Reason in Human Affairs

The Harry Camp Lectures at
Stanford University, 1982

The Harry Camp Memorial Fund was established in 1959 to make possible a continuing series of lectures at Stanford University on topics bearing on the dignity and worth of the human individual.

HERBERT A. SIMON

Reason in Human Affairs

STANFORD UNIVERSITY PRESS 1983
STANFORD, CALIFORNIA

Stanford University Press
Stanford, California

Printed in the United States of America
ISBN 0-8047-1179-8
LC 82-62448

To the Memory of
JASCHA MARSCHAK
who had an unshakeable faith in
human reason, and an unmatched
store of human warmth

Preface

THE NATURE of human reason—its mechanisms, its effects, and its consequences for the human condition—has been my central preoccupation for nearly fifty years. When the invitation came to deliver the Harry Camp Lectures at Stanford University, I wondered if I had anything left to say on the subject. And if there were some such topic, had it not already been thoroughly investigated by such friends on the Stanford campus as Kenneth Arrow, James March, and Amos Tversky—to mention just a few who work in one part or another of this domain? Putting aside this concern, real though it is, I decided to use the occasion of the lectures to explore some byways that seemed to me interesting and important, but that had until now been off the main paths of my own explorations.

Three topics, especially, were the objects of my inquiry: the relation of reason to intuition and emotion, the analogy between rational adaptation and evolution, and the implications of bounded rationality for the operation of social and political institutions. In the chapters that follow, I report on these topics within the framework provided by the general viewpoint of bounded rationality.

I am indebted to Stanford University for the occasion

and opportunity to prepare these pages, and for the hospitality and stimulation I always enjoy on my visits to the Stanford campus. I am grateful, too, to Donald T. Campbell, Richard C. Lewontin, and Edward O. Wilson, who provided valuable criticisms of a draft of Chapter 2, though it is not to be assumed that they would agree with everything in the final version of that chapter. To them, and to many friends who have helped guide my education on evolutionary theory and on other topics addressed in these pages, I offer my warm thanks.

H.A.S.

Contents

Reason in Human Affairs

1. Alternative Visions of Rationality

One kind of optimism, or supposed optimism, argues that if we think hard enough, are rational enough, we can solve all our problems. The eighteenth century, the Age of Reason, was supposed to have been imbued with this kind of optimism. Whether it actually was or not I will leave to historians; certainly the hopes we hold out for reason in our world today are much more modest.

It is my purpose in these pages to explore from a contemporary standpoint the uses and limits of reason in human affairs. In order to avoid the kind of unwarranted optimism I have just mentioned, my first two chapters will be addressed more to the limitations of reason than to its uses. I will try to redress the balance in my third chapter, but as I develop my topic I think you will see why I have taken up the limitations first. Only as we understand those limitations can we devise procedures to use effectively the powers that human reasoning capabilities do give us.

In the first chapter, I will focus initially on the very powerful formal models of rationality that have been constructed in this century and that must be counted among the jewels of intellectual accomplishment in our time. Since these models are well known, I will describe them

only briefly, devoting most of my discussion to showing why, in application to real human affairs, they deliver somewhat less than they appear to promise. But my intent here is not mainly critical. The last half of the chapter will develop a more realistic description of human bounded rationality, and will consider to what extent the limited capability for analysis that is provided by bounded rationality can meet the needs for reason in human affairs.

In the second chapter, I will discuss the thesis, nowadays often associated with the discipline of sociobiology, that the deficiencies of reason will be corrected, for better or for worse, by the sterner rationality of natural selection. Two questions will be of particular concern in that discussion: first, whether and to what degree altruism can survive in a system subjected to the forces of natural selection, and, second, to what extent selection processes resemble optimization processes.

In the light of the conclusions reached in these two chapters, I will turn in the third to the question of how reason can be employed effectively in human social affairs.

In science one is supposed to deliver new truths. The most crushing verdict that can be pronounced on a scientific paper is the fabled referee's report, scribbled in the margin, "What's new here is not true, and what's true is not new." But these pages are not intended as reports of scientific discoveries and will not seek novelty. I will be satisfied if what I have to say is mainly true, even though it will not be at all new. As I shall argue in my discussion of human rationality, attention needs to be called periodically to important old truths.

At the same time, I do not wish simply to repeat here

things that I have said at length in my previous books, and especially in *Administrative Behavior* and *The Sciences of the Artificial*, both of which are deeply concerned with the concept of human rationality. In the former, I examined the implications of the limits of human rationality for organizational behavior. In the latter, I described properties that are common to all adaptive ("artificial") systems, giving us a basis for constructing a general theory of such systems. In the present volume I have drawn on this previous work to the extent necessary to provide a framework for my discussion. But within this framework I have concentrated on topics that remain problematic or controversial and that are of critical importance for understanding the role of rationality in human affairs. I have already indicated briefly what some of those problematic topics are.

THE LIMITS OF REASON

Modern descendants of Archimedes are still looking for the fulcrum on which they can rest the lever that is to move the whole world. In the domain of reasoning, the difficulty in finding a fulcrum resides in the truism "no conclusions without premises." Reasoning processes take symbolic inputs and deliver symbolic outputs. The initial inputs are axioms, themselves not derived by logic but simply induced from empirical observations, or even more simply posited. Moreover, the processes that produce the transformations of inputs to outputs (rules of inference) are also introduced by fiat and are not the products of reason. Axioms and inference rules together constitute the fulcrum on which the lever of reasoning rests; but the particular structure of that fulcrum cannot be justified by the

methods of reasoning. For an attempt at such a justification would involve us in an infinite regress of logics, each as arbitrary in its foundations as the preceding one.

This ineradicable element of arbitrariness—this Original Sin that corrupts the reasoning process, and therefore also its products—has two important consequences for our topic here. First, it puts forever beyond reach an unassailable principle of induction that would allow us to infer infallible general laws, without risk of error, from specific facts, even from myriads of them. No number of viewings of white swans can guarantee that a black one will not be seen next. Whether even a definite probability statement can be made about the color of the next swan is a matter of debate, with the negatives, I think, outnumbering the affirmatives.

Further, the foundations of these inductions—the facts—rest on a complex and sometimes unsteady base of observation, perception, and inference. Facts, especially in science, are usually gathered in with instruments that are themselves permeated with theoretical assumptions. No microscope without at least a primitive theory of light and optics; no human verbal protocols without a theory of short-term memory. Hence the fallibility of reasoning is guaranteed both by the impossibility of generating unassailable general propositions from particular facts, and by the tentative and theory-infected character of the facts themselves.

Second, the principle of "no conclusions without premises" puts forever beyond reach normative statements (statements containing an essential *should*) whose derivation is independent of inputs that also contain *should*'s. None of the rules of inference that have gained acceptance

are capable of generating normative outputs purely from descriptive inputs.[1] The corollary to "no conclusions without premises" is "no *ought*'s from *is*'s alone." Thus, whereas reason may provide powerful help in finding means to reach our ends, it has little to say about the ends themselves.

There is a final difficulty, first pointed out by Gödel, that rich systems of logic are never complete—there always exist true theorems that cannot be reached as outputs by applying the legal transformations to the inputs. Since the problem of logical incompleteness is much less important in the application of reason to human affairs than the difficulties that concern us here, I shall not discuss it further. Nor will I be concerned with whether the standard axioms of logic and the rules of inference themselves are to some extent arbitrary. For the purpose of this discussion, I shall regard them as unexceptionable.

Reason, then, goes to work only after it has been supplied with a suitable set of inputs, or premises. If reason is to be applied to discovering and choosing courses of action, then those inputs include, at the least, a set of *should*'s, or values to be achieved, and a set of *is*'s, or facts about the world in which the action is to be taken. Any attempt to justify these *should*'s and *is*'s by logic will simply lead to a regress to new *should*'s and *is*'s that are similarly postulated.

VALUES

We see that reason is wholly instrumental. It cannot tell us where to go; at best it can tell us how to get there. It is a

[1] I will not undertake to make the argument here. It was stated well many years ago by Ayer, in *Language, Truth, and Logic*, rev. ed. (New York, 1946), chap. 6.

gun for hire that can be employed in the service of whatever goals we have, good or bad. It makes a great difference in our view of the human condition whether we attribute our difficulties to evil or to ignorance and irrationality—to the baseness of goals or to our not knowing how to reach them.

Method in Madness

A useful, if outrageous, exercise for sharpening one's thinking about the limited usefulness of reasoning, taken in isolation, is to attempt to read Hitler's *Mein Kampf* analytically—as though preparing for a debate. The exercise is likely to be painful, but is revealing about how facts, values, and emotions interact in our thinking about human affairs. I pick this particular example because the reader's critical faculties are unlikely, in this case, to be dulled by agreement with the views expressed.

Most of us would take exception to many of Hitler's "facts," especially his analysis of the causes of Europe's economic difficulties, and most of all his allegations that Jews and Marxists (whom he also mistakenly found indistinguishable) were at the root of them. However, if we were to suspend disbelief for a moment and accept his "facts" as true, much of the Nazi program would be quite consistent with goals of security for the German nation or even of welfare for the German people. Up to this point, the unacceptability of that program to us is not a matter of evil goals—no one would object to concern for the welfare of the German people—or of faulty reasoning from those goals, but rests on the unacceptability of the factual postulates that connect the goals to the program. From this viewpoint, we might decide that the remedy for Nazism

was to combat its program by reason resting on better factual premises.

But somehow that calm response does not seem to match the outrage that *Mein Kampf* produces in us. There must be something more to our rejection of its argument, and obviously there is. Its stated goals are, to put it mildly, incomplete. Statements of human goals usually distinguish between a "we" for whom the goals are shaped and a "they" whose welfare is not "our" primary concern. Hitler's "we" was the German people—the definition of "we" being again based on some dubious "facts" about a genetic difference between Aryan and non-Aryan peoples. Leaving aside this fantasy of Nordic purity, most of us would still define "we" differently from Hitler. Our "we" might be Americans instead of Germans, or, if we had reached a twenty-first-century state of enlightenment, our "we" might even be the human species. In either case, we would be involved in a genuine value conflict with *Mein Kampf*, a conflict not resolvable in any obvious way by improvements in either facts or reasoning. Our postulation of a "we"—of the boundary of our concern for others—is a basic assumption about what is good and what is evil.

Probably the greatest sense of outrage that *Mein Kampf* generates stems from the sharpness of the boundary Hitler draws between "we" and "they." Not only does he give priority to "we," but he argues that any treatment of "they," however violent, is justifiable if it advances the goals of "we." Even if Hitler's general goals and "facts" were accepted, most of us would still object to the measures he proposes to inflict on "they" in order to nurture the welfare of "we." If, in our system of values, we do not

regard "they" as being without rights, reason will disclose to us a conflict of values—a conflict between our value of helping "we" and our general goal of not inflicting harm on "they." And so it is not its reasoning for which we must fault *Mein Kampf*, but its alleged facts and its outrageous values.

There is another lesson to be learned from *Mein Kampf*. We cannot read many lines of it before detecting that Hitler's reasoning is not cold reasoning but hot reasoning. We have long since learned that when a position is declaimed with passion and invective, there is special need to examine carefully both its premises and its inferences. We have learned this, but we do not always practice it. Regrettably, it is precisely when the passion and invective resonate with our own inner feelings that we forget the warning and become uncritical readers or listeners.

Hitler was an effective rhetorician for Germans precisely because his passion and invectives resonated with beliefs and values already present in many German hearts. The heat of his rhetoric rendered his readers incapable of applying the rules of reason and evidence to his arguments. Nor was it only Germans who resonated to the facts and values he proclaimed. The latent anti-Semitism and overt anti-Communism of many Western statesmen made a number of his arguments plausible to them.

And so we learned, by bitter experience and against our first quick judgments, that we could not dismiss Hitler as a madman, for there was method in his madness. His prose met standards of reason neither higher nor lower than we are accustomed to encountering in writing designed to persuade. Reason was not, could not have been, our principal shield against Nazism. Our principal shield was contrary factual beliefs and values.

De Gustibus Est Disputandum

Recognizing all these complications in the use of reason, hot or cold, and recognizing also that *ought*'s cannot be derived from *is*'s alone, we must still admit that it is possible to reason about conduct. For most of the *ought*'s we profess are not ultimate standards of conduct but only subgoals, adopted as means to other goals. For example, taken in isolation a goal like "live within your income" may sound unassailable. Yet a student might be well advised to go into debt in order to complete his or her education. A debt incurred as an investment in future productivity is different from a gambling debt.

Values can indeed be disputed (1) if satisfying them has consequences, present or future, for other values, (2) if they are acquired values, or (3) if they are instrumental to more final values. But although there has been widespread consensus about the rules of reasoning that apply to factual matters, it has proved far more difficult over the centuries to reach agreement about the rules that should govern reasoning about interrelated values. Several varieties of modal logic proposed for reasoning about imperative and deontic statements have gained little acceptance and even less application outside of philosophy.[2]

In the past half century, however, an impressive body of formal theory has been erected by mathematical statisti-

[2] I state the case against modal logics in Section 3 of my *Models of Discovery* (Dordrecht, 1977) and in "On Reasoning about Actions," chap. 8 of H. A. Simon and L. Siklóssy, eds., *Representation and Meaning* (Englewood Cliffs, N.J., 1972).

cians and economists to help us reason about these matters—without introducing a new kind of logic. The basic idea of this theory is to load all values into a single function, the utility function, in this way finessing the question of how different values are to be compared. The comparison has in effect already been made when it is assumed that a utility has been assigned to each particular state of affairs.

This formal theory is called subjective expected utility (SEU) theory. Its construction is one of the impressive intellectual achievements of the first half of the twentieth century. It is an elegant machine for applying reason to problems of choice. Our next task is to examine it, and to make some judgments about its validity and limitations.

SUBJECTIVE EXPECTED UTILITY

Since a number of comprehensive and rigorous accounts of SEU theory are available in the literature,[3] I will give here only a brief heuristic survey of its main components.

The Theory

First, the theory assumes that a decision maker has a well-defined *utility function*, and hence that he can assign a cardinal number as a measure of his liking of any particular scenario of events over the future. Second, it assumes that the decision maker is confronted with a well-defined *set of alternatives* to choose from. These alternatives need not be one-time choices, but may involve sequences of choices or strategies in which each subchoice will be made only at a specified time using the information available at that time.

[3] For example, L. J. Savage's classic, *The Foundations of Statistics* (New York, 1954).

Third, it assumes that the decision maker can assign a consistent *joint probability distribution* to all future sets of events. Finally, it assumes that the decision maker will (or should) choose the alternative, or the strategy, that will *maximize the expected value*, in terms of his utility function, of the set of events consequent on the strategy. With each strategy, then, is associated a probability distribution of future scenarios that can be used to weight the utilities of those scenarios.

These are the four principal components of the SEU model: a cardinal utility function, an exhaustive set of alternative strategies, a probability distribution of scenarios for the future associated with each strategy, and a policy of maximizing expected utility.

Problems with the Theory

Conceptually, the SEU model is a beautiful object deserving a prominent place in Plato's heaven of ideas. But vast difficulties make it impossible to employ it in any literal way in making actual human decisions. I have said so much about these difficulties at other times and places (particularly in the pages of *Administrative Behavior*) that I will make only the briefest mention of them here.

The SEU model assumes that the decision maker contemplates, in one comprehensive view, everything that lies before him. He understands the range of alternative choices open to him, not only at the moment but over the whole panorama of the future. He understands the consequences of each of the available choice strategies, at least up to the point of being able to assign a joint probability distribution to future states of the world. He has reconciled or balanced all his conflicting partial values and syn-

thesized them into a single utility function that orders, by his preference for them, all these future states of the world.

The SEU model finesses completely the origins of the values that enter into the utility function; they are simply there, already organized to express consistent preferences among all alternative futures that may be presented for choice. The SEU model finesses just as completely the processes for ascertaining the facts of the present and future states of the world. At best, the model tells us how to reason about fact and value premises; it says nothing about where they come from.

When these assumptions are stated explicitly, it becomes obvious that SEU theory has never been applied, and never can be applied—with or without the largest computers—in the real world. Yet one encounters many purported applications in mathematical economics, statistics, and management science. Examined more closely, these applications retain the formal structure of SEU theory, but substitute for the incredible decision problem postulated in that theory either a highly abstracted problem in a world simplified to a few equations and variables, with the utility function and the joint probability distributions of events assumed to be already provided, or a microproblem referring to some tiny, carefully defined and bounded situation carved out of a larger real-world reality.

SEU as an Approximation

Since I have had occasion to use SEU theory in some of my own research in management science, let me throw the stone through my own window. Holt, Modigliani, Muth, and I constructed a procedure for making decisions about

production levels, inventories, and work force in a factory under conditions of uncertainty.[4] The procedure fits the SEU model. The utility function is (the negative of) a cost function, comprising costs of production, costs of changing the level of production, putative costs of lost orders, and inventory holding costs. The utility function is assumed to be quadratic in the independent variables, an assumption made because it is absolutely essential if the mathematics and computation are to be manageable. Expected values for sales in each future period are assumed to be known. (The same assumption of the quadratic utility function fortunately makes knowledge of the complete probability distributions irrelevant.) The factory is assumed to have a single homogeneous product, or a set of products that can legitimately be represented by a single-dimensional aggregate.

It is clear that if this decision procedure is used to make decisions for a factory, that is very different from employing SEU theory to make decisions in the real world. All but one of the hard questions have been answered in advance by the assumption of a known, quadratic criterion function and known expected values of future sales. Moreover, this single set of production decisions has been carved out of the entire array of decisions that management has to make, and it has been assumed to be describable in a fashion that is completely independent of information about those other decisions or about any other aspect of the real world.

I have no urge to apologize for our decision procedure

[4] C. C. Holt, F. Modigliani, J. R. Muth, and H. A. Simon, *Planning Production, Inventories and Work Force* (Englewood Cliffs, N.J., 1960).

as a useful management science tool. It can be, and has been, applied to this practical decision task in a number of factory situations and seems to have operated satisfactorily. What I wish to emphasize is that it is applied to a highly simplified representation of a tiny fragment of the real-world situation, and that the goodness of the decisions it will produce depends much more on the adequacy of the approximating assumptions and the data supporting them than it does on the computation of a maximizing value according to the prescribed SEU decision rule. Hence, it would be perfectly conceivable for someone to contrive a quite different decision procedure, outside the framework of SEU theory, that would produce better decisions in these situations (measured by real-world consequences) than would be produced by our decision rule.

Exactly the same comments can be made about economic models formed within the SEU mold. Their veridicality and usefulness cannot be judged from the fact that they satisfy, formally, the SEU assumptions. In evaluating them, it is critical to know how close the postulated utilities and future events match those of the real world.

Once we accept the fact that, in any actual application, the SEU rule supplies only a crude approximation to an abstraction, an outcome that may or may not provide satisfactory solutions to the real-world problems, then we are free to ask what other decision procedures, unrelated to SEU, might also provide satisfactory outcomes. In particular, we are free to ask what procedures human beings actually use in their decision making and what relation those actual procedures bear to the SEU theory.

I hope I have persuaded you that, in typical real-world

situations, decision makers, no matter how badly they want to do so, simply cannot apply the SEU model. If doubt still remains on this point, it can be dissipated by examining the results of laboratory experiments in which human subjects have been asked to make decisions involving risk and uncertainty in game-like situations orders of magnitude simpler than the game of real life. The evidence, much of which has been assembled in several articles by Amos Tversky and his colleagues, leaves no doubt whatever that the human behavior in these choice situations—for whatever reasons—departs widely from the prescriptions of SEU theory.[5] Of course, I have already suggested what the principal reason is for this departure. It is that human beings have neither the facts nor the consistent structure of values nor the reasoning power at their disposal that would be required, even in these relatively simple situations, to apply SEU principles.

As our next task, we consider what they do instead.

THE BEHAVIORAL ALTERNATIVE

I will ask you to introspect a bit about how you actually make decisions, and I will make some assertions that you can check against your introspections. First, your decisions are not comprehensive choices over large areas of your life, but are generally concerned with rather specific matters, assumed, whether correctly or not, to be relatively independent of other, perhaps equally important, dimensions of life. At the moment you are buying a car, you are probably not also simultaneously choosing next

[5] See A. Tversky and D. Kahnemann, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185: 1124–31 (1974), and references cited there.

week's dinner menu, or even deciding how to invest income you plan to save.

Second, when you make any particular decision, even an important one, you probably do not work out detailed scenarios of the future, complete with probability distributions, conditional on the alternative you choose. You have a general picture of your life-style and prospects, and perhaps of one or two major contemplated changes in the near future, and even of a couple of contingencies. When you are considering buying a car, you have a general notion of your use of automobiles, your income and the other demands on it, and whether you are thinking of getting a new job in another city. You are unlikely to envision large numbers of other possibilities that might affect what kind of car it makes sense to buy.

Third, the very fact that you are thinking about buying a car, and not a house, will probably focus your attention on some aspects of your life and some of your values to the relative neglect of others. The mere contemplation of buying a car may stimulate fond memories or dreams of travel, and divert your attention from the pleasures of listening to stereo or giving dinner parties for friends at home. Hence, it is unlikely that a single comprehensive utility function will watch over the whole range of decisions you make. On the contrary, particular decision domains will evoke particular values, and great inconsistencies in choice may result from fluctuating attention. We all know that if we want to diet, we should resist exposing ourselves to tempting food. That would be neither necessary nor useful if our choices were actually guided by a single comprehensive and consistent utility function.

Fourth, a large part of whatever effort you devote to making your car-buying decision will be absorbed in gathering facts and evoking possibly relevant values. You may read *Consumer Reports* and consult friends; you may visit car dealers in order to learn more about the various alternatives, and to learn more about your own tastes as well. Once facts of this sort have been assembled, and preferences evoked, the actual choice may take very little time.

Bounded Rationality

Choices made in the general way I have just been describing are sometimes characterized as instances of *bounded rationality*. Good reasons can be given for supposing that evolutionary processes might produce creatures capable of bounded rationality. Moreover, a great deal of psychological research supports the hunch to which our introspections have led us, namely that this is the way in which human decisions—even the most deliberate—are made. Let us call this model of human choice the behavioral model, to contrast it with the Olympian model of SEU theory.

Within the behavioral model of bounded rationality, one doesn't have to make choices that are infinitely deep in time, that encompass the whole range of human values, and in which each problem is interconnected with all the other problems in the world. In actual fact, the environment in which we live, in which all creatures live, is an environment that is nearly factorable into separate problems. Sometimes you're hungry, sometimes you're sleepy, sometimes you're cold. Fortunately, you're not often all three at the same time. Or if you are, all but one of these

needs can be postponed until the most pressing is taken care of. You have lots of other needs, too, but these also do not all impinge on you at once.

We live in what might be called a nearly empty world—one in which there are millions of variables that in principle could affect each other but that most of the time don't. In gravitational theory everything is pulling at everything else, but some things pull harder than others, either because they're bigger or because they're closer. Perhaps there is actually a very dense network of interconnections in the world, but in most of the situations we face we can detect only a modest number of variables or considerations that dominate.

If this factorability is not wholly descriptive of the world we live in today—and I will express some reservations about that—it certainly describes the world in which human rationality evolved: the world of the cavemen's ancestors, and of the cavemen themselves. In that world, very little was happening most of the time, but periodically action had to be taken to deal with hunger, or to flee danger, or to secure protection against the coming winter. Rationality could focus on dealing with one or a few problems at a time, with the expectation that when other problems arose there would be time to deal with those too.[6]

Mechanisms for Bounded Rationality

What characteristics does an organism need to enable it to exercise a sensible kind of bounded reality? It needs

[6]A simple formal model of such rationality is provided by my "Rational Choice and the Structure of the Environment," *Psychological Review* 63: 129–38 (1956).

some way of focusing attention—of avoiding distraction (or at least too much distraction) and focusing on the things that need attention at a given time. A very strong case can be made, and has been made by physiological psychologists, that focusing attention is one of the principal functions of the processes we call emotions. One thing an emotion can do for and to you is to distract you from your current focus of thought, and to call your attention to something else that presumably needs attention right now. Most of the time in our society we don't have to be out looking for food, but every so often we need to be reminded that food is necessary. So we possess some mechanisms that arouse periodically the feeling of hunger, to direct our attention to the need for food. A similar account can be given of other emotions.

Some of an organism's requirements call for continuous activity. People need to have air—access to it can be interrupted only for a short time—and their blood must circulate continually to all parts of their bodies. Of course, human physiology takes care of these and other short-term insistent needs in parallel with the long-term needs. We do not have to have our attention directed to a lack of oxygen in our bloodstream in order to take a breath, or for our heart to beat. But by and large, with respect to those needs that are intermittent, that aren't constantly with us, we operate very much as serial, one-at-a-time, animals. One such need is about as many as our minds can handle at one time. Our ability to get away with that limitation, and to survive in spite of our seriality, depends on the mechanisms, particularly emotional mechanisms, that assure new problems of high urgency a high priority on the agenda.

Second, we need a mechanism capable of generating alternatives. A large part of our problem solving consists in the search for good alternatives, or for improvements in alternatives that we already know. In the past 25 years, research in cognitive psychology and artificial intelligence has taught us a lot about how alternatives are generated. I have given a description of some of the mechanisms in Chapters 3 and 4 of *The Sciences of the Artificial.*[7]

Third, we need a capability for acquiring facts about the environment in which we find ourselves, and a modest capability for drawing inferences from these facts. Of course, this capability is used to help generate alternatives as well as to assess their probable consequences, enabling the organism to maintain a very simple model of the part of the world that is relevant to its current decisions, and to do commonsense reasoning about the model.

What can we say for and about this behavioral version, this bounded rationality version, of human thinking and problem solving? The first thing we can say is that there is now a tremendous weight of evidence that this theory describes the way people, in fact, make decisions and solve problems. The theory has an increasingly firm empirical base as a description of human behavior. Second, it is a theory that accounts for the fact that creatures stay alive and even thrive, who—however smart they are or think they are—have modest computational abilities in comparison with the complexity of the entire world that surrounds them. It explains how such creatures have survived for at least the millions of years that our species has survived. In a world that is nearly empty, in which not every-

[7] Second ed. (Cambridge, Mass., 1981).

thing is closely connected with everything else, in which problems can be decomposed into their components—in such a world, the kind of rationality I've been describing gets us by.

Consequences of Bounded Rationality

Rationality of the sort described by the behavioral model doesn't optimize, of course. Nor does it even guarantee that our decisions will be consistent. As a matter of fact, it is very easy to show that choices made by an organism having these characteristics will often depend on the order in which alternatives are presented. If A is presented before B, A may seem desirable or at least satisfactory; but if B is presented before A, B will seem desirable and will be chosen before A is even considered.

The behavioral model gives up many of the beautiful formal properties of the Olympian model, but in return for giving them up it provides a way of looking at rationality that explains how creatures with our mental capacities—or even, with our mental capacities supplemented with all the computers in Silicon Valley—get along in a world that is much too complicated to be understood from the Olympian viewpoint of SEU theory.

INTUITIVE RATIONALITY

A third model of human rationality has been much less discussed by social scientists than the two that I've considered so far, but is perhaps even more prominent in the popular imagination. I've referred to it as the intuitive model. The intuitive model postulates that a great deal of human thinking, and a great deal of the success of human beings in arriving at correct decisions, is due to the fact

that they have good intuition or good judgment. The notions of intuition and judgment are particularly prominent in public discussion today because of the research of Roger Sperry and others, much supplemented by speculation, on the specialization of the left and right hemispheres of the human brain.

The Two Sides of the Brain

In the minds and hands of some writers, the notion of hemisphere specialization has been turned into a kind of romance. According to this romanticized account, there's the dull, pedestrian left side of the brain, which is very analytic. It either, depending on your beliefs, does the Olympian kind of reasoning that I described first, or—if it's just a poor man's left hemisphere—does the behavioral kind of thinking I described as the second model. In either case, it's a down-to-earth, pedestrian sort of hemisphere, capable perhaps of deep analysis but not of flights of fancy. Then there's the right hemisphere, in which is stored human imagination, creativity—all those good things that account for the abilities of human beings, if they would entrust themselves to this hemisphere, to solve problems in a creative way.

Before I try to characterize intuition and creativity (they are not always the same thing) in a positive way, I must comment on the romantic view I have just caricatured. When we look for the empirical evidence for it, we find that there is none. There is lots of evidence, of course, for specialization of the hemispheres, but none of that evidence really argues that any complex human mental function is performed by either of the hemispheres alone under normal circumstances. By and large, the evidence shows

that any kind of complex thinking that involves taking in information, processing that information, and doing something with it employs both of our hemispheres in varying proportions and in various ways.

Of course, brain localization is not the important issue at stake. Regardless of whether the same things or different things go on in the two hemispheres, the important question is whether there are two radically different forms of human thought—analytic thought and intuitive thought—and whether what we call creativity relies largely on the latter.

Intuition and Recognition

What is intuition all about? It is an observable fact that people sometimes reach solutions to problems suddenly. They then have an "aha!" experience of varying degrees of intensity. There is no doubt of the genuineness of the phenomenon. Moreover, the problem solutions people reach when they have these experiences, when they make intuitive judgments, frequently are correct.

Good data are available on this point for chess masters. Show a chess position, from a mid-game situation in a reasonable game, to a master or grand master. After looking at it for only five or ten seconds, he will usually be able to propose a strong move—very often the move that is objectively best in the position. If he's playing the game against a strong opponent, he won't make that move immediately; he may sit for three minutes or half an hour in order to decide whether or not his first intuition is really correct. But perhaps 80 or 90 percent of the time, his first impulse will in fact show him the correct move.

The explanation for the chess master's sound intuitions

is well known to psychologists, and is not really surprising.[8] It is no deeper than the explanation of your ability, in a matter of seconds, to recognize one of your friends whom you meet on the path tomorrow as you are going to class. Unless you are very deep in thought as you walk, the recognition will be immediate and reliable. Now in any field in which we have gained considerable experience, we have acquired a large number of "friends"—a large number of stimuli that we can recognize immediately. We can sort the stimulus in whatever sorting net performs this function in the brain (the physiology of it is not understood), and discriminate it from all the other stimuli we might encounter. We can do this not only with faces, but with words in our native language.

Almost every college-educated person can discriminate among, and recall the meanings of, fifty to a hundred thousand different words. Somehow, over the years, we have all spent many hundreds of hours looking at words, and we have made friends with fifty or a hundred thousand of them. Every professional entomologist has a comparable ability to discriminate among the insects he sees, and every botanist among the plants. In any field of expertise, possession of an elaborate discrimination net that permits recognition of any one of tens of thousands of different objects or situations is one of the basic tools of the expert and the principal source of his intuitions.

Counts have been made of the numbers of "friends" that chess masters have: the numbers of different configurations of pieces on a chessboard that are old familiar

[8] For a survey of the evidence, see my *Models of Thought* (New Haven, Conn., 1979), chaps. 6.2–6.5.

acquaintances to them. The estimates come out, as an order of magnitude, around fifty thousand, roughly comparable to vocabulary estimates for native speakers. Intuition is the ability to recognize a friend and to retrieve from memory all the things you've learned about the friend in the years that you've known him. And of course if you know a lot about the friend, you'll be able to make good judgments about him. Should you lend him money or not? Will you get it back if you do? If you know the friend well, you can say "yes" or "no" intuitively.

Acquiring Intuitions and Judgment

Why should we believe that the recognition mechanism explains most of the "aha!" experiences that have been reported in the literature of creativity? An important reason is that valid "aha!" experiences happen only to people who possess the appropriate knowledge. Poincaré rightly said that inspiration comes only to the prepared mind. Today we even have some data that indicate how long it takes to prepare a mind for world-class creative performance.

At first blush, it is not clear why it should take just as long in one field as in another to reach a world-class level of performance. However, human quality of performance is evaluated by comparing it with the performance of other human beings. Hence the length of human life is a controlling parameter in the competition; we can spend a substantial fraction of our lives, but no more, in increasing our proficiency. For this reason, the time required to prepare for world-class performance (by the people whose talents allow them to aspire to that level) should be roughly the same for different fields of activity.

Empirical data gathered by my colleague John R. Hayes for chess masters and composers, and somewhat less systematically for painters and mathematicians, indicate that ten years is the magic number. Almost no person in these disciplines has produced world-class performances without having first put in at least ten years of intensive learning and practice.

What about child prodigies? Mozart was composing world-class music perhaps by the time he was seventeen—certainly no earlier. (The standard Hayes used for music is five or more appearances of recordings of a piece of music in the Schwann catalog. Except for some Mozart juvenilia, which no one would bother to listen to if they hadn't been written by Mozart, there is no world-class Mozart before the age of seventeen.) Of course Mozart was already composing at the age of four, so that by age seventeen he had already been educating himself for thirteen years. Mozart is typical of the child prodigies whose biographies Hayes has examined. A *sine qua non* for outstanding work is diligent attention to the field over a decade or more.

Summary: The Intuitive and Behavioral Models

There is no contradiction between the intuitive model of thinking and the behavioral model, nor do the two models represent alternative modes of thought residing in different cerebral hemispheres and competing for control over the mind. All serious thinking calls on both modes, both search-like processes and the sudden recognition of familiar patterns. Without recognition based on previous experience, search through complex spaces would proceed in snail-like fashion. Intuition exploits the knowledge we

have gained through our past searches. Hence we would expect what in fact occurs, that the expert will often be able to proceed intuitively in attacking a problem that requires painful search for the novice. And we would expect also that in most problem situations combining aspects of novelty with familiar components, intuition and search will cooperate in reaching solutions.

INTUITION AND EMOTION

Thus far in our discussion of intuitive processes we have left aside one of the important characteristics these processes are said to possess: their frequent association with emotion. The searching, plodding stages of problem solving tend to be relatively free from intense emotion; they may be described as cold cognition. But sudden discovery, the "aha!" experience, tends to evoke emotion; it is hot cognition. Sometimes ideas come to people when they are excited about something.

Emotion and Attention

Hence, in order to have anything like a complete theory of human rationality, we have to understand what role emotion plays in it. Most likely it serves several quite distinct functions. First of all, some kinds of emotion (e.g., pleasure) are consumption goods. They enter into the utility function of the Olympian theory, and must be counted among the goals we strive for in the behavioral model of rationality.

But for our purposes, emotion has particular importance because of its function of selecting particular things in our environments as the focus of our attention. Why

was Rachel Carson's *Silent Spring* so influential? The problems she described were already known to ecologists and the other biologists at the time she described them. But she described them in a way that aroused emotion, that riveted our attention on the problem she raised. That emotion, once aroused, wouldn't let us go off and worry about other problems until something had been done about this one. At the very least, emotion kept the problem in the back of our minds as a nagging issue that wouldn't go away.

In the Olympian model, all problems are permanently and simultaneously on the agenda (until they are solved). In the behavioral model, by contrast, the choice of problems for the agenda is a matter of central importance, and emotion may play a large role in that choice.

Emotion does not always direct our attention to goals we regard as desirable. If I may go back to my example of *Mein Kampf*, we observed that the reasoning in that book is not cold reasoning but hot reasoning. It is reasoning that seeks deliberately to arouse strong emotions, often the emotion of hate, a powerful human emotion. And of course, the influence of *Mein Kampf*, like that of *Silent Spring* or Picasso's *Guernica*, was due in large part to the fact that it did have evocative power, the ability to arouse and fix the attention of its German readers on the particular goals it had in mind.

A behavioral theory of rationality, with its concern for the focus of attention as a major determinant of choice, does not dissociate emotion from human thought, nor does it in any respect underestimate the powerful effects of emotion in setting the agenda for human problem solving.

Emotion in Education

I would like to take a brief excursion at this point in order to consider the role of emotion in education. If literary and artistic works have a considerable power to evoke emotions, as they certainly do, does this power suggest any special role for them in the educational process?

We all know that the humanities feel a bit besieged today. A large proportion of the students in our universities appear to want to enroll in law, business, or medicine, and the humanities suffer neglect, benign or otherwise. One argument that is often advanced by those who would counter this trend is that it may be better, more effective, for students to learn about the human condition by exposure to the artist's and humanist's view of the world than by exposure to the scientist's. Of course my own professional identifications put me on the other side of the argument, but I think we should look at the issue quite carefully. What are the optimum conditions for efficient human learning about central and important matters? Which is better, cold cognition or hot? And whichever is better, will we find that this is the kind we associate with the sciences or the humanities?

I should say here that I have heard physicists argue for a strong infusion of hot cognition in teaching their subject. The problems that excite them, and motivate them to understand rather abstruse matters, are the cosmological and philosophical problems associated with the fundamental particles, and with astrophysics and the architecture of the universe. So perhaps I should not have associated science strictly with cold cognition.

But let me go to a domain where the point can be made more unequivocally and convincingly. Perhaps some of you are familiar with Arthur Koestler's *Darkness at Noon*. It is a novel that describes what happens to a particular person at the time of the Russian purge trials of the 1930's. Now suppose you wish to understand the history of the Western world between the two world wars, and the events that led up to our contemporary world. You will then certainly need to understand the purge trials. Are you more likely to gain such an understanding by reading *Darkness at Noon*, or by reading a history book that deals with the trials, or by searching out the published transcripts of the trial testimony in the library? I would vote for Koestler's book as the best route, precisely because of the intense emotions it evokes in most readers.

I could go down a long list of such alternatives: *War and Peace* versus a treatise on military sociology, Proust and Chekov versus a textbook on personality. If I were in a position where I had to defend the role of the humanities in education, to provide an argument for something like the traditional liberal arts curriculum of the early twentieth century, I would argue for them on the grounds that most human beings are able to attend to issues longer, to think harder about them, to receive deeper impressions that last longer, if information is presented in a context of emotion—a sort of hot dressing—than if it is presented wholly without affect.

But educating with the help of hot cognition also implies a responsibility. If we are to learn our social science from novelists, then the novelists have to get it right. The scientific content must be valid. Freudian theory permeates a great deal of literature today—at the very time when Freudian theories are being revised radically by new psychological knowledge. There are few orthodox Freudians left in psychology today. Hence there is a danger, if we take this route of asking the humanities to provide an emotional context for learning, that a kind of warmed-over Freud will be served to our students in a powerfully influential form. We have to re-evaluate the great humanist classics to see to what extent they suffer from obsolescence through the progress of our scientific knowledge.

Homer is still alive because the *Iliad* and the *Odyssey* treat mainly of matters in which modern social science has not progressed much beyond lay understanding. Aristotle is barely alive—and certainly his scientific works are not, and his logic hardly. And we could have a great argument with philosophers as to whether his epistemology or his metaphysics has anything to say to students today. And Lucretius, of course, talking about atoms, has gone entirely.

The moral I draw is that, whereas works capable of evoking emotion may have special value for us just by virtue of that capability, if we wish to use them to educate, we must evaluate not only their power to rouse emotion but also their scientific validity when they speak of matters of fact.

If the humanities are to base their claims to a central place in the liberal curriculum on their special insights into the human condition, they must be able to show that their picture of that condition is biologically, sociologically, and psychologically defensible. It is not enough, for this particular purpose, that humanistic works move students.

They must move them in ways that will enable them to live with due regard for reason and fact in the real world. I do not mean to imply that the humanities do not now meet this standard; a detailed assessment of the liberal curriculum in any existing university would certainly not give a simple yes-or-no answer to that question. But I do suggest that any examination of the appropriate roles of different fields of knowledge in providing the materials of a liberal education needs to give close attention both to the emotional temperature of material and to its empirical soundness.

CONCLUSION

In this first chapter, I have sought to present three visions of rationality: three ways of talking about rational choice. The first of these, the Olympian model, postulates a heroic man making comprehensive choices in an integrated universe. The Olympian view serves, perhaps, as a model of the mind of God, but certainly not as a model of the mind of man. I have been rather critical of that theory for present purposes.

The second, the behavioral model, postulates that human rationality is very limited, very much bounded by the situation and by human computational powers. I have argued that there is a great deal of empirical evidence supporting this kind of theory as a valid description of how human beings make decisions. It is a theory of how organisms, including man, possessing limited computational abilities, make adaptive choices and sometimes survive in a complex, but mostly empty, world.

The third, the intuitive model, places great stress on the

processes of intuition. The intuitive theory, I have argued, is in fact a component of the behavioral theory. It emphasizes the recognition processes that underlie the skills humans can acquire by storing experience and by recognizing situations in which their experience is relevant and appropriate. The intuitive theory recognizes that human thought is often affected by emotion, and addresses the question of what function emotion plays in focusing human attention on particular problems at particular times.

I have left for the next chapter a fourth theory: the vision of rationality as evolutionary adaptation. The evolutionary model is a *de facto* model of rationality; it implies that only those organisms that adapt, that behave *as if* they were rational, will survive. In the next chapter, I shall examine these claims of the efficacy and centrality of natural selection as applied to the exercise of human rationality.