intelligent, industrious, impulsive, critical, stubborn, envious” or “envious, stubborn, critical, impulsive, industrious, intelligent.” The impressions of the person were more favorable given the first list than the second, even though one list is simply the reverse of the other. The term “intelligent,” in the first list, seemed (to Asch) to color the interpretation of the later terms; of course, “envious” would have the opposite effect.

Although Asch and others found primacy effects in this type of experiment, others, such as Stewart (1965), found recency effects. Recency effects do not necessarily imply that subjects were too open-minded. Another explanation is that subjects did not follow the instructions to consider all of the evidence equally. Stewart had to exclude the data from many subjects because they responded only to the most recent item rather than to all the items. Even subjects who tried to respond to all the items could have had difficulty suppressing the tendency to respond to the most recent item alone. This tendency can make primacy effects difficult to discover.

Most studies that show primacy effects are not relevant to the question of rationality in belief persistence. There are two general reasons why not. First, often a great deal of evidence is presented, and later (or earlier) evidence is differentially forgotten at the time a judgment is made. Although violation of the order principle would be normatively irrational for someone with perfect memory, it would not be irrational when we take memory into account. We must do that, for our prescriptive perspective, because a prescriptive guideline must reflect what is humanly possible. Second, the subjects might have had reason to believe that earlier (or later) evidence was actually more informative. In natural situations, the most important evidence is presented first (or last, depending on the situation), and it would take at least a little effort to convince subjects that this was not true in the laboratory as well.

Peterson and DuCharme (1967) designed an experiment that minimizes both problems. Subjects were told that a sequence of poker chips would be drawn at random from one of two urns, urn C or urn D. The subject’s task was to estimate the probability that the chips were drawn from urn C (or from urn D). Urn C contained three red, two blue, two yellow, one green, and two white chips. Urn D contained two red, three blue, one yellow, two green, and two white chips. After each chip was drawn, subjects provided a probability judgment about which urn was more likely to have held the chip. The most recent chips are surely freshest in the subject’s mind, so a primacy effect cannot be explained in terms of forgetting the most recent evidence. The experimenter arranged in advance the exact order in which the chips would be drawn. The color frequencies of the first thirty chips favored urn C. The next thirty were the mirror image of the first thirty, with red and blue switched, and yellow and
green switched. At the end of sixty trials, then, the subject should have thought the two urns to be equally likely. In fact, it took about twenty more chips favoring urn D before subjects reached this point. They seem to have become committed to their initial favoritism toward urn C. Because they still favored urn C after sixty trials, when the evidence was equal for both urns, they showed a primacy effect.

There is some reason to think that irrational primacy effects, and irrational persistence in general, are found only when subjects make some commitment to the belief suggested by the earliest evidence they receive. If they simply note the evidence and its implications, without forming a desire that its implications be true, they may remain open-minded until all of the evidence is in; they will effectively be using all of the evidence simultaneously, and irrational persistence will not occur. For example, Dailey (1952) found that subjects were less sensitive to new evidence when they made inferences after receiving early evidence in an impression-formation task (with information presented in paragraphs rather than as adjectives) than when they did not need to make any inferences.

Another experiment that suggests a similar phenomenon is that of Bruner and Potter (1964). Each subject was shown a series of pictures on slides. Each slide was presented several times. The first presentation was very much out of focus, but focus became gradually clearer over the series. After each presentation, subjects were asked to state their best hypothesis about what the picture represented. Subjects began stating hypotheses early in the series. They tended to stick to these hypotheses even after the slide came into focus. Their early hypotheses seemed to inhibit their recognition of what was actually on the slide. This was demonstrated by giving the same slides to other subjects but beginning only slightly out of focus. These subjects had little difficulty recognizing the pictures, even though the original subjects, given this degree of focus, could not recognize the pictures — presumably because they were committed to their original hypotheses, and these hypotheses guided their search for evidence and their inferences from the slide.

A different type of experimental design, called total discrediting, has been used to demonstrate irrational belief persistence (Anderson, Lepper, and Ross, 1980). The idea is to present some evidence that induces a particular belief, and then present contradictory evidence. For example, subjects are given questionnaire responses of two different firefighters, one rated as better than the other at the job. Some subjects are given evidence indicating that risk taking is positively associated with fire-fighting performance, other subjects are given the reverse. The evidence is then discredited by telling the subject that the evidence was totally fabricated, and the belief is then assessed. The direction of the initial belief manipulation continues to influence the belief, even after discrediting. Those who were initially led to believe that risk taking is associated with success continued to believe it. This is technically a violation of the order principle, but it is a special case, since the second piece of evidence is meaningless without the first.\footnote{In addition, Wegner, Coulton, and Wenzlaff (1985) point out that part of this effect is due simply to the difficulty of “wiping out” any information from one’s memory. Some measures of belief are affected by evidence even when the subject is warned in advance that the evidence is false.}
In another demonstration of the same type, high school students were given either a good or a poor film instructing them in how to solve a certain kind of problem, and they were then given four problems of this kind to solve (Lepper, Ross, and Lau, 1986). Their success or failure depended on the film they saw, but they tended to attribute it to their own fixed ability. This attribution persisted even after they were shown the film given to the other subjects. Merely being given an explanation of one’s success or failure in terms of the quality of one’s instruction did not remove the tendency to attribute it to one’s ability.

These effects of discredited information seem to be the result of the subject’s searching memory, at the time when the initial belief is created, for other evidence consistent with the belief. Later, despite the discrediting, the additional evidence remains and continues to affect belief. If there is “irrationality” here, it may be in the initial effort to bolster the belief by searching memory for evidence in favor of the belief but not for evidence against it.

Should we regard this effect as truly irrational, for the purpose of designing our prescriptive guidelines? At the time when the initial evidence is presented, I suggest, the bolstering may occur by automatic elicitation of associated memories rather than by an intentional search for supporting evidence. This effect might be very difficult to avoid. Later, at the time when the final belief is tested, the subjects may simply be unable to distinguish which evidence for their current belief was elicited by the original manipulation and which they would have thought of in any case. If they were to be more cautious, they might end up dismissing everything that they could think of as potentially biased. Prescriptively, then, this effect might not be irrational, because it might not be avoidable without unreasonable effort. This issue seems to have been largely settled by C. A. Anderson (1982), who found that asking subjects to use a heuristic designed to avoid bias — considering whether one could argue for the other side — substantially reduced the basic effect. Therefore, the original bias is one that can be overcome, if only we think in a way that is fairer to alternative possibilities.

In sum, beliefs created in an experiment seem to affect the search for and use of subsequent evidence in a way that maintains the beliefs. This is the basic mechanism that leads to violations of the order principle.

The neutral-evidence principle

The neutral-evidence principle might be stated like this: Neutral evidence should not strengthen belief. By “neutral evidence” I mean evidence that is, on the whole, equally consistent with a belief and its converse. Neutral evidence might consist of mixed evidence, that is, some evidence in favor of a belief and equal evidence against it.

The neutral-evidence principle would be violated if we tended to interpret ambiguous evidence as supporting a favored belief. For example, each side of an international conflict often believes that the other side is up to no good. An offer of concessions may be interpreted as a sign of weakness or trickery rather than as evidence
against the favored belief. Likewise, if the evidence is mixed, one side may attend only to the evidence that supports its initial belief, so that the belief is strengthened by this part of the evidence but not weakened by unfavorable evidence.

The neutral-evidence principle was clearly violated by subjects in experiments that Pitz and his colleagues carried out, using Bayesian probability theory as a normative standard (Pitz, 1969; Pitz, Downing, and Reinhold, 1967). Subjects observed a series of balls drawn from one of two “bingo baskets” — devices for the random selection of balls. The proportion of balls of different colors in the two baskets differed. For example, one basket had 60% red balls (and 40% black), and the other 40% red. The subjects knew that all the balls were drawn from one of the two baskets, and they knew the proportions, but they did not know which basket was used. After each draw, the ball drawn was returned to the basket, so that the proportion of red balls in the basket stayed the same for all draws. After each draw, the subject made a judgment of the probability that the balls were drawn from one of the baskets. When two successive balls were of different colors, the normative model (Bayes’s theorem) specifies that no overall change in probability of the subject’s hypothesis should occur, yet subjects usually increased the probability assigned to their more likely hypothesis after seeing two balls of different colors. If they thought the balls were from the first basket, for example, they counted a red ball as evidence in favor of this hypothesis, but they failed to count a black ball as equally strong evidence against it.

This “inertia effect” was present only when subjects were asked to make a judgment after each draw; when the judgment was delayed until the whole series was over, confidence was a function of the difference between the number of red and black balls drawn, just as Bayes’s theorem says it should be. (When an inertia effect was present, confidence increased with the number of draws, even when the difference was constant.) Pitz (1969) suggests that subjects need to commit themselves to a judgment in order to display resistance to evidence against that judgment.

Lord, Ross, and Lepper (1979) showed a violation of the neutral-evidence principle in a situation where such errors in daily life have deadly serious consequences. They selected subjects who had indicated that they favored, or opposed, capital punishment in responses to a questionnaire. Each subject was then presented with mixed evidence on the effectiveness of capital punishment in deterring crime. Each subject read two reports, one purporting to show effectiveness and the other purporting to show ineffectiveness. (Although the reports appeared to be actual journal articles, they had been fabricated by the experimenters.) One report compared murder rates in various states in the country before and after adoption of capital punishment. The other compared murder rates in states with and without capital punishment. The results were manipulated so that only the first report showed deterrence for half the subjects and only the second report showed deterrence for the other half.

The effect of each report on the subject’s belief was stronger when the report agreed with the belief than when it did not. The authors call this biased assimilation of evidence. Each subject rated the report that agreed with the subject’s opinion as “more convincing,” and found flaws more easily in the reports that went against the
subject’s belief. (Of course, neither kind of report is conclusive evidence, but both kinds are better than no evidence at all.) In the end, subjects polarized: that is, they became stronger in their initial beliefs, regardless of their direction. If anything, mixed evidence should have made subjects less sure of their beliefs.

Is biased assimilation nonnormative? Arguably, not by itself. For good reason (let us suppose), I do not believe in extra-sensory perception. When I am confronted with a study showing that someone in one room can guess, significantly above chance, what a person in another room is looking at, I usually do not bother to read it. I assume that it must have a flaw. If I do read it, I look for flaws and find them. When I see a study showing that behavior is random, I believe it and do not look for flaws. This behavior in itself may be normatively correct, especially if we take the cost of my time into account. I cannot spend time exploring the evidence for every implausible hypothesis. Yet, given that I behave this way, it is certainly nonnormative to take my filtered perception at face value. If I see one study on each side and dismiss one of them because of my prior belief, I have no good reason to strengthen my original belief. It would be as if I cooked my evidence to order and then forgot that I did that. As Lord, Ross and Lepper (1979, p. 2107) put it, “Our subjects’ main inferential shortcoming . . . did not lie in their inclination to process evidence in a biased manner . . . . Rather, [it] lay in their readiness to use evidence already processed in a biased manner to bolster the very theory or belief that initially ‘justified’ the processing bias.”

This polarization study is disturbing because it suggests that evidence is useless in settling controversial social questions. Of course the results may be limited to certain types of cases. People do not always have a chance to find flaws in evidence, and the result could be dependent on the greater effort to find flaws in arguments on the other side. Also, attitudes toward capital punishment may be as much a function of basic moral beliefs as of beliefs about its effectiveness as a deterrent. Opponents of capital punishment tend to feel that “two wrongs do not make a right” (even if the second wrong does prevent other wounds), and proponents tend to feel that “the punishment should fit the crime” (even if it does not prevent other crimes). Counterevidence is therefore easily resisted, in this case, by simply attending to the moral reasons for or against capital punishment, but such an attention shift is unnecessary when the evidence on deterrence is favorable to the choice consistent with one’s moral belief.

It is important to note that this “polarization effect” can be detected only when the bias against counterevidence is extreme. Normatively, we might expect that beliefs would move toward the middle of the range when people are presented with mixed evidence. If people have stronger (or more) evidence for the side they favor, then mixed evidence, which is equally strong on both sides, would add proportionately more strength to the other side. If we could apply a precise normative model to belief revision, it might specify some exact amount of movement toward neutrality. When beliefs do not move toward neutrality at all, they may move less than they should according to such a model. When we cannot apply such a model — as we cannot, in the case of the capital punishment experiment — we cannot detect such resistance.
to evidence unless it leads to polarization. However, the effect can be found in other cases aside from beliefs about capital punishment. For example, it is found with attitudes toward pertussis vaccination, which produces serious side effects as well as preventing potentially deadly disease (Meszaros, Asch, Baron, Hershey, Kunreuther, and Schwartz-Buzaglo, 1996).

Note that the illusory correlation effect (described in Chapter 8) could lead to violation of the neutral-evidence principle. If people interpret zero correlation as consistent with their belief in a positive correlation, they will maintain that belief more tenaciously than they should. The experiments on illusory correlation, together with the experiment by Lord and his colleagues in which subjects tended to find flaws only in the evidence that went against them, suggest that a major mechanism of irrational persistence involves distortion of one’s perception of what the evidence would mean to an unbiased observer.

An extreme example of the violation of the neutral-evidence principle was found by Batson (1975). In his study, the evidence presented was not even neutral, but was entirely against the belief in question, for the relevant subjects. Fifty female high school students who attended a church-sponsored youth program were given a questionnaire that included items concerning the divinity of Jesus. For example, “Jesus actually performed miracles,” and “Jesus was only human.” The students were then divided into two groups, according to their answer to the question, “Do you believe Jesus is the Son of God?” (Of the fifty girls, forty two answered yes; eight answered no.) Subjects in the two groups were then asked to read, discuss, and evaluate some material purportedly “written anonymously and denied publication in the New York Times at the request of the World Council of Churches because of the obvious crushing effect it would have on the entire Christian world” (p. 180). The writings claimed to show, on the basis of newly discovered scrolls, that the New Testament was fraudulent. Of the forty-two believers, eleven accepted the veracity of the article. This group became even more convinced of the divinity of Jesus than they had been before reading the article. (The believers who did not accept the veracity of the article did not change their belief, and the nonbelievers also strengthened their disbelief in the divinity of Jesus, even though most of them did not accept the article either.) The believers who accepted the article had the greatest need to strengthen their belief in the divinity of Jesus, and they did so despite being given nothing but negative evidence.

**Effect of active open-mindedness on outcomes**

These are some of the basic demonstrations of irrational belief persistence. Before looking more closely at how it works, let us ask whether it matters. Does actively open-minded thinking help produce better decisions?

Several studies have looked for a correlation between good thinking and good outcomes. Herek, Janis, and Huth (1987) examined the thinking of U.S. presidents (and their advisers) about how the United States should respond to nineteen interna-
tional crises from the Greek civil war in 1947 to the Yom Kippur war in 1973. Historical records of decisions were evaluated for several symptoms of defective decision making, including “gross omissions in surveying alternatives” (inadequate search for possibilities); “gross omissions in surveying objectives” (inadequate search for goals); “failure to examine major costs and risks of the preferred choice” (inadequate search for evidence); and “selective bias in processing information at hand” (biased interpretation). Experts in international affairs assessed the outcomes, taking into account the best interests of the United States and (separately) the best interests of the world. The symptoms of poor decision making correlated with poor outcomes (from either point of view). It is possible that the judgments of symptoms were influenced by the judges’ knowledge of the outcome, but the correlations were high, and some of the crises were quite obscure, so the effect is probably real. Good thinking does seem to correlate with good outcomes.

Another type of study has examined correlations between measure of actively open-minded thinking and measures of the ability to solve various problems. Some of these problems are of the sort given in school, and others are of the sort used throughout this book to illustrate biases in thinking. I shall suggest that actively open-minded thinking helps to reduce some of these biases, and these studies provide first-blush evidence for that suggestion. They show that people who think in a way that is actively open-minded are better at solving the problems, or less biased.

In one study, Stanovich and West (1998) reported several experiments, involving the presentation of several tasks to each of hundreds of college students. One task, the Argument Evaluation Test, measured myside bias in the evaluation of arguments. Each item began with Dale (a fictitious person) stating an opinion about a social issue, for example, “The welfare system should be drastically cut back in size.” The subject indicated agreement or disagreement (to indicate the subject’s side). Dale then gave a justification, for example, “because welfare recipients take advantage of the system and buy expensive foods with their food stamps.” A critic then presented a counterargument, for example, “Ninety-five percent of welfare recipients use their food stamps to obtain the bare essentials for their families.” Finally, Dale rebuts the counterargument, for example, “Many people who are on welfare are lazy and don’t want to work for a living.” The subject then evaluated the strength of the rebuttal on a four point scale. The subject’s answer was compared to answers given by experts — philosophy professors at the University of California, Berkeley, and Stanovich and West. To estimate myside bias, the authors tried to predict the subject’s ratings from both the expert ratings and the subject’s own opinion about the issue. Myside bias was defined as a positive effect of the subject’s beliefs. That is, subjects showing myside bias were those who tended to deviate from the expert ratings in the direction of their own opinions, rating arguments as better when they agreed with that opinion. Most subjects showed some myside bias, but some were more biased than others. The question is whether the less biased subjects do better on other tasks (compared to the more biased subjects).

This happened for several tasks, but not for all the tasks used. Myside bias may be a common problem in reasoning, but it isn’t the only problem. Students with less my-
side bias did better on a test of logical syllogisms and the Wason four-card problem (Chapter 4), a test involving attention to statistical evidence rather than anecdotes, a measure of efficient hypothesis testing, and a measure of the perception of correlations. No measure of myside bias correlated with inappropriate extreme confidence or correct use of Bayes’s theorem in probability judgment (Chapter 6). Subjects low in myside bias also got better scores on tests of general ability, such as the Scholastic Achievement Test and the Ravens Progressive Matrices (Chapter 2).

**Determinants and related phenomena**

Let us now consider some possible determinants of irrational belief persistence.

**Beliefs about thinking**

General beliefs about thinking itself can play a role. People have their own standards for thinking, some of which encourage poor thinking. The heuristics that we use to form our beliefs are maintained by certain explicit beliefs about how thinking should be conducted — beliefs transmitted through the culture (Baron, 1991; Perkins, Allen, and Hafner, 1983). People differ in their beliefs about how one should draw conclusions. Some think that changing one’s mind is a sign of weakness and that a good thinker is one who is determined, committed, and steadfast. Such people, if they followed their own standards, would be more likely to persist in beliefs irrationally. Others believe that good thinkers are open-minded, willing to listen to the other side, and flexible. Most of us probably subscribe somewhat to both of these beliefs. Whatever our beliefs, most of us desire to be good thinkers, so we try to follow our own standards.

The last chapter discussed the work of Kuhn on the kinds of standards that people apply to their own thinking. Other evidence for a role of beliefs about thinking comes from the study of Stanovich and West (1998), just described. They found that subjects with less myside bias had beliefs about thinking itself that tended to favor active open-mindedness. They endorsed items such as, “People should always take into consideration evidence that goes against their beliefs,” and they disagree with items such as “Changing your mind is a sign of weakness.”

My own research has found similar results (Baron, 1989). I measured subjects’ beliefs about good thinking, in two different ways. First, subjects were asked how they thought people ought to respond to challenges to their beliefs. How, for example, should college students respond when they meet new ideas about religion or politics? Subjects were classified according to whether or not they thought people ought to think further, with a view to revising their beliefs if it is warranted. Second, subjects were asked to give grades (A through F) to hypothetical thinking protocols for the quality of thinking. Some protocols considered arguments on only one side of an issue (for example, on the question of whether automobile insurance rates should be higher for city dwellers than for suburbanites: “My first thought is that each group
of people should pay for its own accidents. City dwellers surely have more accidents, and their cars get broken into and stolen a lot more”). Other arguments presented evidence on the opposite side as well (for example, “On the other hand, it doesn’t seem fair to make people pay for things they can’t help, and a lot of people can’t help where they live”). Subjects’ thinking itself was also measured by looking at whether they themselves produced two-sided or one-sided arguments when asked to consider some question, such as the question about ocean-floor minerals described in Chapter 3. Those thinkers who gave higher grades to two-sided protocols, and who thought that we should be open-minded when our beliefs are challenged, were more likely than other subjects to produce two-sided thinking themselves. It appears that people’s beliefs about thinking affect the way they themselves think.

Why do some fail to realize that two-sided thinking is better than one-sided thinking? It is possible that belief in one-sided thinking is the result of the evolution of institutions, such as organized religions and nations. To keep its adherents from one generation to the next, each of these institutions must convince them that its views are correct, even though many outsiders will argue otherwise. Those institutions that inculcate an ideology in which defense of one’s belief is a virtue and questioning is a vice are the ones most likely to overcome challenges from outside.

Another possibility is that people confuse two different standards for thinking, which we might call the “good thinker” (active open-mindedness) and the “expert.” Because experts know the answer to most questions, they usually do not have to consider alternatives or counterevidence. If we admire experts, we may come to admire people who are “decisive” in the sense of being rigid. When a news commentator criticizes a political candidate for waffling and being unsure (as might befit a good thinker faced with many of the issues that politicians must face), the implication is that the candidate is not expert enough to have figured out the right answer. Similarly, a person who adopts a know-it-all tone — speaking without qualification or doubt — is giving a sign of expertise. Some parents (perhaps because they are experts about the matter under discussion) talk this way to their children, who come to think of it as a “grown-up” way to talk.

This confusion of expertise with good thinking may reinforce the institutional pressures. Those who are considered wise and respected members of the institution or group may talk like experts, encouraging their followers to “know” rather than to think. And how are the followers supposed to “know”? By listening to the experts, of course.

A third possibility is that people confuse the standards of the thinker with those of an advocate. A good lawyer is an advocate for her client. She tries to defend her own side, and she considers the other side of the case only for the purpose of rebutting it. It is inconceivable that she would change her mind, at least in court. She deliberately takes sides, knowing that there is another lawyer on the other side, and a judge to ensure that the opponent is treated fairly. Similarly, in democratic groups, public-spirited people often advocate a point of view they do not necessarily accept but feel is neglected, knowing that the other side of the issue will be well defended. Thus the individual can approach an issue in a one-sided way with the comfort of knowing that
the group as a whole will “think well,” in the sense of considering alternatives and counter-evidence. There is room for one-sided advocacy as part of a larger process of two-sided (or many-sided) group thinking. Even in groups, however, respect and tolerance for the other side is required if the group is to function well. The danger is that people’s standards for thinking may be confused with standards for skill as an advocate. That is why debating teams do not necessarily encourage good thinking.

Distortion of beliefs by desires

We now consider the ways in which beliefs are affected by desires (long-term personal goals or temporary goals). These effects may help to explain irrational belief persistence, and they are also of interest in their own right. They have long been known to psychotherapists as types of bias that can seriously interfere with personal functioning, but they are probably just as insidious in the realm of politics.

Self-deception and wishful thinking. Persistence in an irrational belief can be a kind of self-deception in which we make ourselves believe something through the use of heuristics or methods of thinking that we would know (on reflection) are incorrect. By this view, if we were aware that our thinking was biased when we did it, we would not accept its results. This account assumes that irrational persistence occurs even in people who can recognize good thinking in general when they see it.

The best evidence for self-deception as a phenomenon in its own right comes from a study that has nothing to do with belief persistence. Quattrone and Tversky (1984) first asked each subject to take a cold pressor pain test, in which the subject’s arm was submerged in cold water until the subject could no longer tolerate the pain. After that, the subject was told that recent medical studies had discovered two types of hearts, one type being associated with longer life and fewer heart attacks than the other. The two types could be distinguished by the effect of exercise on the cold pressor test. Some subjects were told that exercise would increase tolerance in the good type of heart; others were told that it would decrease tolerance in the good type. Subjects then repeated the cold pressor test, after riding an exercycle for one minute.

In general, subjects’ tolerance changed in the direction that they were told was associated with a good heart. If they were told that exercise increased tolerance in people with good hearts, they managed to tolerate the cold water a bit longer, and vice versa. Only nine of the thirty-eight subjects indicated (in an anonymous questionnaire) that they had purposely tried to change in the direction associated with a good heart. The remaining twenty-nine showed just as large a change in tolerance (in the good direction) as the nine who admitted that they tried to change. In general, the nine who admitted trying to control their results did not believe that they really had a good heart, but the twenty-nine who did not admit to “cheating” did believe it. The nine admirers therefore failed in their attempt to deceive themselves, because they were caught in the act (by themselves, of course), and therefore they could not accept the results of the deception. The twenty-nine others were successful in keeping from themselves what they had done to create their beliefs.
This experiment illustrates the essential features of all self-deception (see Elster, 1979, 1983): the presence of a desire to have a certain belief; an action or inaction designed to create or strengthen that belief; and an unawareness of the relation between the ultimate belief and the motivated action that gave rise to it. If you neglect to mention disturbing symptoms to your doctor, you must forget that you have done this, if you want to be cheered up when she pronounces you to be in excellent health.

Psychologists have found other examples of beliefs distorted in the direction of desires. McGuire (1960) showed wishful thinking in a complicated task involving syllogisms. Svenson (1981) found, for example, that most drivers believe that they are safer and more skillful than average, and Weinstein (1980) found that most people believe they are more likely than average to live past eighty. Babad and Katz (1991) found that people who bet on sports events are more likely to bet on a win by their favored team, the team they hope will win. Some betting arrangements allow odds to be set by the bets made (rather than, say, expert opinion). In such cases, it should theoretically be possible to make money by betting against the team or outcome that most bettors favor.

Wishful thinking can affect decisions. Weeks and her colleagues (1998) looked at the decisions of patients who had colon cancers that had spread to their livers or who had advanced lung cancer. All the patients had a very short life expectancy. Patients and their doctors estimated each patient’s probability of surviving for six months. Doctors were accurate on the whole, but patients overestimated their probability of survival, sometimes drastically. These patients were all faced with a choice of palliative therapy to relieve their pain and discomfort or aggressive treatment designed to extend life, such as chemotherapy. The aggressive treatment typically makes life worse and has little benefit in these patients. The patients who disagreed the most with their doctors in overestimating their chance of survival were the most likely to choose aggressive therapy.

Some wishful thinking is not irrational, however, as we noted in Chapter 3, for certain unpleasant beliefs may themselves make us unhappy or unable to carry out our other plans. The prescriptive difficulty is to set policies for ourselves that will permit wishful thinking and self-deception only when they are harmless or useful.

Dissonance resolution: eliminating conflict among beliefs. A related phenomenon occurs when we make a difficult decision. Often there are reasons favoring the path not taken. After the decision, we seem to give these reasons less weight than we gave them before the decision. Festinger (1962) asked adolescent girls to rate a number of popular records for attractiveness. The experimenter then asked each girl to choose one of two records that she had rated as moderately attractive. When the records were rated again, the record chosen was rated higher than before, and the record rejected was rated lower than before. Presumably the subjects were more convinced, the second time, that they had good reasons for their original decisions.

In a similar experiment (Festinger, 1962), subjects were paid to write an essay advocating a position (in politics, for example) with which they disagreed. Those paid only a little to do this tended to change their opinion in the direction of the essay they wrote. Those paid a lot did not change. Those paid only a little desired to
believe that they had written the essay for a good reason. Because they did not do it for the money, they apparently convinced themselves that they had done it because they really agreed with the position more than they initially thought.

In another classic study, Festinger and Carlsmith (1959) induced subjects to participate in a psychology experiment that they deliberately made boring and tedious. After the experiment, each subject was asked to convince the next subject (actually a confederate) that the experiment was interesting and fun. Half of the subjects were paid $1 for their participation in the experiment; the other half were paid $20. After this, the subjects were interviewed about their true opinion of the experiment. The group paid $1 had a more favorable opinion of the experiment.

Festinger (1962) explains these results as a process of “reduction in cognitive dissonance.” When the choice is difficult, the reasons for one decision are “dissonant” with the reasons for the other, and the dissonance can be reduced by playing down the reasons for the choice not made or inventing reasons for the choice made (for example, that the dull experiment was really interesting).

Surely we try to eliminate conflict among our beliefs. Most of our attempts to do this, however, are completely rational. When we find evidence against a belief that we favor, for example, we often reduce the strength of the belief, so as to “reduce the dissonance.” These experiments, however, seem to show some sort of irrationality.

What is irrational here? The idea of “dissonance reduction” does not by itself seem to capture it.

The reason for this sort of postdecisional change could be that people like to believe that they are “good” decision makers — both morally good and intelligent. They change their beliefs about their reasons for having made a decision so that their beliefs fit their desire that they be good in these ways. When they write an essay opposing their real view for only a small amount of pay, it is easiest to justify that decision (a bad financial deal) by thinking that they truly have some sympathy for that position. Likewise, they may justify doing a boring task for a small amount of pay by thinking that they actually liked the task. Similarly, when they make a difficult decision between two choices, they may later have doubts about whether they made the best decision unless, in retrospect, they see the decision as not so difficult at all; they therefore play down the value of the rejected choice. These experiments seem to illustrate a form of wishful thinking, where the “wish” is the desire to have been a good decision maker, and the beliefs that are affected are those about reasons for having made a decision.

A number of experiments support this view. Cooper (1971), for example, found that the effect of past decisions on beliefs was larger when the outcome of the decision could be foreseen than when it could not. The decision in question involved agreeing to work with another subject (a confederate) in a task in which the amount of payment depended on the performance of both subjects. In order to receive high payment, both subjects had to solve aptitude test problems and, after doing each problem, indicate accurately whether they had answered it correctly or not. The subject was told that her partner was either “too timid to publicly state that she had [the problem] correct” or else was “a little too sure of herself.” In fact, the partner did
(always) lower the score for both subjects by being either too timid or too overconfident, but half of the time the source of the difficulty was the opposite of what the subject expected. The interesting result was that the subject liked the partner more (thus justifying her own decision to take part in the experiment) when the difficulty had been foreseen than when it had not. When the difficulty was the opposite of what was expected, the subject could not have foreseen it and therefore had less reason to convince herself that her decision had been a good one.

Another study (Aronson, Chase, Helmreich, and Ruhnke, 1974) showed that subjects can even feel responsible for unforeseen (but potentially foreseeable) outcomes and that this responsibility can lead them to justify their decision in hindsight. Subjects were asked to make a videotape recording in which they made arguments (from an outline the experimenter gave them) advocating government regulation of family size — a position with which they disagreed — ostensibly for use in another experiment. In one condition, subjects were told, before they agreed to make the tape, that those who would watch the tape would be either highly persuadable (other students who were unsure of their opinion) or not at all persuadable (students who were strongly against government regulation). Subjects who thought that their audience was persuadable changed their opinion more in the direction of favoring government regulation than the other subjects. Presumably, this is because the subjects who thought that they might persuade their audience thought that they would be doing a bad thing to convince someone else of a totally erroneous view, so they convinced themselves that the view was acceptable, and thereby also maintained the belief that they themselves were good people.

In another condition, the subjects were told nothing about the audience until after they had made the tape. The same effect was found, although it was smaller. That is, subjects changed their own attitude more in the direction of favoring regulation when they were told, after making the videotape, that the audience was persuadable, even though they had known nothing about the persuadability of the audience when they agreed to make the tape. Apparently these subjects blamed themselves for being duped. The effect was not found when the subjects were told initially that the tape would not be shown to anyone and were later told that it would be. In this condition, subjects were able to let themselves “off the hook” because the experimenter’s lie had been so blatant.

In another experiment, Scher and Cooper (1989) asked college students if they were willing to write essays either for or against an increase in student fees (a proposal that most students opposed). The essays would be shown to the “Dean’s Committee on Policy” as part of a study of the decision-making process of that committee. After writing the essay, each student was told either that the essay would be one of the first shown to the committee or one of the last (picked at random). Subjects were also told — and here is the trick — that previous research indicated that the first few essays had a “boomerang effect,” that is, “they tend to convince the committee to take the other side.” Subjects whose essays would be read early thus thought that their essay would move the committee in the opposite direction from the position advocated in the essay.
The results were clear: The effect of the essay on the student’s own opinion depended on the effect that the essay was supposed to have on the committee, not on what the essay said. In particular, subjects who wrote against the fee increase but whose essays would be read early changed their opinion about the fee increase just as much as subjects who wrote against the increase and whose essays would be read late. This study shows clearly that the effect of writing the essay is not in the “dissonance” between the content of the essay and the student’s original belief. The results are easily explained by the hypothesis that subjects wanted to avoid self-blame for moving the committee in the wrong direction. If the direction were not so wrong, less self-blame would be needed.

In general, then, people do not like to think of themselves as liars or bad decision makers, and they manipulate their own beliefs so as to convince themselves that they are not, and were not in the past. This appears to be a type of wishful thinking, possibly also involving self-deception.

What is the relation between the effects we have been discussing and the general phenomenon of the irrational persistence of belief? For one thing, the “dissonance” experiments are a type of irrational persistence in their own right. What seems to persist is each person’s belief that he is a good decision maker, moral and intelligent. This belief is maintained, however, in a peculiar way. When a person runs into evidence against the belief, evidence suggesting that a bad decision may have been made, the person changes his beliefs about his own desires (“I must really have wanted it, or I wouldn’t have done it for so little money,” or “put in so much effort,” and so forth). These beliefs about desires, in turn, may influence the desires themselves, as we have just seen.

Just as we want to think of ourselves as good decision makers, we want to think of ourselves as good belief formers. When a belief is challenged, our first impulse is often to bolster it (Janis and Mann, 1977), in order to maintain our belief in our earlier intelligence. We want to have been right all along — whereas it would be more reasonable to want to be right in the present (even if that means admitting error). This is what makes us into lawyers, hired by our own earlier views to defend them against all accusations, rather than detectives seeking the truth itself.

Related results

Two other results show essential mechanisms underlying irrational persistence. Selective exposure is the tendency to search selectively for evidence that will support current beliefs. Belief overkill is the tendency to deny conflicting arguments, even if they do not need to be denied.

Selective exposure

People maintain their beliefs by exposing themselves to information that they know beforehand is likely to support what they already want to believe. Liberals tend to read liberal newspapers, and conservatives tend to read conservative newspapers.
Those who voted for George McGovern for president of the United States in 1972 watched eagerly as the winner, Richard Nixon, was raked over the coals in the Watergate affair of 1973, while those who had supported Nixon were relatively uninterested (Sweeney and Gruber, 1984).

In an experiment conducted during the 1964 election campaign, subjects were given an opportunity to order free brochures either supporting the candidate they favored or supporting his opponent (Lowin, 1967). Subjects received samples of the contents of each brochure. When the arguments in the sample were strong and difficult to refute, subjects ordered more brochures supporting their own side than brochures supporting the other side. When the arguments in the sample were weak and easy to refute, however, subjects tended to order more brochures on the other side. People can strengthen their own beliefs by convincing themselves that the arguments on the other side are weak or that their opponents are foolish, as well as by listening to their own side. Many other studies have found this sort of bias toward information that can strengthen desired beliefs (Frey, 1986).

Selective exposure can lead to self-deception. Imagine that you want to believe that some course of action is correct, so you ask a friend to tell you all the reasons why it is good. If you then say to yourself, “This must be a great plan, because it has only good points and no bad points,” you are neglecting the fact that you have not asked for the bad points. People who select biased information and then believe it as though it were unbiased are manipulating their own beliefs.

This kind of manipulation is particularly easy, because people tend to change their beliefs in response to one-sided evidence even when they know it is one sided. They under-compensate for their knowledge that there is another side. Brenner, Koehler, and Tversky (1996) presented subjects with one side of the evidence concerning a legal dispute. The subjects predicted that other subjects given the role of jurors, who would hear both sides, would favor the side presented. This effect was sharply reduced if the subjects were asked, “Recall that the ‘jury’ subjects read both the plaintiff’s arguments and the defendant’s arguments. On the basis of what you’ve read above, do you expect the defendant’s arguments to be weaker or stronger than the plaintiff’s arguments?” Evidently, the subjects did not ask themselves this question when they heard only one side.

Belief overkill

A second related phenomenon is belief overkill. Many controversial issues are controversial because there are good arguments on both sides. A rational decision would involve balancing the arguments in a quantitative way, a way that takes into account their relative strengths. But people find ways to avoid this balancing. Through wishful thinking, they convince themselves that all the good arguments are on one side. Robert Jervis (1976, pp. 128–142) provides many examples of this kind of overkill in

Early studies of selective exposure failed to find an overall preference for information on the subject’s side, but the arguments on the other side in these studies were often easy to refute, so subjects who chose information on that side might have done so in order to refute it (Freedman and Sears, 1965).
judgments about foreign policy. In discussions of a ban on testing nuclear weapons, “People who favored a nuclear test-ban believed that testing created a serious medical danger, would not lead to major weapons improvements, and was a source of international tension. Those who opposed the treaty usually took the opposite position on all three issues. Yet neither logic nor experience indicates that there should be any such relationship. The health risks of testing are in no way connected with the military advantages, and a priori we should not expect any correlation between people’s views on these questions.”

Attitudes about capital punishment provide a good example of overkill (Ellsworth and Ross, 1983). It is possible in principle to believe that capital punishment is morally wrong yet effective as a deterrent against serious crimes, or morally acceptable yet ineffective. Yet almost nobody holds these combinations of belief. Those who find it morally wrong also think it is ineffective, and vice versa.

Here is an example of the process at work, from a subject in a study of reasoning, asked his opinion about animal experimentation: “We actually have no right to do such things. It’s not even necessary. If it was necessary, maybe there would be a reason for it, but, there’s no need for it, I don’t think. We’re sort of guardians here.” The subject’s intuition was that animals have a right to our protection. He could believe this and still also think that we could gain some benefit from experimenting on them, a benefit that we ought to forgo. But, instead, he convinced himself that no benefit exists (dos Santos (1996).

Factors that moderate belief persistence

We have been discussing the major causes of irrational belief persistence. Several other factors have been shown to increase or decrease the amount of this bias that subjects show — or the extent to which thinking is actively open-minded — although none of these factors could provide a sufficient account of the existence of the bias.

Elastic justification

Desires affect beliefs more easily when other determinants of belief are weaker. Hsee (1996), for example, gave American subjects a test of “language intuition” in which they had to pair twenty Chinese symbols with their meanings. The subjects scored their own tests. Hsee told the subjects in the “inelastic” condition to count only the odd-numbered items. He told the subjects in the the “elastic” condition that some of the symbols looked “yin” and others looked “yang,” that the meaning of these terms was a matter of individual perception, and that their score should be computed from the ten items that looked most yang (which subjects could, if they wanted, determine with the scoring key in view). The subjects were paid through a lottery in which the number of dollars they could win was their score on the test (0–100). The subjects wrote their scores on a separate sheet for submission to the lottery, so that nobody could check the honesty of their scoring.
The subjects in the inelastic condition were quite honest and gave the actual mean score of the group as their average score (despite the fact that cheating could not be detected). The subjects in the elastic condition, however, gave themselves higher scores than the mean, and higher than those of the inelastic group. In other words, they cheated. But they may not even have known they were cheating, because the cheating took the form of selecting the “yang” items in a way that achieved the highest score. Hsee suggests that this phenomenon can occur whenever people are conflicted between a “should” response and a “want” response. They will tend toward the “want” response when it is easier to search in a biased way for evidence that supports that response, that is, when justification is elastic.

Value conflict
People think in a more actively open-minded way when they must make a judgment or decision involving values (goals) that are both strong and conflicting. Tetlock (1986) showed this effect by asking college students to write down their thoughts about questions that involved conflicting values, such as “Should the [Central Intelligence Agency] have the authority to open the mail of American citizens as part of its efforts against foreign spies?” — a question that pits national security against individual freedom. Each subject was asked to rank all the values that were pitted against each other by the various questions. Tetlock measured “differentiation” of the response, which was, in essence, the tendency to consider both sides. The differentiation of subjects’ thinking was higher when the values underlying the question were ranked close together (and when they were both highly ranked). Subjects thus tended to give a differentiated answer to the question about opening mail if they valued both national security and individual freedom. Subjects who ranked only one of these values highly found the question easy to answer and were less prone to consider evidence on both sides.

Accountability
When we express our beliefs to others, we usually have the goal of being liked by those people. We therefore tend to accommodate our statements to the beliefs of our audience. When we must justify our views to an unknown audience, we are inclined to imagine various possible audiences, so that we try to accommodate our statements to many different points of view. In this way, accountability for one’s judgments increases active open-mindedness.

Tetlock (1983b) found that accountability reduced the primacy effect in a judgment task. Subjects read evidence concerning the guilt or innocence of a defendant in a criminal case. When subjects did not need to justify their judgments, they showed a strong primacy effect: They judged the defendant as more likely to be guilty when the evidence pointing toward guilt came first than when it came second. (Subjects had no reason to think that the order of the evidence mattered, so this experimental condition demonstrates a clear violation of the order principle, described earlier in
FACTORS THAT MODERATE BELIEF PERSISTENCE

In a similar study, Tetlock and Kim (1987) gave each subject another person’s answers to some items of a personality questionnaire and asked the subject to predict the person’s answers to other items and to give confidence ratings. When the experimenter told the subjects that they would have to justify their judgments in a tape-recorded interview after the experiment, their judgments were more accurate, their overconfidence was reduced, and their thinking was more “integratively complex.” These beneficial effects of accountability on the quality of thinking imply that thinking will be best (and irrational belief persistence minimized) in situations that require such accountability — such as a judge writing a legal opinion — and worst in situations that do not require it — such as a citizen voting by secret ballot.

In general, making people accountable to unknown audiences tends to reduce some biases and increase others (Tetlock, 1992). We may understand these effects by noting that accountability makes people think of more arguments on both sides. When a bias results from failing to take something into account — such as base rates in probability judgments — accountability helps to reduce the bias. However, if subjects think that other people tend to be biased in one direction more often than in the opposite direction, then accountability may increase the amount of bias in that direction. People’s theories about the biases of others — whether these theories are correct or not — may sometimes cause accountability to increase biases rather than reduce them. In sum, accountability to unknown others usually leads to an increase in actively open-minded thinking and a reduction in biases, but it can also lead to a tendency to accommodate to the real or imagined biases of others.

Stress

Janis and Mann (1977) proposed that the quality of decision making is affected by “stress,” which occurs when the decision maker cannot see how to avoid extremely negative outcomes. According to their “conflict theory model,” decisions are easy, involving little stress, when doing nothing (not changing from the status quo or default) involves little risk, or when there are serious risks of not changing but no risk of changing. These patterns are called “unconflicted adherence” and “unconflicted change.” When either option (change or no change) has risks, and when it is realistic to hope to find a better solution and sufficient time to do so, this yields a style of “vigilant” decision making, in which the decision maker seeks information and weights the options. Vigilant decision making occurs under moderate stress. If it is not realistic to hope to find a better solution, that is, either option will be worse than the status quo (although one might still be better than the other), the most common pattern is “defensive avoidance,” not thinking about the decision at all. Finally, if there is time pressure, a condition of frantic and disorganized search called “hypervigilance” may result, in which the decision maker considers one option after another, with little search for evidence. When the decision maker does seek evidence, the search is un-
systematic, and the most useful evidence is often overlooked. Defensive avoidance and hypervigilance are both examples of high-stress decision making. A unique feature of the conflict-theory model, for which much support exists, is the claim that decision making is highly influenced by situational factors. The same person may make rational, methodical decisions in one situation and awful decisions in others.

In some cases, the effect of stress results from time pressure alone. When time is short, decision makers are forced to restrict their search. Kruglanski and Freund (1983) have found that several biases, including the primacy effect, are increased by time pressure.

Keinan (1987) found that stress can impair thinking, even when time pressure is absent. Subjects were asked to solve analogy problems, such as “Butter is to margarine as sugar is to ... beets, saccharin, honey, lemon, candy, chocolate” (p. 640). The problems appeared on a computer display. The subjects had to examine each of the six alternative answers, one at a time, by pressing the corresponding number on the computer keyboard. The computer recorded the number of alternatives that were examined and the order in which they were examined for each problem.

When subjects were told to expect painful electric shocks during the experiment, they often responded before examining all of the alternatives (on the average, on about five of the fifteen problems). Other subjects, who expected no shocks, rarely did this (on the average, only once out of fifteen problems). Subjects who expected shock were less likely to scan the alternatives in a systematic order (for example, 1, 2, 3, 4, 5, 6), and they were less likely to answer correctly (36% correct versus 59% for the group that did not expect shocks). (In fact, no subject received any shocks.) These negative effects of stress occurred even when subjects thought they would receive shocks only if they answered incorrectly too often, so that the threat of shock provided an incentive to answer correctly. Keinan suggests that stress — in the form of fear — can distract our attention and cause us to search less thoroughly.

Groupthink

Biases such as irrational belief persistence are found when thinking is done by groups as well as by individuals. Since much important thinking is done by groups, this field is important. As we noted in Chapter 1, the thinking of groups has analogies with the thinking of individuals. Possibilities and goals are suggested; evidence (arguments) is brought forward; and conclusions are drawn. If, when the group first starts its work, everyone already has a fixed belief, it will be hard to give a fair hearing to other possibilities. There are “hanging juries” as well as “hanging judges.”

To a certain extent, the factors that operate in group thinking are the same as those that operate in individual thinking. After all, a group is simply a collection of individuals, and if the individuals making up a group tend to show a certain bias, then the group tends to show it as well.

In other ways, group thinking differs from individual thinking. Groups have an opportunity to overcome some of the biases shown by individuals, because it is possible to choose the members of the group so as to represent a variety of points of
FACTORS THAT MODERATE BELIEF PERSISTENCE

view. On the other hand, the individual members may be too willing to assume that this has been done when it has not been done. A group consensus sometimes seems much more obviously “right” than the same conclusion reached by an individual, even though the members of the group are all alike in sharing the characteristics that lead to the consensus. Would it be any surprise if a group of automobile workers in the United States “agreed” that importing of Japanese cars should be stopped, or if a group of Toyota dealers agreed that importing should not be stopped?

Janis (1982) studied the rationality of group decision making and the biases that distort group decisions by reviewing the history of major foreign-policy decisions made by the president of the United States and his advisers. Some of these decisions displayed poor thinking, and others displayed good thinking. As it turned out, the former led to poor outcomes and the latter led to good outcomes. It is to be expected that better thinking will lead to better outcomes on the average, but Janis is aware that the correlation is not perfect. He tried to select his cases according to the kind of thinking that went into them rather than according to the outcome. The examples include what he regarded as the poorly made decision of President Kennedy and his advisers to attempt the Bay of Pigs invasion in Cuba in 1961; the well-made decisions of practically the same group during the Cuban missile crisis in 1962; the poorly made decisions of President Johnson and his advisers to escalate the Vietnam War over several years; and the poorly made decisions of President Nixon and his advisers to withhold information concerning White House involvement in the Watergate break-in from 1972 to 1974.

Janis’s selection of cases was supported by Tetlock’s study (1979) analyzing the “integrative complexity” of public statements by prominent decision-making groups. Tetlock found that the poorly made decisions were associated with statements at or near the lowest level, but the well-made decisions were associated with well-differentiated statements. (Other recent studies of groupthink include those of McCauley, 1989, and Tetlock, Peterson, McGuire, Chang, and Feld, 1992. These studies deal with the nature of conformity in groupthink and with its causes, respectively.)

Janis characterized poor group thinking as “groupthink” (borrowing the term from George Orwell). He identified three major causes of groupthink, presented here in outline form:

Type I. Overestimation of the group
1. Illusion of invulnerability
2. Belief in the inherent morality of the group

Type II. Closed-mindedness
3. Collective rationalization
4. Stereotypes of out-groups

Type III. Pressures toward uniformity
5. Self-censorship
6. Illusion of unanimity
7. Direct pressure on dissenters
8. Self-appointed mind-guards
By contrast, when good thinking occurs in groups, there is a commitment of the group to a friendly (and sometimes not so friendly) interchange of arguments pro and con, not to a decision already tentatively made. Loyalty to the group is defined in terms of loyalty to the process of making the best decision, not loyalty to a decision already made. Visitors to President Kennedy’s inner circle during the Cuban missile crisis were often surprised at the freedom that members had to bring up seemingly irrelevant ideas and suggestions. (Kennedy apparently had learned something about group decision making from the Bay of Pigs.) Information was sought out from a variety of sources, especially people expected to disagree with the group, and these people were questioned thoroughly. (Janis suggests that assigning one member of the group to be devil’s advocate can help to prevent groupthink.)

Janis does not deny that there were other causes of poor decision making in his examples. For instance, he noted an excessive concern not to appear “soft on communism,” for domestic political reasons. Had Presidents Kennedy and Johnson considered the possible outcomes of their poor decisions motivated in this way, Janis argues, they would have realized that they were ultimately undercutting even this goal. One of the advantages of Janis’s analysis is that it can explain poor decision making while allowing that very good decision making could occur in similar circumstances. Decision makers are not simply the victims of their political biases, and the purported existence of these biases does not provide a full explanation of poor decisions. In good decision making, questioning is always possible.

Conclusion

This chapter has provided the main evidence for my claim that we tend to be biased in favor of our initial ideas. I showed this by comparing our responses to evidence with normative principles such as the order principle and the neutral-evidence principle. The prescriptive policy to avoid these biases is actively open-minded thinking. We have explored some of the factors that facilitate and inhibit such thinking.

Many of the biases discussed in this chapter are prevalent in conflict situations between two groups: for example, the Bosnians and the Serbs, the Israelis and their Arab opponents, each nation and its trading partners (or would-be partners), and the advocates and opponents of abortion, free trade, and many other public policies. 3 If people learned to think more rationally — to consider counterevidence and to form their ideologies with more sympathy for the variety of goals that people pursue — such conflicts could be reduced. We often suppose that only the other side thinks poorly, that they, not us, are the ones in need of education. Even if this is true, no harm is done by making sure that our own house is in order. Careful attention to the quality of our own arguments can even uplift — by example — the reasoning of our opponents.

3 For those interested in further reading on biases and conflict, Jervis (1976) provides an excellent discussion.
Are the biases discussed in this chapter an inevitable part of nature? Apparently not. First, we have discussed the evidence for individual differences. Some people show these biases a lot more than others, and it may be that some people do not show them at all in some situations.

Second, some evidence suggests that people can be taught to think in a more actively open-minded way. Beginning with Selz (1935), several studies have shown that something resembling actively open-minded thinking can be trained and that it transfers to new situations (Baron, Badgio, and Gaskins, 1986; Perkins, Bushey, and Faraday, 1986). Such training does not need to be outside of regular courses. It can often be done by changing the design of traditional course, often in ways that even facilitate the learning of the course content (Baron, 1993a).