HUMAN PROBLEM SOLVING

ALLEN NEWELL
University Professor
Carnegie-Mellon University

HERBERT A. SIMON
Professor of Computer Science and Psychology
Carnegie-Mellon University

PRENTICE-HALL, INC., Englewood Cliffs, New Jersey

13–445403–0

Library of Congress Catalog Card Number: 79–15252

Printed in the United States of America

Current printing (last digit):
10 9 8 7 6 5

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

PREFACE

The problem solving that went into this volume on human problem solving extends over a seventeen-year period—the actual writing of it over more than a decade: one of us claims to remember a portion of a draft dated 1956 and both remember a draft dated 1959. Since our aims and intentions for the book are spelled out in the introduction and we do not think that we should further delay giving the reader the results of our deliberations, we will limit this preface to acknowledgments.

An alarming number of debts to colleagues can be incurred during seventeen years. Our more general intellectual obligations are sketched in the historical addendum with which the book closes, but we should like to mention by name some of those with whom we have worked most closely, and from whom we have borrowed most specifically, in the preparation of this volume.

There is first of all Cliff Shaw who, as the dedication states, "helped us start all this." He participated as a full member of the triumvirate from the beginning of the project, in 1955, until about 1959, when his own interests and activities gradually took him off in other directions. His grasp of the basic principles of

information processing was profound and original, providing indispensable ingredients for our approach.

For the most part the others who participated directly in the work reported in this book were our doctoral students: David Bree, Andrew Chenzoff, and Peter Houts, who collected most of the logic and cryptarithmetic protocols; Arnold Winikoff, who carried out the eye-movement studies; George Baylor, who programmed MATER II and gathered several chess protocols; and Michael Barenfeld, who did much of the programming for the chess perception system reported in Chapter 13.

But there are many others—students and faculty colleagues—whose contributions were less direct, but certainly not less appreciated. We have worked in close association on related projects with Lee Gregg and Walter Reitman, and more recently with William Chase and Donald Waterman. We have enjoyed continuing communication and interchange of ideas with Adriaan de Groot of the University of Amsterdam. We have learned much from our doctoral students working on topics in psychology and artificial intelligence outside the area of problem solving: Geoffrey Clarkson, Stephen Coles, Donald Dansereau, George Ernst, Edward Feigenbaum, Julian Feldman, Kenneth Laughery, Robert Lindsay, James Moore, Laurent Siklóssy, Frederick Tonge, Donald Williams, Thomas Williams, and Richard Young.

Several organizations and agencies have provided generous financial support for our research. During the first years of work, the largest part of the support came from the RAND Corporation, on whose computers, JOHNNIAC and its successors, most of the simulation was then done. At Carnegie-Mellon University we had initially been granted a substantial research allowance from the Carnegie Corporation. Since the middle sixties, the principal support has come from the National Institute of Mental Health (Project MH-07722), and from the Advanced Research Projects Agency of the Office of the Secretary of Defense (F44620-67-C-0058). We are grateful to all of these organizations for the confidence they have placed in us, and the broad terms upon which they have offered their support. Throughout the period of our investigations, we have been in the ideal but embarrassing position for researchers: the rate at which we could progress has been determined by the rate of generating research ideas, rather than by limits on the magnitude of the financial resources available to us.

Most of the empirical data analyzed in detail here comes from our own laboratory. However, in Chapter 7, we have made some use of protocols published by Professor F. C. Bartlett in his book, *Thinking*; and in Chapter 13, some of Adriaan de Groot's chess protocols. In Chapters 9 and 10, we present analyses of data provided to us by O. K. Moore and Scarvia Anderson, gathered in the course of problem-solving studies carried out at Yale University. Our debt to them is acknowledged more specifically in the chapters themselves.

We would be embarrassed to count up the number of drafts that have passed through the typewriters and editorial hands of our secretaries, Evelyn Adams and Mildred Sisko. They have been most patient and helpful through this whole long process. Finally, we want to thank the members of the editorial and production staffs of Prentice-Hall, who showed great skill in interpreting our intentions—some of them all too implicitly indicated in our manuscript.

Some portions of our data and analysis have been published previously, but most of this work has been extensively revised for the present volume. The last half of Chapter 4 is based on Newell, Shaw, and Simon, 1957; Chapter 6 on Newell, 1967c and 1968b; Chapter 8, with much revision, on Newell, Shaw, and Simon, 1960a; parts of Chapter 9 on Newell and Simon, 1961a; Chapter 11 on Newell, Shaw, and Simon, 1958c—again with much modification; Chapter 12 on Newell and Simon, 1965b; and parts of Chapter 13 on Baylor, 1965, Baylor and Simon, 1966, and Simon and Barenfeld, 1969.

CONTENTS

1 INTRODUCTION

The aim of this book is to advance our understanding of how humans think. It seeks to do so by putting forth a theory of human problem solving, along with a body of empirical evidence that permits assessment of the theory.

No single work advances understanding very far. The aims of a scientific work are limited by the formal character of the theory, by the phenomena it encompasses, by the experimental situations it uses, by the types of subjects it studies, and by the data it gathers. Of course, a theory may speak beyond its initial base—all scientists hope for just that. But science is a series of successive approximations. Not all things can be done at once, and even if one aspires to go far, he must start somewhere. If one aims at covering all of human thinking in a single work, the work will necessarily be superficial. If one aims at probing in depth, then many aspects of the subject, however important, will be left untouched.

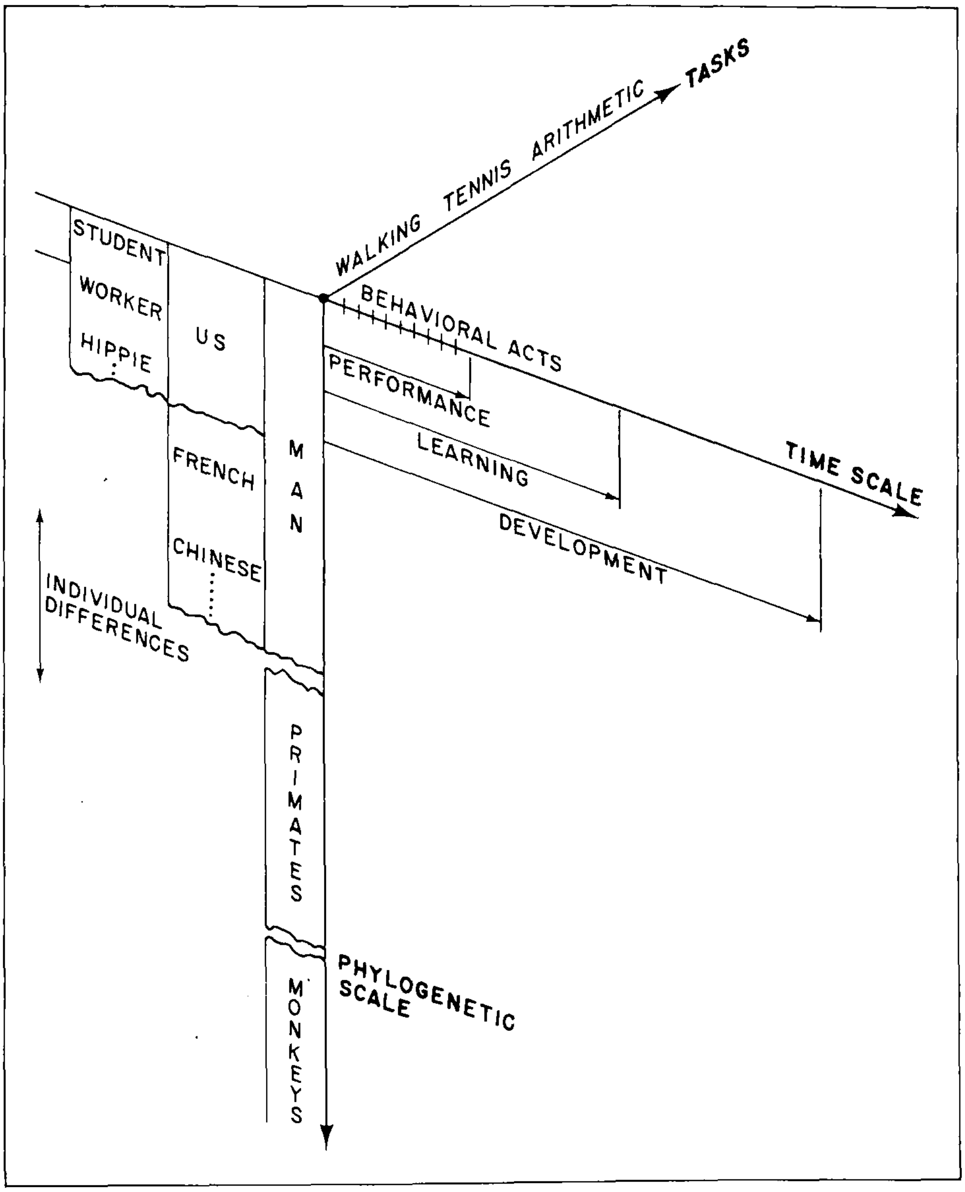
THE SCOPE OF THE BOOK

Our first task, then, is to indicate the scope of this work—what it includes and what it deliberately excludes. The boundaries need not be carefully marked; but the central focus should be made clear.

Figure 1.1 attempts to compress in one diagram many of the dimensions

FIGURE 1.1

dimensions of variation of the human system



along which the total human system can vary. Its purpose is to demark the particular portions included in the present work, not to provide a new or total view. The focus of the diagram is the individual human being. He is a system consisting of parts: sensory subsystems, memory, effectors, arousal subsystems, and so on. It makes little difference for immediate purposes that we are unsure just what is the most appropriate way to enumerate the subsystems. We can limit study—and we will—to one or a few parts.

Task Dimension. A human behaves in a number of different classes of situations, which we will come to call *task environments*. In the figure, we have represented these in a single dimension, but clearly a whole geography is intended: a human does mathematics, he walks, he interacts with his fellow man, he drives cars, he makes love, he argues, he buys food, he dies.

Performance-Learning-Development Dimension. Holding the task environment constant, we usually distinguish a human who is *performing* a task from one who is *learning* to perform a task, or one who is *developing* with respect to a task. Within the first group, those who are performing, we can focus on successively smaller acts, so that we need not even be concerned with an entire performance. It is not important that the distinctions are imperfect. (For example, it is often impossible to distinguish with confidence between learning and maturation.) The distinctions among these three kinds of activities are to some extent correlated with time scale, and we have so depicted them.

Individual-Difference Dimension. A single man may be viewed as a member of various *populations*. Each man differs—both in systematic ways and simply by virtue of his unique genetic endowment and historical fate—from all other men. Differences dependent on age are strongly related to development and are indicated along the temporal axis. But cultures also differ, and within a culture various demographic variables, such as socioeconomic status, isolate distinct populations. More broadly, of course, Man is only one among innumerable species, and so the vertical axis extends to other organisms.

The Book's Focus

From the many directions of variation expressed or implied in Figure 1.1 limited regions can be carved out for attention: the development of intelligence as a function of organismic level (in the evolutionary sense); the learning of personal roles in social groups of college students; and so on. Each region provides a starting point for investigation and description; each leaves out most of the human phenomena.

The present study is concerned with the performance of intelligent adults in our own culture. The tasks discussed are short (half-hour), moderately difficult problems of a symbolic nature. The three main tasks we use—chess, symbolic logic, and algebra-like puzzles (called cryptarithmetic puzzles)—typify this class of problems. The study is concerned with the integrated activities that constitute problem solving. It is not centrally concerned with perception, motor skill, or what are called personality variables. The study is concerned primarily with performance, only a little with learning, and not at all with development or differences related to age. Finally, it is concerned with integrated activities, hence deemphasizes the details of processing on the time scale of elementary reactions (that is, half a second or less). Similarly, long-term integrated activities extending over periods of days or years receive no attention.

These restrictions on the scope of our study delineate a central focus, not a set of boundaries. Thus, the possibility that learning is taking place cannot be excluded in complex tasks that last tens of minutes. Likewise, it is an empirical question, not a matter of definition or fiat, to what extent perceptual mechanisms are important in problem solving—although limiting our investigation to symbolic tasks helps to restrict the importance of perception. Again, how the basic mechanisms of immediate memory and immediate processing affect integrated behavior at the next higher level is an open question. One recent author on cognitive processes (Neisser, 1967) has taken the description of the immediate processing mechanisms as the important first task, devoting only a last chapter of his book to the integrated level of behavior that is central to this volume.

Information Processing Theory

The reasons for our choices are various and largely opportunistic (although with an opportunism that has lasted twenty years and thus may constitute philosophic conviction). For several decades psychology, responding to different opportunities and convictions, focused on learning, lower organisms, and tasks that are simple from an adult human viewpoint. Within the last dozen years a general change in scientific outlook has occurred, consonant with the point of view represented here. One can date the change roughly from 1956: in psychology, by the appearance of Bruner, Goodnow, and Austin's *Study of Thinking* and George Miller's "The magical number seven"; in linguistics, by Noam Chomsky's "Three models of language"; and in computer science, by our own paper on the Logic Theory Machine.

As these titles show, the common new emphasis was not the investigation of problem solving, but rather the exploration of complex processes and the acceptance of the need to be explicit about internal, symbolic mechanisms. Nor do all four of these works stem from a specific common lineage, unless it be the whole of applied mathematics and technology: control theory, information theory, operational mathematics including game theory and decision theory, computers and programming. These topics emerged in World War II and ramified through the late forties and early fifties in many directions, the new approach to the study of man being only one.

We ourselves, through the forties and early fifties, were largely concerned with human organizational behavior and were influenced strongly by the growing technologies just mentioned. But the specific opportunity that has set the course of the present work is the development of a science of information processing (now generally termed computer science). Thus, this study is concerned with thinking—or that subspecies of it called problem solving—but it approaches the subject in a definite way. It asserts specifically that thinking can be explained by means of an information processing theory. This assertion requires some explanation.

The present theory views a human as a processor of information. Both of these notions—*information* and *processing*—are long-established, highly general concepts. Thus, the label could be thought vacuous unless the phrase *information processing* took on additional technical meaning.

One may try to provide this meaning by saying that a computer is an instance of an information processor. This would suggest that the phrase is a metaphor: that man is to be modeled as a digital computer. Metaphors have their own good place in science, though there is neither terminology nor metatheory of science to explicate the roles of metaphors, analogs, models, theories and descriptions, or the passage from one category to another (Simon and Newell, 1956). Something ceases to be metaphor when detailed calculations can be made from it; it remains metaphor when it is rich in features in its own right, whose relevance to the object of comparison is problematic. Thus a computer can indeed be a metaphor for man; then it becomes relevant to discover whether man is all bits on the inside.

But an alternative to metaphor is at hand. An abstract concept of an information processing system has emerged with the development of the digital computer. In fact, a whole array of different abstract concepts has developed, as scientists have sought to capture the essence of the new technology in different ways (Minsky, 1967). The various features that make the digital computer seem machinelike—its fast arithmetic, its simply ordered memory, its construction by means of binary elements—all have faded in the search for the essential. Thus, in this book we will introduce a suitable abstract information processing system to describe how man processes task-oriented symbolic information. This is not the most abstract possible way to describe an information processing system, but it is tailored to our scientific needs.

An information processing theory is not restricted to stating generalities about Man. With a model of an information processing system, it becomes meaningful to try to represent in some detail a particular man at work on a particular task. Such a representation is no metaphor, but a precise symbolic model on the basis of which pertinent specific aspects of the man's problem solving behavior can be calculated. This model of symbol manipulation remains very much an approximation, of course, hypothesizing in an extreme form the neatness of discrete symbols and a small set of elementary processes, each with precisely defined and limited behavior. This abstraction, though possibly severe, does provide a grip on symbolic behavior that was not available heretofore. It does, equally, steer away from physiological models, with their concern for fidelity to continuous physiological mechanisms, either electrical, chemical, or hormonal. Perhaps the nonphysiological nature of the theory is not as disadvantageous as one might first believe, for the collection of mechanisms that are at present somewhat understood in neuropsychology is not at all adequate to the tasks dealt with in this book. We could not have proceeded to construct theories of human behavior in these tasks had we restricted ourselves to mechanisms that can today be provided with physiological bases.

The Choice of Scope

We have brought out the general grounding of our work in computer science to explain the limitation of our aims—asserting that this limitation is really an opportunity, since information processing systems provide our first precise notion of what symbols and symbol manipulation could mean. However, once we have chosen to study symbolic systems in a technical and precise way, many alternative paths still remain open along both the task dimension and the performance-learning-development dimension. Something needs to be said of our choices here.

Relation to Artificial Intelligence. The most important influence upon our choice of tasks such as chess and symbolic logic is the development of the field of artificial intelligence. This is the part of computer science devoted to getting computers (or other devices) to perform tasks requiring intelligence. As will become clear, a theory of the psychology of problem solving requires not only good task analyses but also an inventory of possible problem solving mechanisms from which one can surmise what actual mechanisms are being used by humans. Thus, one must work with task environments in which artificial intelligence has provided the requisite array of plausible mechanisms. The task areas represented in this book satisfy these conditions: game players, theorem provers, and puzzle solvers constitute a large fraction of the existing artificial intelligence systems. Many other task areas that are attractive on other grounds have not yet been studied extensively from the standpoint of artificial intelligence, hence are less useful for our purposes than those mentioned above.

On two counts the previous paragraph is insufficient to explain our choice of tasks. First, it assumes that artificial intelligence moves independently of psychology. This is demonstrably not the case. Much of the work in artificial intelligence started from psychological concerns. Second, and more important, the explanation simply raises the new question of why artificial intelligence research should take up the types of tasks represented in this book and not others.

A partial answer is that many other tasks have in fact been worked on. Language processing is an example. But while there has been a great deal of work in linguistics, almost all of it has focused on grammatical analysis (*competence*), rather than on the use of language (*performance*). Programs dealing with semantics and pragmatics are fewer, more recent, and somewhat less developed than the problem solving programs represented here. Thus, we will have little to say about language processing behavior as such. However, we shall see that our theory of problem solving has strong implications for what the linguist calls the deep structure of language.

Pattern recognition is another area in which there has been much work, both theoretical and empirical, even more extensive than the body of work upon which we draw. We make little use of this research here, because sequential, integrated behavior is at the heart of most thinking, while most artificial intelligence pattern recognition machines are built around the single act of recognition. While many of the schemes do involve some sequential processing, none of them is adequate for modeling general sequential behavior.

There are many kinds of thinking that one might like to study: designing a house, discovering a new scientific law, preparing a law case, arguing over political parties, creating new music, daydreaming while watching the clouds, preparing a five-year economic plan, and so on. Detailed theories of these and many other kinds of thinking are largely beyond the current state of the art. Of course, there have been investigations into some of these areas, many of them still in midstream. Only their incomplete state and our limits of space and energy have inhibited us from including some of them in this work, since they are in fact part of the same story we wish to tell.

Emphasis on Performance. Turning to the performance-learning-development dimension, our emphasis on performance again represents a scientific bet. We recognize that what sort of information processing system a human becomes depends intimately on the way he develops. The kernel from which development starts—say, the neonate—already contains a genetically determined set of mechanisms of immense complexity. How complex they are is easily appreciated by anyone who follows the acts of self organization that take place in the embryo as it progresses to the neonatal starting position. Still, by common scientific assent, the emerging system is remarkably content-free, and without the powers of integrated action shown by the normal adult. Many constraints on the nature of the fully developed system arise from the requirement of self organization—help from the external environment (say via language) can only be used after the system has developed itself to a point where it is capable of such assimilation. Yet, acknowledging this, it still seems to us that we have too imperfect a view of the system's final nature to be able to make predictions from the development process to the characteristics of the structures it produces.

Similar remarks apply to learning. Humans learn continuously, and much that they do involves using in obvious ways information gathered for a specific purpose, rather than solving difficult problems like those studied in this book. One enters a department store: "Where do I find men's suits?" "Third floor, down the center aisle and to your right"; "Thank you"; and off one goes, following directions. Several phenomena here are closely allied to the interests of this book: language production and reception; deciding to ask for information to solve a problem; following directions, once assimilated; perhaps (if the directions were imperfect) solving some smallish subproblems along the way. Certainly this is the behavior of an information processing system. But the task in this case is carefully contrived to permit simple learning to substitute for most of the work of problem solving.

Learning is a second-order effect. That is, it transforms a system capable of certain performances into a system capable of additional ones (usually better ones; usually accomplished without losing much of the preexisting performance capability; and usually integrated with existing capability so they can be evoked on appropriate occasions). The study of learning, if carried out with theoretical precision, must start with a model of a performing organism, so that one can represent, as learning, the changes in the model.[1] The mathematization of learning theory in the last decade shows this very well (Atkinson, Bower, and Crothers, 1965). In the prototype version of mathematical learning theories, the organism is represented by a set of probabilities of occurrence of a fixed set of responses; learning involves changes in these probabilities under the impact of experience.

The study of learning takes its cue, then, from the nature of the performance system. If performance is not well understood, it is somewhat premature to study learning. Nevertheless, we pay a price for the omission of learning, for we might otherwise draw inferences about the performance system from the fact that the system must be capable of modification through learning. It is our judgment that in the present state of the art, the study of performance must be given precedence, even if the strategy is not costless. Both learning and development must then be incorporated in integral ways in the more complete and successful theory of human information processing that will emerge at a later stage in the development of our science.

Other Omissions. The omission of both sensory and motor skills, and many aspects of perception, from our study is perhaps plausible on the surface: we are concerned with central symbolic activities. Nevertheless, at least one scientist, Bartlett, in *Thinking* (1958), made motor skills the key metaphor for attempting to understand thinking. And a whole school, the Gestaltists, have used perception as the touchstone of central activity (Wertheimer, 1945). On the perceptual side, a basis for our choice has already been indicated: our concern with sequential behavior. However, as even this book will reveal, the boundary between perceptual and symbolic behavior may not be maintainable for long, especially if work pushes in the direction of better models of the immediate processor.

On the motor side the situation is peculiar. Bartlett chose to compare thinking with motor skills precisely because he felt sequential behavior was central both in thinking and in motor skills, where it had been studied intensively. Our own feeling is that, while there has been much experimental investigation, motor skills have not yet found a mechanistic representation having anything like the power of the information processing representation exploited here. Furthermore, motor skills seem in considerable part to be nonsymbolic—and that makes them a poor model for a system where symbols are central.

Our final choice is to exclude motivational and personality variables—what Abelson (1963) covered in part by the term "hot cognition." We omit them by reason of convictions, not about the importance or unimportance of the phenomena, but about the order in which theory should develop. Many motivational and emotional phenomena operate through the lens of the cognitive system, as the work of Schachter and Singer (1962) and others has indicated. A plausible scientific strategy is to put our cognitive models in order before moving to these other phenomena. Our one exploratory foray in the direction of motivation and emotion was precisely in this vein (Simon, 1967a).

[1] For empirical purposes, of course, one can always get by for a while by talking in the language of experimental protocols—simply describing the changed behavior in the experimental situation.