

Chapter 3

Good thinking: The nature of rationality

In the case of any person whose judgment is really deserving of confidence, how has it become so? Because he has kept his mind open to criticism of his opinions and conduct. Because it has been his practice to listen to all that could be said against him; to profit by as much of it as was just, and expound to himself . . . the fallacy of what was fallacious.

John Stuart Mill (1859)

A normative view

The purpose of this chapter is to introduce a coherent normative model for thinking. I shall design the model to reflect what we want it to do: help us think in the way that best achieves our personal goals. I shall then make some assumptions about where thinking most often departs from this normative model, descriptively. These assumptions lead to a general prescriptive theory of thinking.

The best kind of thinking, which we shall call *rational thinking*, is whatever kind of thinking best helps people achieve their goals. If it should turn out that following the rules of formal logic leads to eternal happiness, then it is “rational thinking” to follow the laws of logic (assuming that we all want eternal happiness). If it should turn out, on the other hand, that carefully violating the laws of logic at every turn leads to eternal happiness, then it is these violations that we shall call “rational.” When I argue that certain kinds of thinking are “most rational,” I mean that these help people fulfill their goals. Such arguments could be wrong. If so, some other sort of thinking is most rational.

Using the search-inference framework of Chapter 1, let us consider how our best thinking is done. Take decision making. We do not achieve our goals best if we

neglect some of them as we are evaluating the “possibilities.” If, for example, in choosing a college course to take as an elective, I forget to consider the difficulty of the course, I may choose a course that is too hard. Other things being equal, the more goals I consider, the better my decision is likely to be.¹ (Of course, other things usually are not equal. Time may be short, and it takes time to search for goals. Let us come back to that later.)

Likewise, we do not achieve our goals best if we neglect possibilities that might achieve them better than the one we adopt. Even if Political Science 101 would serve my goals best, I cannot choose it if I do not think of it. Good thinking requires a thorough search for possibilities — other things being equal.

For the same reason, we must search thoroughly for evidence. The more I find out about the college courses I am considering, the more likely I am to pick one that does in fact satisfy my goals. We must also seek evidence in a way that is most helpful in finding the possibility that best achieves our goals. We must not seek evidence for any other reason. In particular, we should not seek evidence because we know it will turn out to favor possibilities that are already strong in our minds. If I seek evidence only about the good qualities of a course I am thinking of taking — because I want my initial hunch to be right, for example — I will miss the evidence about its bad qualities.

We must use evidence in a way that best achieves our goals, not in any other way. Again, we should not use evidence in a way that favors possibilities that are already strong, just because we want them to be the ones ultimately chosen. We should be willing to change if the evidence points that way.

There is a problem with the idea that more search is always better. The search for possibilities, evidence, and goals takes time and effort. If our search is lengthy, eventually a point may be reached at which the effort is not worthwhile, in terms of achieving our goals. One of our goals is not to spend our whole life lost in thought. Sometimes, when a quick decision is required, our goal is to think quickly. If registration ends tomorrow, I must pick a course now, making do with the possibilities and evidence at hand. Thus, we must balance the benefit of thorough search against the cost of search itself. Ordinarily, search is most useful at the beginning of thinking, and there is a “point of diminishing returns” beyond which search is no longer useful (Baron, Badglio, and Gaskins, 1986).

The real danger is not in thinking too little, then, but rather in behaving as though we had great confidence in conclusions that were reached with little thought. We should not make momentous decisions on the basis of unexamined beliefs, and we should not express strong confidence in hasty conclusions that fly in the face of conclusions reached by others who have thought much more.

In sum, good decision making involves *sufficient search* for possibilities, evidence, and goals, and *fairness* in the search for evidence and in the use of evidence (Baron, 1985b, 1991). Search is “sufficient” when it best serves the thinker’s per-

¹In the case of thinking about beliefs, the goal is usually fixed (as described in Chapter 1), so search for goals is less important.

sonal goals, including the goal of minimizing the cost of thinking. Search and inference are “fair” when they are not influenced by factors other than the goals of the thinking itself. Good decision making also requires use of the best methods of inference. We can think about these methods themselves.

This normative model of thinking is not very helpful as a practical, prescriptive model. In order to arrive at a prescriptive model, we ought to find out where people depart from the normative model. Then we can give practical advice to correct these departures, as well as whatever other advice will help people do their best thinking. In Chapter 9, I shall argue that people typically depart from this model by favoring possibilities that are already strong. We must thus make an effort to counteract this bias by looking actively for other possibilities and evidence on the other side. I call this *actively open-minded thinking*.

One comment before going on. It is possible to reflect on one’s goals in life. Therefore, what achieves one’s goals as they *are* may not achieve them as they *will be* after one thinks about them. We can still call the thinking good, though, if it takes into account a person’s goals as they are.

Another thing we should note is that rational thinking can be defined relative to a person at a given time, with a given set of beliefs and goals. People may think rationally on the basis of irrationally formed beliefs. For example, if I believe (a delusion) that the Mafia is pursuing me, I might still make rational decisions for coping with that situation. Similarly, people who pursue irrational goals may do poor thinking about their goals, but, given their goals, they may still do good thinking about how to achieve them. If my goal is to escape from the Mafia, I may pursue it well or badly.

Rationality

Rationality concerns the methods of thinking we use, not the conclusions of our thinking. Rational methods are those that are generally best in achieving the thinker’s goals. It is true that when we say someone is “irrational,” we usually disagree with this person’s conclusion — but we disagree in a particular way. We think that better methods ought to have been used in reaching that conclusion. When we call someone “irrational,” we are giving advice (to this person or anyone else who listens) about how he ought to have thought.

The meaning of rationality

Rationality is, therefore, not the same as accuracy, and irrationality is not the same as error. We can use good methods and reach erroneous conclusions, or we can use poor methods and be lucky, getting a correct answer. There are even cases — such as thinking about one’s life goals — where, although there is no reasonable standard of “correctness,” we can still speak of rationality and irrationality.

Rationality is a matter of degree. It makes sense to say that one way of thinking is “more rational” or “less rational” than another. Also, there may be no single “best” way of thinking. There may be several ways of thinking that are indistinguishable in terms of their value in helping people achieve their goals, but still better than many other ways of thinking.

A useful theory of rational thinking, such as the one I have begun to outline here, ought to provide advice to people that they can follow. It does no good to try to teach people by saying, “Be correct!” or “Make good decisions!” That is like telling investors to buy low and sell high. This advice only tells people what their goals are, not how to achieve them. An appropriate response to such advice is, “Gee, thanks.”

We can speak of the rationality of social institutions and of whole societies, as well as the rationality of individuals. Again, the criterion is whether these groups make collective decisions in ways that achieve the goals of their members, taken together. Such social decisions are affected both by the practices of institutions and cultures — the way they are organized for decision making — and by the decisions of individuals within those groups. We can therefore judge the rationality of individual decision making from the group’s point of view as well as from the point of view of the individual’s own goals. The judgments made from these two points of view need not always agree.

Rationality and luck

In discussing rational decision making, we must distinguish between good decision making and good outcomes. A *good decision* is one that makes effective use of the information available to the decision maker at the time the decision is made. A *good outcome* is one that the decision maker likes. Such an outcome can result from a good decision, but it can also result from good fortune, following a bad decision. Of course, the whole point of good thinking is to increase the probability of good outcomes (and true conclusions), but many other factors affect outcomes aside from good thinking. Some of these have to do with good thinking on earlier occasions. Others have to do with luck — factors beyond the person’s control (hence, beyond any general advice we could give).

If we want to promote good decision making, we should ensure that people do the best they can with what is knowable. We cannot insist on clairvoyance. Prudently made investment decisions can lead to surprising losses. A decision to perform surgery could have been a rational one, even if the patient is the one in a thousand who dies on the operating table from that operation. In offering advice or instruction on good decision making, it does not do much good to say, “Do whatever achieves a good outcome.” When we think a decision was badly made, we try to learn some lesson from it for our future decisions. If we think a decision was well made but turned out badly, there is no lesson to be learned. In such a case, we may need to make special efforts to emphasize the quality of the decision, lest the unfortunate outcome dissuade us from our good decision-making practices.

Similarly, when we judge how well a decision was made, we must bear in mind the possibility that well-made decisions turn out badly. Of course, when a decision turns out well, it is more likely to have been well made than if it turned out badly. (We do not appoint commissions of inquiry to study the causes of good outcomes.) Still, people do tend to judge decisions by their outcomes even when they know everything that the decision maker knew, as though decision makers were held responsible for their luck (Baron and Hershey, 1988).

These comments apply to belief formation, as well as to decision making. In general, good thinking leads us to true beliefs, but it can mislead us, and poor thinkers can hit on the truth by chance.

Objections to rationality

The definition of rationality as “the kind of thinking that helps us achieve our goals” answers a number of objections to the concept of rationality as a guide and dispenses with some caricatures of the concept.

First, rationality is not the same as cold calculation of self-interest. Many people think of rationality as exemplified by Dr. Strangelove (in the Stanley Kubrick film), whose single-minded devotion to winning a nuclear war enabled him to think quite coolly about the annihilation of most of the rest of humanity.

Rational thinking need not be cold. Emotions, in fact, are one type of evidence. A bad feeling about a choice is a reason not to make it — although not an overriding reason. Often, bad feelings are signals that some more tangible reason will reveal itself with further search. Even when the more tangible reason does not reveal itself, it may be rational to give uneasy feelings some weight, for the reason may still be there, even though we do not know what it is.

The seeking of pleasant emotions and the avoidance of unpleasant ones are surely goals that most people have (and would want to have, after thinking about their goals). Because these goals are things we want, we often think about how to achieve them.

Moreover, rationality need not be self-interested. Moral goals, including concern for the feelings of others, surely are among the goals we have and ought to have. More generally, rationality does not need to be single-minded; single-mindedness corresponds to the failure to search for more than one goal. The political leader who worries only about maximizing the Gross National Product is not the one who is rational. The rational leader is, rather, the one who worries about such things as people’s feelings, the satisfaction of their desires, in all their variety (and in the long run).

Nor is rational thinking the same as thinking too much. When people really think rationally, the amount of thinking that they do is appropriate to the situation, insofar as possible. Rational thinkers, we have said, are moderate.

A more serious objection to rationality, some claim, is that it stands in the way of commitment, which is sometimes necessary. Apparently, President Richard Nixon of the United States believed that the most effective way of preventing a first strike

by the Soviet Union was to maintain a powerful nuclear arsenal. Nixon appeared to be a madman, who was crazy enough to respond to a nuclear attack even at the risk of destroying civilization. The best way to appear to be mad in this way is to be mad in fact.

Another way to state this objection more generally is to argue that the ultimate objectives of rational decision making (decisions that best serve our goals) are not achieved by *trying* to make rational decisions but by trying to achieve some other goal — such as patriotism — which, by itself, is not necessarily always rational. In some cases, that is, rational thinking is *self-defeating* (Parfit, 1984). Like trying to “be spontaneous,” trying to be rational may ensure that we cannot succeed in doing so. This objection does not undercut the idea of rationality as such. It does, however, imply that people might be better off not *knowing* about rationality and not trying to achieve it.

This approach, in my view, endangers the survival of rationality in a society. If rational thought is useful at all, then it must be maintained as a practice. Parents must teach it to their children, teachers must teach it to their students, and people must respect each other for their rationality. If the practice of rational thought is not to be lost, some group of people, at least, will have to maintain it. If that group is not to be an entire society, then it will have to be some sort of elite that perpetuates itself from generation to generation. This is not a foolish idea. It has been tried before in history, and it is still being tried. It is clearly inconsistent, however, with the ideal of an open society, in which all are given the tools and the opportunity to participate in decisions that affect them. Suppression of the teaching of rationality can therefore interfere with the existence of an open society itself. Without an elite that makes all important decisions, there is no way to ensure that people will make decisions that serve their own goals except by teaching them to think rationally.

A final objection to the concept of rationality is the claim that rational thinkers cannot be happy. By this argument, happiness requires a certain amount of self-deception. If one questions one's beliefs too closely, one may discover that one is not as successful, competent, or well liked as one thought. This is consistent with Alloy and Abramson's finding (1979; discussed in Chapter 7) that depressives correctly perceive their lack of control. Perhaps if we all correctly perceived the world, we would all be depressed.

It is true that some of us maintain an overly rosy view of ourselves through a kind of irrationality, in which we ignore the evidence against our rosy views. We convince ourselves that everything is just dandy, without asking whether it could be better. Many people may live their entire lives this way, happy as clams. If rational thinking were defined as whatever led to happiness, we might well have to change our view of what rational thinking is. Instead of respect for evidence, neglect of evidence might turn out to be rational.

In my view, happiness does not require such irrationality. Often, the happiness that results from irrationally formed beliefs goes along with irrationally formed goals. For example, people who think that they are universally liked often have the goal of being liked by everyone. Although it is surely rational to want to be liked, it

is, for most people, hopeless to try to be liked by *everyone*. A balanced, rational view of how things actually are needs to be combined with a balanced, realistic view of how they ought to be, if we are not to be disappointed. If one's goals are as rationally formed as one's beliefs about how well one's goals are being achieved, accurate beliefs need not be disappointing. If I desire to be liked by *most* of the people I meet, I probably will not need to deceive myself in order to convince myself that this goal is being achieved reasonably well.²

Rationally formed beliefs have other advantages. On reflection, the combination of accurate beliefs and realistic goals may be more desirable than the combination of irrational beliefs and unrealistic goals, even though both combinations are capable of making us happy for the moment (and perhaps for longer, if we are lucky).

Rationality and emotion

Rationality is often contrasted with emotion. If we think of “emotion” as a way of making decisions *without thinking*, then this contrast is reasonable. Sometimes it pays to think. We have already seen, however, that emotions enter into thinking itself in a variety of ways. In particular, we noted that emotions can serve as evidence and that the creation or avoidance of certain emotions can serve as goals of behavior, and therefore as goals of thinking. Let us look a little further into the relation between emotion and rationality.

What is emotion? Roughly, we can take an emotion to be “a state that is subjectively experienced as pleasant or unpleasant, that drives or motivates certain kinds of behavior specific to the emotion, and that tends to be elicited by a certain kind of situation.” Anger, for example, is usually unpleasant. In extreme anger, our muscles tighten and our hearts pound; we are more inclined to strike out; we want to hurt certain people or to see them hurt. Anger is typically induced by what we consider unfair treatment (of ourselves or of others). Fear, while also unpleasant, is induced by danger. Fear can increase our belief that harm will occur, and it can reduce our tendency to adopt risky options. Other emotions are elation, sadness, embarrassment, pride, regret, and rejoicing. Some emotions are related to moral behavior in particular, such as guilt feelings, anger, and empathic sadness or joy. The situations that induce emotions, as well as their effects on behavior, can differ from person to person. Likewise, people may use the same term for an emotion in different ways. (Sabini, 1992, has an interesting discussion of emotions.)

Much of our behavior seems to be designed to let us feel desirable emotions. The way we do this is often indirect. The first parachute jump, researchers tell us, evokes terror followed by relief. After a few jumps, however, the terror decreases, and the relief becomes euphoria that may last for days (Solomon & Corbit, 1974). A single extremely positive experience (such as winning the state lottery), on the other hand, can *reduce* the capacity to experience future pleasures, and vice versa

²Practically all of the “irrational beliefs” mentioned by Ellis (for example, 1987) as causes of psychological disturbance take the form of goals that are impossible to achieve.

(Brickman, Coates, and Janoff-Bulman, 1978). Direct attempts to induce pleasant emotions, then, are sometimes self-defeating. Many desirable emotions are essentially by-products of actions taken for other reasons (Elster, 1983).

Although emotions can serve as goals, they are certainly not the *only* goals we strive to achieve in our behavior. If you find this hard to believe, consider the fact that many people strive for goals that will not be reached until after their own deaths and thus cannot possibly give them any future emotional experience. They put money aside for their children's inheritance, or they work for long-term causes in their old age, perhaps even knowing that their death is near. It may even be that the desire for emotional experiences plays a very unimportant role in the major goals of most people. Some Buddhists systematically strive to eliminate emotions, on the grounds that, on the whole, they are just not worth it (Kolm, 1986).

Emotions are to some extent unavoidable, but they are also partly under our control. Many actors can induce or suppress emotions in themselves almost on cue. Some people try to reshape their character — often with the help of therapists — so that their emotional responses change. Moreover, emotions often have undesired effects; for example, teachers who get angry at their students may fail to teach well, as a result. Emotions can also have desired effects, as when the emotions of athletes make them try harder.

Are emotions rational? When emotions are *not* under our control, this question makes no more sense than asking whether the knee-jerk reflex is rational. Even if we decide that we do not like this reflex, there is nothing we can do about it, so the question is empty. If emotions are *partly* under our control, however, we can at least think about whether we should try to control them. (We could want them to be stronger or weaker.) The decision about whether or not to try may be made well or badly, like any decision. In thinking about whether to try to control our emotions, we must consider the cost of the effort, which may be substantial. It could be better to live with a slightly mixed-up emotional system than to spend years in therapy trying to fix it. On the other hand, therapy for some kinds of undesirable emotional responses, such as phobias, could be well worth the effort involved (see Beck, 1976).

Emotions may help us achieve our goals in the long run, even when they seem to prevent us from achieving our goals in the short run (Frank, 1988). For example, suppose that you are in an experiment called the "ultimatum game," in which another subject is instructed to make you an offer of some part of \$20. You know that if you accept the offer, then she gives you the amount offered and keeps the rest of the \$20 for herself. If you refuse the offer, then neither of you gets any money. Now suppose that she offers \$1. Would you accept? Most people would not (Thaler, 1988), even though it would be in their short-run interest to do so. (They would gain a dollar.) Very likely, an offer of only \$1 out of \$20 makes people angry, and this causes them to reject it. Now if the other subject knows that you are the sort of person who will hurt yourself in order to hurt someone who has made you an unfair offer, she will make you better offers in the future (e.g., \$10). So a short-run loss can ensure long-run gains. Displaying your anger at an unfair offer (if only by rejecting it) is a way to maintain this kind of reputation. If people see that you are in the grip of an emotion,

they will know that you mean it.

Because of the long-run advantages, people with such emotions may have reproduced more in the past, so emotions may have been maintained by natural selection (Frank, 1988). Although emotions can have this sort of benefit, and may well have had it while the human species was evolving, the same emotions can lead to harm in other cases, as when anger — combined with biased judgment of what is fair — prevents people from negotiating an agreement to end a conflict.

Our knowledge of our emotions can become part of our thinking itself. For example (Chapter 11), in thinking about risky choices such as buying stock, we could take into account the regret we would feel if the value of the stock were to go down after we bought it. We could think of this emotion of regret as a risk we take *in addition* to the financial loss itself. If we know that we usually cannot control this emotion but are bothered by it, our unwillingness to feel so much regret could give us a good reason not to buy the stock, even if we were willing to take the risk on financial grounds alone. On the other hand, we may know that we can control this emotion. If the stock goes down, some people are able to avoid the feeling of regret by telling themselves (truthfully, perhaps) that the risk was worth taking, even though the venture did not work out. Control of emotion, therefore, can be a rational choice in its own right.

The emotions we have — and those we expect to have — can be influenced by our beliefs in a variety of respects. For example, if you come to believe that smoking is morally wrong, you will get less pleasure from smoking, you will feel more guilty from smoking, and you will (after you quit) be more likely to get angry at smokers. You will also anticipate these emotions, and that anticipation will affect your decisions. For example, in order to avoid anger, you will not go to a restaurant that has no section for nonsmokers. Sometimes these effects of anticipated emotions can become disabling: Fear of having a panic attack keeps some people from going out in public (Chambless and Gracely, 1989), even when the panic itself is caused by clearly exaggerated fears of heart attacks or other unlikely misfortunes (Cox, Swinson, Norton, and Kuch, 1991).

Changing people's beliefs can change both their emotions and their anticipation of these emotions. Thus, a woman's panic attacks may perhaps be reduced by convincing her that her heart palpitations do not (as she thought) indicate an incipient heart attack (Salkovskis and Clark, 1991). Likewise, when a children's vaccine can cause death from side effects that rarely occur, some people say they would decide not to vaccinate a child because they fear the guilt feelings that would result if their child died from the vaccine, even if the vaccine reduced the overall risk of death. These people would change their decision if they were convinced that vaccination was the morally right thing to do under these conditions. They would then worry more about the guilt feelings that would result if their unvaccinated child died from the disease (Baron, 1992).

Can emotions *make* us think irrationally? Janis and Mann (1977) present a number of cases in which this seems to occur. They show, for example, that people experiencing fear often do not think effectively about how they can deal with the

real danger that causes the fear. It may be misleading, though, to call such an effect “irrational.” Once the emotion is present, its effect on thinking could be unavoidable. If there is nothing to be done about the effect, it is empty to call it irrational. If there is any irrationality, it may be in failing to control the emotion itself, or in failing to shape one’s character so that panic does not easily occur.

In sum, the relation between rational thinking and emotion is more complex than a simple contrast between the two, once we stop using the term “emotion” as a substitute for the word “irrationality.” We need to think about emotions in the psychological sense in which I have been using the term, as states with certain causes and effects. Emotions — in this sense — are often the goals of our decisions. They also affect our decisions in ways that are sometimes desired and sometimes undesired. They may help us in the long run, even if they hurt us in the short run. They may make us behave more morally, or less morally, than we would without them. They can impair thinking itself, but, in moderation, they can also help it.

Rationality and belief

When we form beliefs, we generally have the goal of believing what is true. We therefore look for beliefs that fit the evidence we have. When we have time to think thoroughly and openly, we look for evidence against beliefs we are considering — that is, evidence that they are not true — and we seek alternative beliefs. Chapter 9 argues that such thinking — which is “actively open-minded” — is a good prescriptive method to counteract the biases that favor pre-existing beliefs.

Rational belief formation

In general, then, actively open-minded thinking is most likely to lead to true beliefs. In addition, when we cannot be sure that a belief is true, good thinking will ensure that our *confidence* in the belief is in proportion to the evidence available. Appropriate confidence is, in most cases, a more realistic goal than certainty.

The main advantage of true beliefs, or beliefs that we hold with appropriate confidence, is that they allow us to make better decisions, decisions more likely to achieve our goals. This is illustrated most clearly in the discussion of probability and utility theory in Chapters 5 and 10, where we shall see that coherent and consistent probability judgments are the ones most likely to give us good results. The same point may be made more generally, even when numerical probabilities are not at issue. If our confidence depends appropriately on the evidence we have, we will take the calculated risks that we ought to take, and when action requires certainty, we shall hold back if we cannot be certain.

There may well be other reasons to have rationally formed beliefs. We could have the goal of pursuing truth “for its own sake,” regardless of the help it gives us in making decisions. There is surely nothing irrational about having such a goal.

Self-deception

Although our goal in belief formation is usually to believe the truth (or to have appropriate confidence), sometimes it would seem better to believe what is false. It might therefore be more rational sometimes to think in a way that leads to false beliefs. This amounts to self-deception. Although self-deception can at times be best, at other times it lies behind the most insidious forms of irrationality, as when people convince themselves that some idea of theirs is right, despite the evidence against it.

What is self-deception? To some, the idea implies that we really have two selves — the “deceiver” (perhaps the unconscious), who knows the truth, and the “deceived.” The deceiver must have some reason to carry out his deception. For example, he might want the deceived to feel that she has been right all along. (The deceived might find changing her mind to be painful, and the deceiver might be sympathetic.)

Although this idea of a dual self has its appeal, it is not needed to explain self-deception. All I need to do to deceive myself is to do something in order to control my belief, without being aware that I have done it. For example, the philosopher Pascal argued (see Chapter 10) that one ought to try to believe in God, since if God exists and one does not believe, one might be damned, and this is too big a risk to take. Pascal felt that someone who understood this argument could voluntarily become a believer by honestly trying to live the Christian life. In doing this, one would eventually, through studying the Bible and associating with other Christians, come to believe in Christianity and in the Christian God. Eventually one would very likely forget that the whole thing was inspired by the ulterior motive of avoiding eternal damnation.

All that is necessary for self-deception, then, is that when we form our belief we do not take account of the fact that self-deception has occurred. If we do, it will not work. If we keep in our minds the fact that we began to go to church only because we were afraid of hell, we will not be so easily persuaded by what we hear there.

Our beliefs are manipulated more frequently than Pascal’s rather extreme example suggests. If you go to law school and become a lawyer, you will very likely come to believe that lawyers are good people who serve a valuable function in society (even if you go to law school for some other reason, without believing this at the outset). Similarly, if you have a child, you will very likely come to believe in the frivolity of those who voluntarily remain childless. Any course of life you choose, in other words, is likely to affect your beliefs. If you want to control your beliefs, then, you can do so by choosing your course of life for that purpose. On a more mundane level, some people set their clocks 5 minutes fast, in order to get to work on time — hoping to deceive themselves, if only for a panicky moment, into thinking that time is short, so they will get off to work quickly.

Can self-deception ever be rational? On the one hand, self-deception seems to be one of the major means we have for maintaining (at times with great confidence) false and harmful beliefs. If we want to believe that smoking is harmless, for example, we can make ourselves believe this by seeking evidence in favor of our belief and

ignoring evidence against it. We must be sure not to take fully into account the fact that we have done these things, for, if we do, we will see that the evidence we use was as good as useless, “cooked” to order. It might as well have been made up.

This kind of biased search can become so much a matter of habit for some people that they do not know that they behave this way. Perhaps they formed the habit unconsciously, because it was more comfortable. If they never questioned their beliefs, they never had to suffer the pain of changing them.

It is difficult to know how many of our beliefs are maintained in this way. If you wonder about a particular belief, try to think about it in an actively open-minded way over a period of time. This is the only cure for self-deception, and it is a cure that has few undesirable side-effects, even for those who did not really have the disease.

Self-deception is thus clearly irrational, in some cases. It is almost necessarily a part of poor thinking. If people *know* that their thinking is poor, they will not believe its results. One of the purposes of a book like this is to make recognition of poor thinking more widespread, so that it will no longer be such a convenient means of self-deception.

On the other hand, in certain cases we can be reasonably sure that the benefits of self-deception outweigh the costs:

1. We can sometimes manipulate our own behavior to our benefit through self-deception. A simple case is one already described, setting one's clock ahead in order to get to work on time. Similarly, the behavior of liars may be more convincing if they believe the lies they tell. If they can make themselves believe their lies, they will be more effective. Actors, of course, deliberately try to deceive themselves in order to act more convincingly. Loyal spouses may maintain their belief that their spouse is the best one for them by never “experimenting” with other possibilities.
2. Beliefs can affect goals, particularly the strength of the goals; therefore, we may deceive ourselves in order to control our goals. Athletes may convince themselves that they are likely to win, in order to make themselves undergo the rigors of training or take the risks that they must take in playing. The potential cost here is that the effort and risk will be futile and hopeless, but again, the benefits might outweigh this. On the other hand, some people convince themselves that their goals are unattainable in order to avoid the anguish of trying to attain them (Elster, 1983).
3. Beliefs themselves can make us happy or unhappy, and sometimes the beliefs that make us unhappy have no compensating advantages. It can be reasonable not to want to be told that one is dying from an incurable disease, especially if one would only continue to live one's life as best one could. Workers who are subject to occupational hazards sometimes convince themselves that the risks are minor (Akerlof and Dickens, 1982), avoiding the stress of worrying about them. In general, it seems reasonable not to want to know bad things that we cannot do anything about.

From a normative point of view, whether self-deception should be attempted in a given case depends on the balance of costs and benefits in that case. From a prescriptive point of view, it may be wise to try to seek the truth as a general rule. First, self-deception is often unnecessary. There are other ways of getting to work on time besides setting the clock ahead. Marital fidelity and bliss do not necessarily require the belief that one's spouse was the best possible choice.

Second, self-deception can have harmful effects. For every athlete who can win by “psyching herself up” for an important match, there are countless other average athletes who convince themselves that they will make it to the Olympics and waste years trying. (They may then tell themselves it was all somehow worthwhile, thus continuing the deception.) Excessive self-esteem can result in optimistic self-deception that leads, in turn, to excessive risk taking. In a video-game experiment with monetary prizes, subjects with high self-esteem generally performed better and chose targets that they were capable of achieving. In an “ego threat” condition, however, all subjects were told, “Now if you are worried that you might choke [lose your nerve] under pressure or if you don't think you have what it takes to beat the target, then you might want to play it safe and just go for two dollars.” In this condition, subjects with high self-esteem tried for the larger prize and, in the end, made less money than those with lower self-esteem (Baumeister, Heatherton, and Tice, 1993).

Finally, if we get the idea that it is OK to deceive ourselves, we may well overdo it, because cases in which self-deception is irrational can be hard to recognize. When we overdo it, we are prevented from knowing the truth even when we want to. A person who has been ill but who has said to his doctor, “If it's really bad, don't tell me,” may not be able to be truly reassured if his doctor tells him that he will indeed recover. In extreme cases, habitual self-deceivers may wake up one day in terror, not knowing which of their beliefs are real. Those who set out on the path of self-deception should proceed with caution.

Beliefs as a cause of desires

Beliefs can produce a “sour grapes” effect. When something seems impossible or difficult for us to attain, we *want* it less. Sometimes this is rational: If we can adjust our desires to reality, we will be happier. Changing our belief is irrational, though, when we are too easily persuaded to give up a goal that we could achieve (Elster, 1983).

Whether this effect is rational or not, it seems to occur. Harris (1969) asked subjects to rate the desirability of a number of phonograph records, one of which they would be given at the end of the experiment. Then subjects were told that a certain record would not be given away, and the ratings were done again. The excluded record was given lower ratings than it had been given when it was a possible goal. We might call this the “Pangloss effect,” after Dr. Pangloss, the “sage” in Voltaire's *Candide*, who, after each tragic episode in the story, explains at great length why it was “all for the best.”

Are people ever really irrational?

Much of the research described in this book involves attempts to show that people do not follow normative or prescriptive models — that is, they do not think in the best way. We are tempted, when reading about such studies, to come to the defense of the researchers' subjects. After all, calling someone "irrational" is not nice, especially if it is a false accusation.

Sometimes we attempt to defend these subjects by arguing that the behavior in question is functional — that it serves some purpose other than "rationality." If people are illogical, for example, perhaps it is because they "want" to be and feel better when they behave this way, or because illogic leads them to a "deeper" truth.

There are indeed situations in which people can be rational while appearing to be irrational. The theory of rationality that is used to judge their thinking may be wrong, or some important goal (such as the subject's emotions) may have been neglected in applying the theory of rationality.

On the other hand, we cannot assume that people always have good reasons for appearing to be irrational. People can really be irrational sometimes. In this book I shall assume that the "burden of proof" is not on one side or the other. This is because I shall also assume that our main interest is in helping people to think better (or to maintain those aspects of good thinking that they already use). Therefore, the two kinds of mistakes that we can make — deciding falsely that others are or are not irrational — are both costly ones. If we falsely conclude that people are *irrational* in some way, we may waste our effort in trying to help them — and we may even make them worse. If we falsely conclude that people are rational when they are not, we lose an opportunity to help them. (Of course, to paraphrase Pogo, "them" is "us.")

For every argument of the form "If we're so irrational, how did we ever get to the moon?", there is another argument of the form "If we're so rational, how come we [pick your favorite complaint about the world situation or about people]?" It is, in a way, from our point of view, a cause for optimism to discover biases and irrationalities, for we can teach people to avoid them. If the errors of humanity — collective and individual — cannot be prevented, our future is precarious.

To conclude, the purpose of the research discussed in this book is not to give grades to the human race or to Western culture. History gives the grades. Our job is to try to figure out *why* the human race is getting C's rather than A's — and whether anything can be done about it.

Conclusion

The idea of rationality presented in this chapter, and developed in the rest of the book, is not an arbitrary standard that some dictator of the mind is trying to impose on the world. It is a standard that we would all want to meet because we want to achieve our own goals. If you think I am wrong, you must argue that the standards I propose do not help us achieve our goals. All of this follows from the idea that the purpose of thinking is to achieve our goals.

Chapter 4

Logic

Nothing is better than eternal happiness.

A ham sandwich is better than nothing.

Therefore, a ham sandwich is better than eternal happiness.

Nickerson (1986)

Past writers (for example, Arnauld, 1662/1964) have taken formal logic as a normative model of thinking. Today, people sometimes use the word "logical" as if it simply meant "reasonable" or "rational." Logic has influenced education — where it served as the basis for the teaching of thinking for centuries — and it has provided us with much of our language for talking about thinking: "premise," "assumption," "contradiction."

Today logic has lost its central place as the normative model for thinking. But the development of logic and its psychology provide an example of the approach emphasized here, which compares human reasoning to normative models. Because logic is so old, it is not often challenged as a normative model. Errors in logical reasoning are generally accepted as errors or biases. Later I will discuss models that are not (yet) so well accepted. Claims about bias may seem more questionable. All claims, of course, are questionable, but we must remember the example of logic before we conclude that the whole enterprise is so tenuous as to be meaningless.

What is logic?

Matthew Lipman has written a series of philosophical "novels" to introduce children in elementary and secondary schools to philosophical thinking. In one of these, *Harry Stottlemeier's Discovery* (1974; say the name aloud and think of a famous Greek philosopher), Harry is daydreaming in class. Suddenly he hears the teacher asking him whether Halley's comet is a planet, and he struggles to come up with an