Thinking AND Deciding

FOURTH EDITION

Jonathan Baron

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521862073

This page intentionally left blank

Thinking and Deciding, Fourth Edition

Beginning with its first edition and through three subsequent editions, *Thinking and Deciding* has established itself as the required text and important reference work for students and scholars of human cognition and rationality. In this, the fourth edition, Jonathan Baron retains the comprehensive attention to the key questions addressed in the previous editions — How should we think? What, if anything, keeps us from thinking that way? How can we improve our thinking and decision making? — and his expanded treatment of topics such as risk, utilitarianism, Bayes's theorem, and moral thinking. With the student in mind, the fourth edition emphasizes the development of an understanding of the fundamental concepts in judgment and decision making. This book is essential reading for students and scholars in judgment and decision making and related fields, including psychology, economics, law, medicine, and business.

Jonathan Baron is Professor of Psychology at the University of Pennsylvania. He is the author and editor of several other books, most recently *Against Bioethics*. Currently he is editor of the journal *Judgment and Decision Making* and president of the Society for Judgment and Decision Making (2007).

Thinking and Deciding

Fourth Edition

Jonathan Baron
University of Pennsylvania

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521862073

First published in print format 2007

ISBN-13 978-0-511-46043-2 eBook (Adobe Reader)

ISBN-13 978-0-521-86207-3 hardback

Contents

Preface to the fourth edition

The fourth edition retains many of the features of the first three editions:

1. Knowledge about judgment and decision making has been scattered among a number of different fields. Philosophers, psychologists, educators, economists, decision scientists, and computer scientists have different approaches to the theory. The approach in this book represents my own effort to draw together some of the key ideas from these different disciplines. Much of what I present is not original or new. If it were either of these, I would not be so confident that it is basically correct.

2. I retain the idea that all goal-directed thinking and decision making can be described in terms of what I call the *search-inference framework*: Thinking can be described as inferences made from possibilities, evidence, and goals that are discovered through searching.

3. I also argue that one main problem with our thinking and decision making is that much of it suffers from a lack of *active open-mindedness*: We ignore possibilities, evidence, and goals that we ought to consider, and we make inferences in ways that protect our favored ideas.

In the course of this book, I apply these ideas to the major concepts and theories in the study of thinking. I begin, in Part I, with general considerations: the nature of rationality; methods for studying thinking; and logic. Part II is concerned with belief formation, which is a form of thinking in which the goal of thinking is held constant. In this part, I introduce probability theory as a formal standard. Part III concerns decision making, including the making of decisions about personal plans and goals, and decisions that affect others, such as those that involve moral issues or matters of public concern. This part introduces utility theory, which formalizes many of the ideas that run throughout the book.

The fourth edition continues the trend of increasing the emphasis on judgment and decision making and correspondingly reducing the discussion of problem solving and logic. Nonetheless, I have retained the original title with the expectation that this edition will be the last, so it is no time to change that. Because I want this edition to be useful for a while, I have also attempted to emphasize fundamental concepts. I make less of an attempt at keeping up to date with current literature. In a few cases, however, my crystal ball says that some recent ideas in the literature will last, so I have tried to explain them. The same fallible crystal ball tells me that other ideas of some current interest are passing fads. Because I cannot cover everything, I have used this fallible judgment as a guide for exclusion.

Otherwise, the changes, although extensive, are mostly at the level of detail. The only major change in structure is in the chapter on morality. I have also made an effort to organize what some have claimed to be a disorganized heap of biases. The organization is listed in a table on p. 56, and I have attempted to refer back to this in much of the discussion.

Many people have provided useful comments and other assistance. For the first three editions, Judy Baron, Kathie Galotti, and anonymous reviewers each gave useful advice about several chapters. Other chapters or sections were helpfully read by George Ainslie, David Baron, Judy Baron, Dorrit Billman, Colin Camerer, Allan Collins, Craig Fox, Deborah Frisch, Robin Gregory, John C. Hershey, Joel Kupperman, Liang Zhuyuan, David Messick, Chris Poliquin, Peter Ubel, and Peter Wakker. Many students brought errors to my attention. Christie Lerch, as an editor for Cambridge University Press, provided the final, most demanding, most detailed, and most helpful set of criticisms and constructive suggestions concerning all levels of writing and organization. The book was formatted using $\LaTeX$, and figures were drawn (over many years) with Systat, Metapost, *R*, Xfig, and raw PostScript.

I am also grateful to many colleagues who have influenced my thinking over the years, including Jane Beattie, Colin Camerer, Deborah Frisch, John C. Hershey, Howard Kunreuther, David Perkins, Ilana Ritov, John Sabini, Jay Schulkin, Mark Spranca, and Peter Ubel.

I dedicate this edition to the memory of two colleagues whom I shall never forget: Jane Beattie and John Sabini.

Part I

THINKING IN GENERAL

Part I is about the basics, the fundamentals. Chapters 1 through 3 present the concepts that underlie the rest of the book. Chapter 1 defines thinking, introduces the main types of thinking, and presents what I call the search-inference framework for describing thinking. Chapter 2 introduces the *study* of thinking and decision making, including the three types of questions we shall ask:

1. The *normative* question: How should we evaluate thinking, judgment, and decision making? By what standards?

2. The *descriptive* question: How do we think? What prevents us from doing better than we do according to normative standards?

3. The *prescriptive* question: What can we do to improve our thinking, judgment, and decision making, both as individuals and as a society?

These three questions define the content of the book. We can ask them about every topic. The third chapter introduces a theory of the nature of *good* thinking and of how we tend to think poorly. By using the normative theory to evaluate our actual thinking, we can know how it must be improved if it is found wanting. In this way, we can learn to think more *rationally*, that is, in a way that helps us achieve our goals.

Chapter 4 briefly introduces the study of logic. This is an older tradition in both philosophy and psychology. It is of interest because it has, from the time of Aristotle, taken roughly the approach I have just sketched. Logic provides a standard of reasoning. Although people often reason in accord with this standard, they sometimes depart from it systematically. Scholars across the centuries thus have asked, “How can we help people to think more logically?”

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 e*valu*ation, 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).

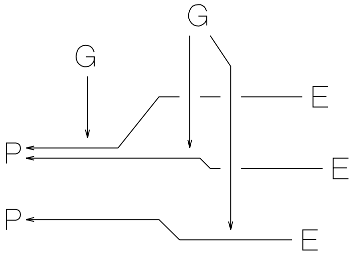
Goals are criteria used for evaluating possibilities. In the apartment-hunting example, goals include factors such as rent, distance from work or school, safety, and design quality. The goals determine what evidence is sought and how it is used. It is not until you think that safety might be relevant 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:



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.[1] 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.

[1] For 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?[2] 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

[2] Hint: What else are such numbers also divisible by? Another hint: What is the smallest number of this form? Another hint: A and B can both be 0. Another hint: Is it divisible by 13?

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 "naive" is that they are now superceded by better theories. Many scientific theories today will turn out to be naive 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.

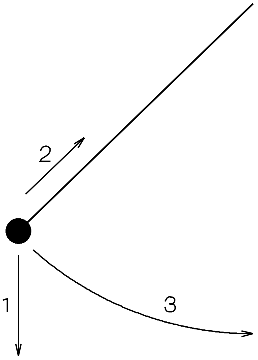
Vosniadou and Brewer (1987) point out that when the modern view is finally adopted, the change is truly radical. First, the concepts themselves are replaced with new concepts. For instance, whereas some young children believe that the sun is alive, sleeps at night, and could stop shining if it wanted to, older children learn that the sun is a star like the others that shine at night. Second, the relationships among the child's concepts change. The earth is no longer physically at the center of the universe; the light from the moon is understood as related to the positions of the earth and sun. Finally, the new system explains different phenomena. It explains the relationship between sun, moon, and stars, but *not* the relationship between sun, clouds, and wind (which might have been understood as being interrelated in the old system). This last point is of interest because it suggests that something is lost by adopting the new system — not just the innocence of childhood, but also a way of understanding certain things. The appearance of the earth as being flat and of the sun as being much smaller than the earth now become mysteries. Eventually, these things too will be explained, of course. In sum, as children adopt adult astronomy there are changes of belief that are as radical as those that have occurred in history, from the system of ancient astronomy to the system of Copernicus.

Just as children have naive theories of astronomy, children and adults seem to have naive theories in other subject areas, such as physics, which must be replaced, sometimes with great difficulty, in order for a person to learn a modern scientific theory. Often these naive theories correspond to systems proposed by early scientists such as Aristotle. (We shall see in the last two parts of this book that there may also be naive theories of judgment and decision making, theories held by most adults.)

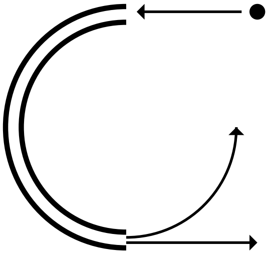
Clement (1983), for example, found that students who had taken a physics course held a theory (sometimes even after they had finished the course) about physical forces that was similar to one held by Galileo in his early work but that later both he and Newton questioned. The students believed that a body in motion always requires a force to keep it in motion. In contrast, we now find it more useful to suppose that a body keeps going unless it is stopped or slowed by some force. Of course, wagons and cars *do* require force to keep them going, but that is because they would otherwise be slowed down by friction.

Clement asked his subjects what forces were acting on a coin thrown up into the air, during the time it was rising but after it had left the thrower's hand. Most students (even after taking the course) said that there was a force directed upward while the coin was rising, in addition to the force of gravity pulling the coin down. This view fits in with the "motion implies force" theory that the students held. Physicists find it more useful to suppose that there is only one force, the force of gravity, once the coin is released.

Clement also asked the students about the forces acting on the bob of a pendulum while the pendulum is swinging. Many students said that there were three forces: (1) the force of gravity, pulling the bob straight down; (2) the force of the string, pulling the bob directly toward the point where the string was attached; and (3) a force acting in the direction in which the bob was moving, as shown in the following diagram. A modern physicist would say that the third force is unnecessary to explain the motion of the bob.



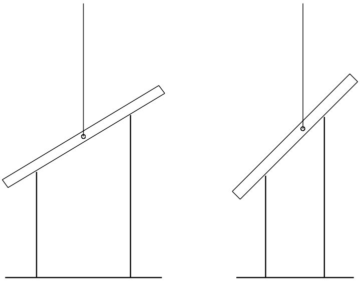
In another series of studies, McCloskey (1983) asked undergraduates, some of whom had studied physics, to trace the path of a metal ball shot out of a curved tube at high speed, as shown in the following illustration. The tube is lying flat on top of a surface; therefore the effect of gravity can be ignored. Many of the subjects, including some who had studied physics, said that the path of the ball would be curved. In fact, it would be straight. McCloskey argues (on the basis of interviews with subjects) that these students held a theory in which the ball acquires some sort of "impetus," a concept something like the mature concept of momentum, except that the impetus can include the curvature of the path. A similar theory was apparently held during medieval times.



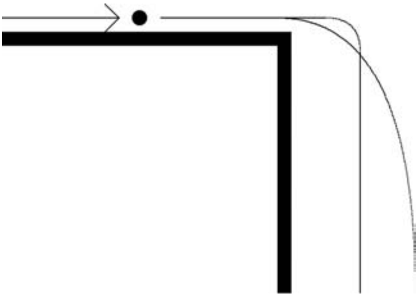
Roncato and Rumiati (1986) showed subjects a drawing like the following,[3] which shows two bars, each supported by a cable from the ceiling, attached to a pivot through the bar, around which the bar could turn freely. The thicker lines under each bar are supports. What happens when the supports are carefully removed? (Think about it.) Most subjects thought that the first bar would become horizontal and the second bar would remain tilted (although perhaps at a different angle). They seemed to think that the angle of the bar would indicate the discrepancy between the weight on the two sides, as a balance scale would do. In fact, the first bar would

[3] Reproduced with the authors' permission.

not move. (Why should it?) The second would become vertical. Perhaps this naive theory results from reliance on a false analogy with a balance.



Another of McCloskey's studies involved asking subjects to trace the path of a ball after it rolls off the edge of a table. One incorrect answer is shown in the following illustration, along with the correct answer. Subjects seem to think that the impetus from the original force takes a little time to dissipate, but that once it does, gravity takes over and the ball falls straight down (much like movie cartoon characters, who usually look first, then fall). In fact, the momentum from the original push keeps the ball moving at the same speed in the horizontal direction, and the path changes direction only because the downward speed increases.



McCloskey (pp. 321–322) argues that these naive theories are not entirely harmless in the real world:

> An acquaintance of ours was recently stepping onto a ladder from a roof 20 feet above the ground. Unfortunately, the ladder slipped out from under him. As he began to fall, he pushed himself out from the edge of the roof in an attempt to land in a bush about 3 feet out from the base of the house .... However, he overshot the bush, landing about 12 feet from the base of the house and breaking his arm. Was this just a random miscalculation, or did our acquaintance push off too hard because of a naive belief that he would move outward for a short time and then fall straight down?

Naive theories may also have some advantages. Kempton (1986) found that people tend to hold two different theories of home heat control. The physically correct theory for the vast majority of homes in the United States is the *feedback theory*. By this theory, the thermostat simply turns the heat on and off, depending on the temperature. As one Michigan farmer put it, "You just turn the thermostat up, and once she gets up there [to the desired temperature] she'll kick off automatically. And then she'll kick on and off to keep it at that temperature." By this theory, it does no good to turn the thermostat way up to warm up the home quickly. People who hold this feedback theory leave the thermostat set at a fixed value during the day.

Many people hold a different view, the *valve theory*. By this view, the thermostat is like the gas pedal of a car. The higher you turn it up, the more heat goes into the house, and the faster the temperature changes. People who hold this theory turn the thermostat way up when they come into a cold house, and then, if they remember, turn it down after the house warms up. The valve theory may well lead to wasted fuel, but it does give people a simple reason why they ought to turn the thermostat down when they are out of the house: less fuel will be used when the setting is lower.

The feedback theory is technically correct, as can be ascertained by looking inside a thermostat; however, it has some serious drawbacks. First, the valve theory does a better job than the basic feedback theory of explaining certain phenomena. In many homes, thermostats do need to be set higher to maintain the same feeling of warmth when it is very cold outside. This is easily explained by the valve theory, but in the feedback theory other concepts must be invoked, such as the fact that some rooms are less well heated than others and that some of the feeling of warmth may come from radiant heat from the walls and ceiling. Likewise, it may be necessary to turn the thermostat up higher than normal when entering a cold room, because the walls and furniture take longer to come to the desired temperature than the air does, and the room will still feel cold even after the air (which affects the thermostat) has reached the desired temperature.

Second, the feedback theory does not easily explain why it is a good idea to turn the heat down when one is out of the house. One valve theorist felt that the heat should be turned down when one is out, but, she said, "My husband disagrees with me. He ... feels, and he will argue with me long enough, that we do not save any fuel by turning the thermostat up and down .... Because he, he feels that by the time you turn it down to 55 [degrees], and in order to get all the objects in the house back up to 65, you're going to use more fuel than if you would have left it at 65 and it just kicks in now and then." Now the husband's reasoning here is physically incorrect. The use of fuel is directly proportionate to the flow of heat out of the house, and this, in turn, depends only on the temperature difference between the inside of the house and the outside. Thus, the house loses less heat, and uses less fuel, when it is at 55 degrees Fahrenheit than when it is at 65 degrees; but, as Kempton notes, the physically correct argument requires a more abstract understanding than most people typically achieve. If they act according to the valve theory, they may actually save more energy than if they act in terms of a rudimentary feedback theory, such as that held by the husband in this example.

We might be tempted to suppose that the valve theory is maintained by its functional value in saving fuel rather than by the ready availability of analogies with other valves (accelerators, faucets) and by its explanatory value. This conclusion does not follow. To draw it, we would need to argue that the functional value of the valve theory *causes* the theory to be maintained (Elster, 1979). Are people really sensitive to the amount of fuel they use? People's beliefs sometimes are for the best, but, as McCloskey argues, sometimes they are not.

This example is a particularly good illustration of naive theories, because it seems likely that the subjects have actually thought about how thermostats work. They have had to face the issue in learning how to use them. In the previous examples from college physics, it is not clear that the subjects really "had" any theories before they were confronted with the problems given them by the experimenters. They may simply have constructed answers to the problems on the spot. The fact that their answers often correspond to traditional theories simply reflects the fact (as it would on any account) that these theories explain the most obvious phenomena and are based on the most obvious analogies. After all, balls thrown with spin on them keep spinning; why should not balls shot out of a curved tube keep curving as well?

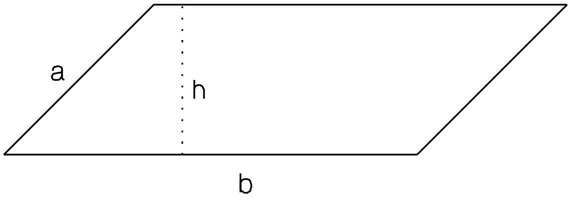
The home heat-control theories seem to provide yet another example of restructuring (assuming that some people change from the valve theory to the feedback theory). Like the Copernican theory of astronomy, the fully correct theory requires new concepts, such as the concept of heat flow over a temperature difference and that of radiant heat. The theory establishes new relationships among concepts, such as thermostat settings and heat flow. It also explains different phenomena, such as the fact that the house temperature stays roughly at the setting on the thermostat.

Understanding

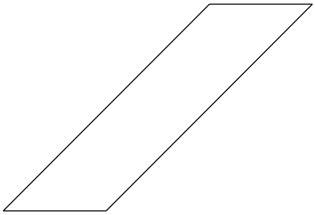
Students and their teachers often make a distinction between *understanding* something and "just memorizing it" (or perhaps just not learning it at all). Everyone wants to learn with understanding and teach for understanding, but there is a lot of misunderstanding about what understanding is. The issue has a history worth reviewing.

Wertheimer and Katona

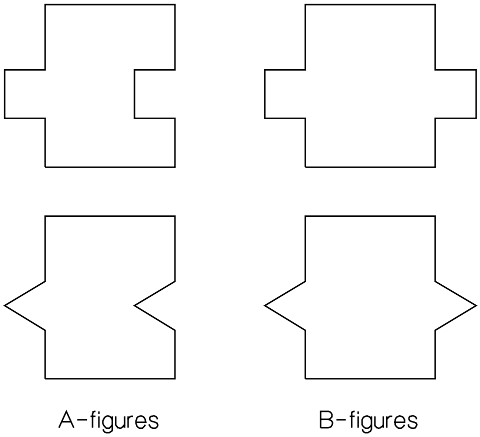
Max Wertheimer (1945/1959), one of the founders of Gestalt psychology in the early part of the last century, is the psychologist who called our attention most forcefully to the problem of understanding. Wertheimer's main example was the formula for finding the area of the parallelogram, $A = b \cdot h$, where A is the area, b is the base, and h is the height. Wertheimer examined a group of students who had learned this formula to their teacher's satisfaction. On close examination, though, they turned out not to understand it. They could apply it in familiar cases such as the following parallelogram:



But they refused to apply the formula to new cases, such as a parallelogram depicted standing on its side, which had not been among the original examples that they had studied:



They also were given other new cases (which Wertheimer called “A problems”) that followed the same principle: A rectangle was made into another figure by removing a piece from one side and attaching it to the opposite side, as shown in the following diagram, just as a parallelogram can be made into a rectangle by cutting a triangle from one side and moving it to the other side (without changing A, b, or h). Some students did indeed apply the formula to such cases, multiplying the base by the height to get the area, but these same students usually also applied the formula to other problems showing figures that could *not* be turned back into rectangles by moving a piece around (B problems). In sum, learning without understanding was characterized either by lack of transfer of the principle to cases where it applied or by inappropriate transfer to cases where it did not apply.



In contrast to these students, Wertheimer reported other cases of real understanding, some in children much younger than those in the class just described. Most of the time, these children solved the problem for themselves rather than having the formula explained to them. They figured out for themselves that the parallelogram could be converted into a rectangle of the same area without changing b or h. One child bent the parallelogram into a belt and then made it into a rectangle by cutting straight across the middle. Wertheimer does not insist that understanding arises only from personal problem solving, but he implies that there is a connection. When one solves a problem oneself, one usually understands why the solution is the solution.

Learning with understanding is not the same as discovery by induction. Wertheimer (pp. 27–28) made this philosophical point concretely by giving the class the values of a (the side other than the base), b, and h from each of the following problems (with parallelograms drawn) and asking the students to compute the area of each parallelogram:

a	b	h	Area to be computed
2.5	5	1.5	7.5
2.0	10	1.2	12.0
20.0	1 1/3	16.0	21 1/3
15.0	1 7/8	9.0	16 7/8

Wertheimer describes what happened:

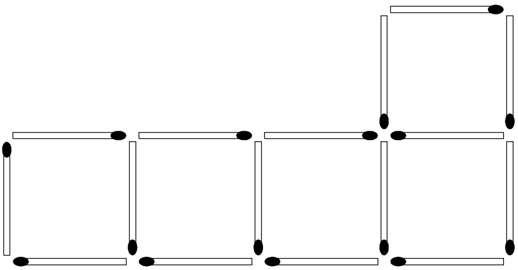
> The pupils worked at the problems, experiencing a certain amount of difficulty with the multiplication.
>
> Suddenly a boy raised his hand. Looking somewhat superciliously at the others who had not yet finished, he burst out: "It's foolish to bother with multiplication and measuring the altitude. I've got a better method for finding the area — it's so simple. The area is $a + b$."
>
> "Have you any idea why the area is equal to $a + b$?" I asked.
>
> "I can prove it," he answered. "I counted it out in all the examples. Why bother with $b \cdot h$. The area equals $a + b$."
>
> I then gave him a fifth problem: $a = 2.5$, $b = 5$, $h = 2$. The boy began to figure, became somewhat flustered, then said pleasantly: "Here adding the two does not give the area. I am sorry; it would have been so nice."
>
> "*Would it?*" I asked.
>
> I may add that the real purpose of this "mean" experiment was not simply to mislead. Visiting the class earlier, I had noticed that there was a real danger of their dealing superficially with the method of induction. My purpose was to give these pupils — and their teacher — a striking experience of the hazards of this attitude.

Wertheimer also pointed out that it is difficult or impossible to understand a principle that is "ugly" — that is, not revealing of certain important relationships in the matter to which it refers. An example would be a formula for the area that reduced to the simple formula only by some algebraic manipulation:

$$A = \frac{(b-h)}{(1/h - 1/b)}$$

Another way to put this, perhaps, is that the process that leads to understanding will fail to learn what cannot be understood. This process is unlikely to accept falsehood, even when propounded by authority, because falsehood is usually incomprehensible.[4] Of course, many facts are essentially arbitrary, so that “understanding,” in the sense in which we are using the term, is impossible. A statement such as this — “The Battle of Hastings was fought in 1066” — must be accepted without understanding.

Katona (1940), a follower of Wertheimer, made several additional observations concerning the relation between understanding and learning. Katona taught subjects how to solve different kinds of problems under various conditions, some designed to promote understanding and others designed not to do so. Certain of his problems concerned rearranging squares made of matchsticks so that a different number of squares could be made from the same number of matchsticks with a minimal number of moves. For example, make the following five squares into four squares by moving only three sticks:



The “no understanding” groups simply learned the solutions to a few such problems. The “understanding” groups were given a lesson in the relationship between number of matches, number of squares, and the number of matches that served as the border between two squares. The most efficient way to decrease the number of squares is to eliminate squares that share sides with other squares, so that the resulting squares touch only at their corners. For example, the square in the lower right is a good one to remove. You can then build a new square on top of the second by using the bottom of the second square for the top side of the new one. Groups that learned with understanding of the common-side principle were able to transfer better to new problems, and they even “remembered” the solutions to example problems more accurately after a delay, even though these examples were not part of the lesson itself. The difference between Katona’s demonstrations and Wertheimer’s is

[4] Scheffler (1965) makes related arguments.

mainly in the fact that subjects could see immediately whether they had solved Katona's problems or not, so they did not propose false solutions. Hence, Katona did not observe inappropriate transfer of the sort observed by Wertheimer.

Katona also pointed out that learning with understanding was not the same as learning with "meaning." Two groups were asked to memorize the number 5812151-92226. One group was given a meaning for the figure: They were told that it was the number of dollars spent in the previous year by the federal government. This group recalled the figure no more accurately than a group told nothing about the figure's meaning. A third group, however, which discovered the inner structure of the number (a series starting with 5, 8, 12, and so forth) did recall it more accurately.

What is understanding?

What does understanding mean? Can we reduce understanding to knowing certain kinds of things? If so, what kinds? One possibility, suggested by Duncker (1945, p. 5) is that "[knowledge of] the functional value of a solution is indispensable for the understanding of its being a solution." By "functional value," Duncker means the means–end, or subgoal–goal, relationship. For example, the parallelogram problem can be solved through the subgoal of making any nonrectangular parallelogram into a rectangle.

The idea that understanding involves knowledge of purposes or goals goes far toward explaining understanding. It helps us, for example, to see why the term "understand" is relative to some context, purpose, or goal. We can have partial understanding of some idea when we know the relationship of the idea to some goals but not others. For example, before I read a revealing article by Van Lehn and Brown (1980), I thought that I fully understood decimal addition. I found, however, that I was unable to answer a rather obvious question about decimal addition: "Why do we carry (regroup), rather than simply add up the numbers in a column whatever the sum (possibly putting commas between columns)?" Why do we not write the sum of 15 and 17 as 2,12? The reason given by Van Lehn and Brown is that if we wrote totals in this way we would violate the "cardinal rule" of our number system, the requirement that each quantity have a unique representation. This rule can be seen as a major goal — or purpose — that we have imposed on ourselves in the design of our number system. Even before I learned this, however, my understanding of some of the purposes of carrying was complete; for example, I did know that one reason for carrying, as opposed to simply writing a single digit and forgetting about the rest, was to "conserve quantity" — another (more crucial) goal imposed in the design of the number system.

Perkins (1986) suggests that there is somewhat more to understanding than knowing purposes. Understanding, Perkins states, involves knowing three things: (1) the structure of what we want to understand; (2) the purpose of the structure; and (3) the arguments about why the structure serves the purpose.

First, we must know the *structure* of the thing to be understood, the piece of knowledge, or what we shall (for reasons given shortly) call the *design*. We must know a general description of the design. Typically, this description refers to other concepts already known (or understood) for the relevant purpose. For example, $A = b \cdot h$ may be described *with reference to* the concepts "formula" (presumably already understood), "equality," and "variable." A full description would also indicate that b represents the base and h the height. Appropriate interpretation of the description is, practically always, facilitated by the use of at least one *model* or example in which the interpretation is given for a specific case.[5]

Second, we must know the *purpose*, or purposes, of the design: for example, to find the area of the parallelogram. The purpose also refers to still other concepts, such as "area" and the concept of "purpose" itself. So far, Perkins's theory fits well with the idea that the critical part of understanding is knowing purposes.

The third part of understanding, and the part that is new, is the group of *arguments* that explain how the design in fact serves the purpose. These arguments consist of other facts and beliefs. For example, in the case of the parallelogram, one argument explaining why the formula gives the area is that the parallelogram has the same area as a rectangle of the same base and height. Subsidiary arguments are relevant to each of these points: sameness of area, sameness of base, and sameness of height.

Perkins originally developed his theory for designs in general rather than knowledge in particular. A "design," as Perkins uses the term, is anything whose structure serves a purpose. It may be invented by people or it may evolve. In Perkins's theory, buttons, forks, and hands are designs; rocks and rainbows are not. Consider a pencil. One of its *purposes* is to write. Perhaps you have a wooden pencil with you (a *model*) that you can use to examine the *structure* of a pencil. What are the *arguments*? The pencil is soft, so that it can be sharpened (another purpose). The wooden shaft is (usually) hexagonal, so that the pencil does not roll off the table when you put it down (still another purpose). The pencil has an eraser on the end, so that . . . What other purposes might be served by the pencil's design? With this question, we can also criticize the ordinary pencil. It wears out. How can we change the design to prevent this? (The mechanical pencil.) Keep at it, and you will be an inventor, a designer.

Perkins's important insight was that *his* design — his theory of design — applied to knowledge as well: Hence, the title of his book, *Knowledge as Design*. We can think of just about any kind of knowledge this way. The decimal system that we all labor over in elementary school is quite an impressive design — far superior, for example, to the Roman system of calculation. The theories and concepts discussed in this book, and in many other academic subjects, are also designs. Often the purpose of such designs is to explain something. In mathematics, designs often have the purpose of helping us to measure something.

The nature of arguments may be understood by thinking of them as *evidence* (defined as part of the search-inference framework). In thinking about the area of the

[5] The danger here is that the interpretation will be too closely associated with a single model. The students described earlier who could not apply the formula to a parallelogram turned on its side might have been suffering from such an overly narrow description of the design, rather than from a more complete failure of understanding.

parallelogram, for example, a student might simultaneously search for possibilities (possible formulas — perhaps ones already known that might be applied); evidence (properties of the figure that might provide a clue); and goals (actually subgoals, assuming that our student does not question the utility of the basic task itself). She might discover the formula for the rectangle (a possibility); the evidence that the parallelogram is visually similar to a rectangle (which would increase the strength of this possibility — for some students, perhaps increasing it enough that search would stop here); and the subgoal of making the parallelogram into a rectangle. This subgoal would in turn initiate a new process of thinking, inside, as it were, the main process.

By this account, a student understands the formula $A = b \cdot h$ when he has learned the following facts:

- The *purpose* is to find the area.
- The *design* is to use the rectangle formula, replacing one side of the parallelogram with the height.
- One *argument* is based on the subgoal (subpurpose) of making the parallelogram into a rectangle.
- The *design* for this subpurpose is to move the triangle from one side to the other.
- An *argument* for this subdesign is that when we do this, the area is unchanged.
- Another *argument* is that the base is unchanged and the height is unchanged.

Sometimes the things we are asked to learn are not well supported and can be improved or replaced. The strategy of thinking as we learn is likely to encourage the discovery of such deficiencies. Students who insist on understanding do not learn simply because they are told that such and such is the truth. They resist false dogma. If they are given a questionable generalization to learn (such as the claim that totalitarian states never become democratic without violence), they will think of their own counterexamples (such as Chile).

Perkins's analysis of understanding in terms of purposes, designs (and models), and arguments has several implications. First, this theory calls attention to the relationship between understanding and inference. Understanding involves knowing what is a good argument. We must have certain standards for this, and these standards are likely to change as we become more educated. A young child often accepts an argument if it is merely consistent with the possibility being argued. An older child, in accepting the argument, often also insists that the argument be *more* consistent with this possibility than with some other possibility. It appears that understanding must be renewed as we become more sophisticated about arguments. There seem to have been no child prodigies in philosophy. This domain, by its nature, insists on the highest standards of evidence and therefore cannot be understood at all in an immature way.

Perkins's theory, like Wertheimer's, is also a remedy for a common misconception about the nature of understanding that is exemplified in the work of Ausubel (1963) and many others. The essence of Ausubel's theory is the idea that new knowledge becomes meaningful when relationships are established between new knowledge and old. Although Ausubel specifies that relationships must not be arbitrary in order for learning to count as meaningful, he fails to define "arbitrary relationships" with sufficient clarity to rule out mnemonically learned relationships (that is, relationships learned through special memorization techniques) that might hinder the acquisition of true understanding in Wertheimer's sense. Relationships of the sort indicated by Ausubel, for example, can just as easily be used to learn a falsehood (such as a formula stating that the area of a parallelogram is the sum of the lengths of the sides) as to learn the correct formula. Ausubel omits any consideration of purpose or of evidence that a given element serves a purpose. These latter restrictions, as we have seen, seem to be required to account for Wertheimer's demonstrations — particularly his argument that the process of learning with understanding resists the learning of falsehoods.

Conclusion

Thinking can help us to make decisions that achieve our personal goals, to adopt beliefs about which courses of action are most effective, and to adopt goals that are most consistent with our other goals (including the general goal of being satisfied with our lives). In the rest of this book, we shall be concerned primarily with the properties of thinking that make it useful for these purposes. Like any goal-directed activity, thinking can be done well or badly. Thinking that is done well is thinking of the sort that achieves its goals. When we criticize people's thinking, we are trying to help them achieve their own goals. When we try to think well, it is because we want to achieve our goals.

Thinking leads to understanding, which is the best way to improve naive theories, but no guarantee that further improvement is impossible. The best defense against baloney is to ask about the purpose and the arguments. The next chapter will discuss how such understanding can prevent errors.

Chapter 2

The study of thinking

Descriptive, normative, and prescriptive

Here is a problem: "All families with six children in a city were surveyed. In seventy-two families, the *exact order* of births of boys (*B*) and girls (*G*) was G B G B B G. What is your estimate of the number of families surveyed in which the *exact order* of births was B G B B B B?"

Many people give figures less than seventy-two as their answers, even if they believe that boys and girls are equally likely (Kahneman and Tversky, 1972). Apparently they feel that the second sequence, which contains only one girl, is not typical of the sequences they expect. In fact, if you believe that boys and girls are equally likely, your best guess should be exactly seventy two. This is because the probability of each sequence is $1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2$ or $1/64$, the same in both cases. In other words, the two sequences are equally likely. The births are independent: The probability of each one is independent of what came before. In this case, multiplication of probabilities yields the probability that the particular sequence will happen. This multiplication rule is a design, which we can understand in terms of its purpose, which we can think of, for now, as corresponding to the correct proportions of possible sequences.

What makes this problem tricky is that the first sequence looks more like the kind of sequence you might expect, because it has an equal number of boys and girls, and the sexes alternate fairly frequently within the sequence.

This problem can help us illustrate three general *models*,[1] or approaches to the study of thinking, which I shall call descriptive models, prescriptive models, and normative models.

[1] The term "model" comes from the idea that one way to understand something is to build a model of it. In this sense, the game of Monopoly is a model of real estate investment. In this book, the term "model" is used loosely to mean "theory" or "proposal." Sometimes, however, the models will be more detailed — for example, computer models or mathematical models.

Descriptive models are theories about how people normally think — for example, how we solve problems in logic or how we make decisions. Many of these models are expressed in the form of *heuristics*, or rules of thumb, that we use in certain situations. One heuristic is the "What if everyone did that?" rule for thinking about moral situations, and another is the use of analogies in making predictions. Other descriptive models are mathematical, describing functional relationships between inputs (such as probabilities) and outputs (such as choices or judgments).

In the probability problem that I just described, the heuristic used is to judge probability by asking, "How similar is this sequence to a typical sequence?" Because the sequence G B G B B G is more similar to the typical sequence than B G B B B B, the former is judged more likely. Like most heuristics, this is a good rule to follow in some situations, and we can "understand" why: When other things are equal, an item that is similar to the members of a category is more likely to be in that category than an item that is not similar to the members. (Here the "category" is "sequences of births.") But that argument does not apply here because the type of similarity that people perceive is irrelevant in this case. In sum, heuristics can lead to errors when they result from incomplete understanding.

Unlike many other fields of psychology, such as the study of perception, where the emphasis is on finding out "how it works," much of the study of thinking is concerned with comparing the way we usually think with some ideal. This difference from other fields is partly a result of the fact that we have a considerable amount of control over how we think. That is not so with perception. Except for going to the eye doctor once in a while, we have very little control over how our visual system works. To answer the question "How do we think?," we also have to answer the question "How do we *choose* to think?" The way we think is, apparently, strongly affected by our culture. Such tools as probability theory, arithmetic, and logic are cultural inventions. So are our attitudes toward knowledge and decision making. Thus, the way we think is a matter of cultural design. To study only how we happen to think in a particular culture, at a particular time in history, is to fail to do justice to the full range of possibilities.

Part of our subject matter is therefore the question of how we *ought* to think. If we know this, we can compare it to the way we *do* think. Then, if we find discrepancies, we can ask how they can be repaired. The way we ought to think, however, is not at all obvious. Thus, we shall have to discuss models or theories of how we ought to think, as well as models of how we do think. Models of how we ought to think will fall, in our discussion, into two categories: prescriptive and normative.

Prescriptive models are simple models that "prescribe" or state how we ought to think. Teachers are highly aware of prescriptive models and try to get their students to conform to them, not just in thinking but also in writing, reading, and mathematics. For example, there are many good prescriptive models of composition in books on style. There may, of course, be more than one "right" way to think (or write). There may also be "good" ways that are not quite the "best." A good teacher encourages students to think (or write) in "better" ways rather than "worse" ones.

Prescriptive models may consist of lists of useful heuristics, or rules of thumb, much like the heuristics that make up many descriptive models. Such heuristics may take the form of "words to the wise" that we try to follow, such as "Make sure each paragraph has a topic sentence" or (in algebra) "Make sure you know what is 'given' and what is 'unknown' before you try to solve a problem." In studying probability, one might learn the general rule "All sequences of equally likely events are equally likely to occur." Knowing this rule would have saved you the effort of calculating the answer to the problem about the families with six children.

To determine which prescriptive models are the most useful, we apply a *normative model*, that is, a standard that defines thinking that is best for achieving the thinker's goals. For probability problems like the one concerning the birth order of boys and girls, the normative model is the theory of probability. By using the theory of probability, we could prove that the rule "All sequences of equally likely events are equally likely to occur" always works.

Normative models evaluate thinking and decision making in terms of the personal goals of the thinker or thinkers. For decision making, the normative model consists of the policy that will, in the long run, achieve these goals to the greatest extent. Such a model takes into account the probability that a given act (for example, leaving my umbrella at home) will bring about a certain outcome (my getting wet) and the relative desirability of that outcome according to the decision maker's personal goals. It is not enough simply to say that the normative model *is* the decision that leads to the best outcome (carrying an umbrella only when it will rain). We need a way of evaluating decisions at the time they are made, so that we can give prescriptive advice to the decision maker who is not clairvoyant.

You might think that the best prescriptive model is always to "try to use the normative model itself to govern your thinking." This is not crazy. Performing musicians often listen to their own playing as if they were an audience listening to someone else, thus applying their best standards to themselves, using such evaluation as feedback. Likewise, in some cases, we can evaluate our own thinking by some normative model. This approach has two problems, though. First, although it may be possible to evaluate a musical performance while listening to it, the application of normative models to thinking and decision making is often time consuming. A normative model of decision making may require calculations of probabilities and desirabilities of various outcomes. (For example, in deciding whether to take an umbrella, I would have to determine the probability of rain and the relative undesirability of carrying an umbrella needlessly or of getting wet.) Because we value time, the attempt to apply normative models by calculating can be self-defeating. If we spend time applying them, we ensure that we will violate them, because the cost of the time itself will undercut the achievement of other goals. For most practical purposes, people can do better by using some simple heuristics or rules of thumb (for example, "When in doubt, carry the umbrella") than by making these calculations. Even if the calculations sometimes yielded a better choice than the choice that the heuristics would yield, the difference between the two choices in desirability usually would be too small to make calculation worthwhile as a general policy.

The second problem with attempting to apply normative models directly is that we sometimes may do better, according to these models, by aiming at something else. For example, we shall see that people tend to be biased toward possibilities they already favor. This kind of bias may affect judgments of probability, which may be part of the attempt to apply a normative model of belief. So it may be necessary to bend over backward to avoid such effects.

In short, normative models tell us how to evaluate judgments and decisions in terms of their departure from an ideal standard. Descriptive models specify what people in a particular culture actually do and how they deviate from the normative models. Prescriptive models are designs or inventions, whose purpose is to bring the results of actual thinking into closer conformity to the normative model. If prescriptive recommendations derived in this way are successful, the study of thinking can help people to become better thinkers.

Methods for empirical research

The development of descriptive models is the business of psychological research. A great variety of methods can help us in this task. Some involve observation of people (or animals) in their usual activities. Other methods involve construction of artificial situations, or experiments.

Observation

When we observe, we collect data but do not intervene, except insofar as necessary to get the data. Sometimes we can get interesting data literally by observing and recording what people do in a natural setting. Keren and Wagenaar (1985), for example, studied gambling behavior by observing blackjack players in an Amsterdam casino over a period of several months, recording every play of every game. Observations of behavior in real-life situations do not encounter the problems that may result from subjects trying to please a researcher; however, there are other problems. Goals in the real world are often complex, and it is difficult to "purify" the situation so as to determine how a subject would pursue a single goal. For example, a subject in a hypothetical gambling experiment can be instructed to imagine that his goal is to win as much as he can, but Keren and Wagenaar (1985) found that in real life gamblers were often as concerned with making the game interesting as with winning.

Process tracing

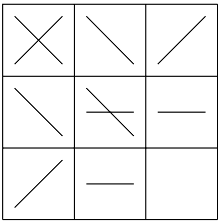
Many methods attempt to describe thinking by tracing the *process* of thinking as it occurs. These methods are not concerned with the subject's conclusion, but with how the conclusion was reached, that is, the steps or "moves" that led to it. Ideally, it would be nice to have a mind-reading machine that displays the subject's thoughts on a television set, in color images and stereophonic sound. Until such a device is invented, we must make do with less direct methods.

One method in this category involves the use of computers and other apparatus to record everything that subjects look at, and for how long, while performing an experiment (e.g., Payne, Bettman, and Johnson, 1988). This method has been used for studying decisions about apartments. The subject is asked to read a table giving data on various apartments. Each column represents an apartment and each row gives figures on matters such as rent, size, and distance from work. If the subject scans across the rent row first and then seeks no other information about the apartments with the highest rent, we can infer that she has eliminated those apartments on the basis of their high rent. To use this method effectively, the experimenter must be clever in setting up the experiment, so that such inferences can be made.

Perhaps the simplest and most direct method for process tracing is to give a subject a task that requires thinking and ask the subject to "think aloud," either while doing the task or as soon afterward as possible. What the subject says is then a *verbal think-aloud protocol*, which a researcher can analyze in many ways. This method has been in almost continuous use since the nineteenth century (Woodworth and Schlosberg, 1954, ch. 26, give some examples).

To get a feeling for this method, try reading the following puzzle problem; then stop reading, and think aloud to yourself as you try to work out the answer. Remember that your task is to do the problem and to say out loud, at the same time, what is going on in your mind, *as it happens*.

Problem: Examine the following three-by-three matrix. Notice that the lower right-hand corner of the matrix is blank. What symbol belongs in that corner?



Here is an example of a verbal protocol in which someone is thinking about this problem. (The different moves are numbered for later reference):

1. Let's see. There's an X, a tilted X, and a bunch of lines — diagonal lines along the top and left side, and horizontal lines in the lower right.

2. It looks like there ought to be another horizontal line in the lower right.

3. That would make a nice pattern.

4. But how can I be sure it's right?

5. Maybe there's a rule.
6. I wonder if the X has something to do with the diagonals. The X is really just the two diagonals put together.
7. That doesn't help me figure out what goes in the lower right.
8. Oh, another idea. Maybe there's two of each thing in each row.
9. Yes, that works. The X is there because there have to be two left diagonals and two right diagonals in the top row. And it works for the columns too.
10. So I guess there has to be a dash and a right diagonal together in the bottom right.

This example is fairly typical. It reveals to us that the thinker is making a search for possibilities, evidence, and goals. Now here is an analysis of the same verbal protocol, describing each step in terms of the search-inference framework:

1. The subject spends some time simply seeking evidence, without any idea about the answer (that is, without being aware of any possibilities for it).
2. A possibility is found for the answer.
3. Evidence is found in favor of the possibility. (Some subjects would stop here, making an error as a result of failing to search further.)
4. Further search for evidence.
5. The subject sets up a subgoal here. The original goal was to say simply what kind of symbol went in the lower right-hand corner. The new goal is to find a rule that will produce the pattern. The search for goals and subgoals — as if the subject said to himself, "Exactly what should I be trying to do?" — is an important, and often neglected, part of thinking.
6. Here is a possibility about the rule, suggested by the evidence that "X is two diagonals put together." It is not a complete possibility, however, for the idea that X has something to do with the diagonals does not say exactly what it has to do with the other diagonals in the matrix. So this possibility, because it is incomplete, sets up a subgoal of making it complete.
7. In this problem, the possibility and the subgoal are put aside because the subject cannot find a way in which they help to satisfy the goal of finding a rule. This failure is a kind of evidence, and the subject uses this evidence to weaken the possibility in question.
8. Another possibility for the rule is found (not unrelated to the first).
9. Evidence for this possibility is sought and found.

10. The subject returns to the original goal of figuring out what goes in the lower right, a task quickly accomplished once the subgoal of finding the rule is achieved.

This analysis shows how the search-inference framework enables us to categorize the moves that a thinker makes in the course of thinking. Notice that a given move can belong to two different categories, because the move may have two different functions; for example, in move 6, the same object is both a possibility and a new subgoal. A given phase in an episode of thinking can contain other episodes of thinking, which can contain others, and so on. For example, the task of searching for the goal might involve trying to understand the instructions, which might involve searching for possibilities and for evidence about the meaning of words such as "matrix" (if one did not know the meaning already). As an exercise, you might find it useful to generate another verbal protocol of your own thinking about some problem and analyze it in this way.

Psychologists have developed a great variety of other methods for analyzing think-aloud protocols. (Ericsson and Simon, 1980, review a number of these.) Different approaches use different *units of analysis*. Some investigators allow the system of analysis itself to define the unit: I did this in the example just given, using as units the categories of the search-inference framework. Other investigators divide the protocol into linguistic units, such as sentences. Others use time measurements, dividing the protocol into 5- or 10-second units and analyzing what is happening (or not happening) in each unit. Approaches also differ in the categories used. The method of analysis is closely linked with the investigator's own goals and the theoretical or practical questions that led to the work.

Despite the extensive use of this method, many doubts have been raised over the years about its adequacy:

1. Some mental processes do not produce much that is accessible for conscious report. Or the processes may go by too quickly for the subject to remember them.

2. The instruction to think aloud may induce subjects to think differently than they ordinarily would. For example, they could think less quickly because of the need to verbalize everything. Verbalization could interfere with thinking, or it could help by forcing thinkers to be more careful. Both of these results have been found, but it has also been found that in many tasks verbalization has no apparent effect (Ericsson and Simon, 1980).

3. Verbal protocols might be misleading with respect to the underlying determinants of the subjects' behavior. For example, suppose that you are deciding whether to buy a used television set, and you say, "The picture is nice, but the sound isn't very good, and $200 is too expensive. I'll keep looking." One might infer from this that you are following a rule that you should not pay more than $200, no matter what. Although this may be true, it may instead be true

that you would be willing to pay more if the sound quality were good enough. You may be trading off quality against price, even though you do not express this in your verbalizations (Einhorn, Kleinmuntz, and Kleinmuntz, 1979).

4. Subjects may be unable to explain how they reached a certain conclusion. For example, in an experiment done by Nisbett and Wilson (1977), passersby in a shopping mall were asked to compare four nightgowns and rate their quality. Most subjects gave the last nightgown they examined the highest rating, regardless of which of the four it was, but all subjects attributed their rating to some property of the nightgown itself rather than to its position in the sequence.

Ericsson and Simon (1980) argue that this last sort of demonstration does not shed any light on the validity of verbal reports, because there is a difference between reporting *what* one is thinking (as in the matrix problem) and explaining the *causes* of one's conclusions. Asking subjects to infer a cause requires that the subjects take the role of scientists, which they may not be able to do. If the subjects simply report their experiences instead of inferring causes, then they cannot be accused of making an error.[2]

Ericsson and Simon argue that verbal reports sometimes provide a quite reliable method for discovering how thinking proceeds. In particular, they assert, "thinking" aloud is useful when there is relevant information to be reported that is in the subject's "working memory" (or immediate consciousness). An example would be the information reported in the matrix task. In such cases, performance is found not to be affected by the requirement to think aloud, and what subjects do is consistent with what they say.

More generally, as Ericsson and Simon point out, verbal reports are just one kind of investigative method. Any method of investigation has defects, and these defects are more serious in some cases than in others. When possible, a good investigator will try to use a variety of methods to check the results from one against the results from another.

Interviews

Sometimes researchers interview subjects extensively. Some interviews are like questionnaires because the interviewer simply reads the questions and the subjects answer them. Interviews give the opportunity to ask follow-up questions or let the subjects explain their answers at greater length. A useful idea is the *structured interview*, which is essentially a set of questions to be answered, with the assumption that the interviewer will ask follow-up questions until the subject gives a satisfactory answer to each question, or gives up. Compared to strict questionnaires, interviews

[2] It is not even clear that the subjects in Nisbett and Wilson's experiments made an error at all. If a subject said she liked the last nightgown she examined because of its texture, she could be correct, even though it is also true that she liked it because it was the last: The fact that it was the last might have caused her to like its texture.

have the advantages of making sure that the subjects understand the questions and answers each as it is intended. They also allow subjects to explain their answers. The disadvantage is that they provide an opportunity for the interviewer to influence the answers.

Use of archival data

Another method for process tracing is the use of historical records of decisions made by groups. Janis (1982), for example, studied group decision making by reading records of how President Kennedy and his advisers made the policy decisions that led to the Bay of Pigs fiasco in 1961. This method is useful when the records are very complete, as they are in this case.

Another application of this is the measure of "integrative complexity" (Schroder, Driver, and Streufert, 1967; Suedfeld and Rank, 1976; Suedfeld and Tetlock, 1977). Using this method, Tetlock measured the complexity of speeches by U.S. senators (1983a) and personal interviews given by members of the British House of Commons (1984). Integrative complexity is scored on a scale of one to seven. The scoring takes two dimensions into account, "differentiation" and "integration," although only the latter dimension is reflected in the name "integrative." In Tetlock's examples (1983a, p. 121), a score of one is given to a statement that expresses only a one-sided view, neglecting obvious arguments on the other side, thus failing to "differentiate" the two sides. For example,

> Abortion is a basic right that should be available to all women. To limit a woman's access to an abortion is an intolerable infringement on her civil liberties. Such an infringement must not be tolerated. To do so would be to threaten the separation of Church and State so fundamental to the American way of life.

A score of three is given when the statement is differentiated — that is, when it includes arguments (evidence or goals) for both sides:

> Many see abortion as a basic civil liberty that should be available to any woman who chooses to exercise this right. Others, however, see abortion as infanticide.

A score of five or higher is given when the person making the argument succeeds in "integrating" opposing arguments, presenting a reflective statement about the criteria by which arguments should be evaluated:

> Some view abortion as a civil liberties issue — that of the woman's right to choose; others view abortion as no more justifiable than murder. Which perspective one takes depends on when one views the organism developing within the mother as a human being.

Tetlock found that moderate leftists got the highest scores, and he interpreted this in terms of the fact that this group was constantly facing issues that put their values in conflict, such as the goals of equality and economic efficiency, which conflict in such questions as whether the rich should be taxed to help the poor (thus reducing economic incentive but increasing equality).

Hypothetical scenarios

Another way to learn how people think or make decisions is to observe the conclusions that people draw or the decisions they make. The investigator then makes inferences from the effects of relevant variables on these responses. (A "variable" is anything that we can measure, label, or manipulate.) Investigators strive both to capture the phenomenon of interest and to control the effective variables, so that they can determine which variables do what.

Sometimes we observe people making real decisions, or what they think are real decisions. Real decisions may involve money that is actually paid at the end of an experiment. Social psychologists often stage realistic deceptions, which the subjects think are real. Most of the studies of decision making described in this book ask subjects what they would do in hypothetical situations. The disadvantage of hypothetical questions is that the results may not tell us much about what people would actually do in the real world. For example, in answering the hypothetical question, people may tell us that they would do what they think *we* (the researchers) would *want* them to do — not what they would really do. This is called a "social desirability" effect. (This is not a serious problem if the experimenter is *interested* in the subjects' views about what is the best decision.)

An advantage of using hypothetical situations for a study is that the researchers can extensively manipulate the situation to find out what variables are affecting the subject's responses. Another advantage is that the experimenter can easily ask subjects for justifications or explanations. Justifications and explanations can suggest new hypotheses for study, provide evidence bearing on other hypotheses, and provide evidence that subjects understand (or fail to understand) the situation as the experimenter intended. Hypothetical situations are also useful in telling us how subjects would respond to situations that are novel to them, and situations that are difficult to stage in the laboratory. Finally, hypothetical decisions may be just as useful as real ones for finding out how people think about certain types of problems.

Most of the research described in this book uses hypothetical decisions. The conclusions of such research are strengthened if we can point to cases in the real world — often cases involving public policy choices — that seem to correspond to the hypothetical cases used in the laboratory. We cannot be sure that the apparent biases in real cases are real biases. Other factors may influence people's judgments and decisions aside from those that are present in the lab. But the combination of real cases and experimental evidence makes a case that may be sufficiently compelling to arouse us to try to do something, especially when the apparent bias makes things worse.

Individual differences

Differences among people are of interest for many reasons. If some people show a bias and others do not, we can look for both causes and effects of these differences. For example, some of the difficulties that people have with decisions about risks, or present/future conflict, are correlated with measures of “cognitive reflection,” as measured by the ability to solve trick problems such as: “A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?” (Frederick, 2005).

Similarly, Peters and her colleagues (2006) assessed numeracy with questions such as: “The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?” Subjects who gave more correct answers to questions like this made better decisions and judgments when numbers were involved. For example, they were more likely to bet on drawing a red jelly bean from a bowl when the bowl contained 1 red bean out of 10 than when it contained 9 out of 100. Together, these results suggest that some biases can be overcome by types of quantitative thinking that are within the capacity of a fair number of people. (See also p. 212.)

More generally, studies of individual differences in biases have repeatedly found that many people do not show biases. A “bias” is a departure from the normative model in a particular direction. Thus, if half of the subjects show a bias and the other half show no bias in either direction, the average subject will show a bias. With enough subjects, the overall average bias will be statistically reliable. It is important to determine whether biases are universal or not. Some have compared judgment biases to optical illusions, which enlighten us about how our visual system works by showing how it systematically fails in some cases. This analogy is a poor one if only half of the people show a bias. Everyone with normal vision sees most visual illusions. Visual illusions may be hard wired into our nervous systems in ways that judgment biases are not.

Another reason for studying individual differences is that some people may show the reverse of the usual bias. If these people are sufficiently rare, then the usual bias will still be found on the average, but, really, there would be two biases going in different directions.

Cultural differences can also enlighten us about the range of human possibility. Individual differences may result from education, subcultural differences, or genetic differences. They may affect people’s success or failure in achieving various goals. Research on differences among people is based on correlations, that is, observation of differences rather than experimental manipulation. Because of this, it is often difficult to determine what the cause of some difference is, or whether it is responsible for some effect. For example, sex differences could result from genetic or environmental causes. Still, such research is often useful.

Training and debiasing

One way to test a prescriptive theory about thinking is to try to improve it through instruction and then show some effect of the instruction on something else. This avoids the problem of looking for correlations between good outcomes and good thinking. Such correlations can result when the outcomes and the thinking are both influenced by the same factors, such as general mental ability. Training studies can use a control group, which is not trained, so that any later differences between the trained group and the control group can be ascribed to the training. Another advantage of these studies is that it often leads to a promising educational technique. These studies are also called "debiasing" because they attempt to reduce a bias.

Transfer of learning is the effect of learning in one situation on learning or behavior in a very different situation. If we are teaching *thinking*, transfer is essential. Because thinking (as defined in this book) is what we do *when we do not know what to choose, desire, or believe*, thinking will be most essential in situations that we have not encountered before. Can teaching of thinking transfer? A number of studies suggest that it can.

Some studies have examined the effects of certain courses on reasoning about judgments and decisions. Schoemaker (1979) found that students who had taken a statistics course gave more consistent answers to questions involving choices of gambles. Students who had not taken the course were also more likely to bid more than the maximum amount that could be won in order to play a gamble in which money could be won, and more likely to require more money than the maximum loss in order to play a gamble in which money could be lost.

Fong, Krantz, and Nisbett (1986, experiment 4) found that statistical training transfers to solving everyday problems involving statistical reasoning. Half of the men enrolled in a college statistics course were interviewed by telephone at the beginning of the course, and the other half were interviewed at the end. The interview (conducted by a woman experimenter) ostensibly concerned sports and began with questions concerning sports controversies (such as what colleges should do about recruiting violations) in order to hide the fact that the basic concern was with statistics. Then subjects were asked such questions as why the winner of the Rookie of the Year award in baseball usually does not do as well in his second year as in his first. A nonstatistical response might be "because he's resting on his laurels; he's not trying as hard in his second year." A statistical response (based on the principle of regression to the mean in numerical prediction explained in Chapter 15) would be "A player's performance varies from year to year. Sometimes you have good years and sometimes you have bad years. The player who won the Rookie of the Year award had an exceptional year. He'll probably do better than average his second year, but not as well as he did when he was a rookie." Students gave more good statistical answers of this sort at the end of the course than at the beginning. Therefore, these students did transfer what they had learned to cases where it is relevant.

Nisbett, Fong, Lehman, and Cheng (1987) carried out other studies in which they examined the effects of various kinds of university instruction on statistical, logical, and methodological reasoning. The researchers measured statistical reasoning by asking subjects to suggest explanations of certain facts, such as that the Rookie of the Year typically does not do as well in his second year or that "a traveling saleswoman is typically disappointed on repeat visits to a restaurant where she experienced a truly outstanding meal on her first visit" (p. 629). Each fact could be explained either in terms of nonstatistical factors ("Her expectations were so high that the food couldn't live up to them") or statistical ones ("Very few restaurants have only excellent meals, odds are she was just lucky the first time"). Subjects were given credit here for mentioning the statistical explanations as possibilities (which they were, whether the other explanations were also true or not).

To measure methodological reasoning, the researchers asked subjects to comment on flawed arguments. One item (from a newspaper editorial) argued for the learning of Latin and Greek on the grounds that students who had studied these languages in high school received much higher than average scores on the Verbal Scholastic Aptitude Test.[3] Another item concerned a claim that the mayor of Indianapolis should fire his police chief because crime had increased since the chief began his tenure in office.[4] Logical reasoning was measured with problems of the sort to be discussed in Chapter 4.

Nisbett and his colleagues found that logical reasoning, as measured by their test, did not improve from the beginning to the end of two introductory college courses in logic (one emphasizing formal logic, the other emphasizing informal fallacies). It did not improve as a result of two years of graduate school in chemistry, either. Logical reasoning did improve, however, as a result of two years of graduate school in law, medicine, or psychology. The researchers suggest that these three fields emphasize logical reasoning to some extent. This improvement was found in two different kinds of studies, one in which students just beginning graduate school were compared to students who had just completed their second year, and the other in which the same students were tested at the beginning and the end of their first two years of graduate training.

Statistical and methodological reasoning showed a somewhat different pattern. The largest improvement by far occurred among psychology students, probably because training in such things as the use of control groups is an important part of graduate work in psychology. Methodological reasoning also improved with medical training, which places some emphasis on the conduct and interpretation of research, but training in law or chemistry had no effect.

Lehman and Nisbett (1990) found that undergraduate training can have similar effects. Courses in the social sciences affected mostly statistical and methodological reasoning. Courses in the natural sciences (including mathematics) and humanities affected mostly logical reasoning. These studies together provide further evidence

[3] Subjects were scored as giving correct responses if they pointed out that students who studied Latin and Greek in high school were unusually competent and would probably do well on the test even without studying these languages.

[4] Subjects were scored as correct if they pointed out the potential relevance of crime-rate increases in other cities over the same period.

that appropriate education, in which certain methods of reasoning are explicitly emphasized, can have general effects on the tendency to use these methods in everyday problems unrelated to the areas in which the methods were taught. They also indicate that some of these effects can be specific to certain methods.

Training studies are not perfect either. Some aspect of the training may be effective, other than the aspect intended. Really both correlational and experimental (training) studies are useful, and to some extent they make up for each others' flaws.

Experimental economics

Much of the research described in this book is closely related to research in economics. Economists develop normative models, and some of those models are the same as those considered here. Traditionally, economists have assumed that people follow normative models, or, at least, that those who use economic theory would do well to assume that people are rational. (If you assume that people are irrational, rational people may find a way to take advantage of your assumptions.) Some economists, however, have tried to test economic theory in the laboratory, using methods much like those of psychology. They are called experimental economists. (See Kagel and Roth, 1995, for an overview.)

In this book, I make no distinction between the work of experimental economists and the work of other researchers. Economic experiments are relevant to many issues discussed here. But the work does have a kind of characteristic approach. First, economists tend to be suspicious of verbal reports or judgments that have no consequences for the subjects. They tend to look at choices, and they provide real consequences to subjects in the form of payoffs, typically money.

More importantly, experimental economics derives its hypotheses from economic theory, and it assumes that this theory is meant to be universally descriptive of human behavior. The theory should therefore apply in the laboratory as well as in real markets. This is an ambitious assumption, but one that is worthy of our attention. If economic theory is universally true in this way, it could help explain other social phenomena, such as politics, crime, and sexual behavior. It is thus important to find out how the simplest assumptions of economic theory must be modified in the light of human psychology.

Another characteristic of much work in experimental economics is that it often assumes that the simplest normative model is one in which people pursue their self-interest rationally. This self-interest assumption is not a necessary part of economic theory, just a simple model that is often useful. The result is that economists are just as interested in what they call "other-regarding preferences" as they are in biases. For example, if you give $5 anonymously to another subject in an experiment because, by chance, you got $10 and she got nothing, you are sacrificing self-interest for fairness. Again, this is not a "bias," but it does contradict the simple model that people pursue rational self-interest only.