The stochastic modeling of elementary psychological processes

JAMES T. TOWNSEND
Purdue University

F. GREGORY ASHBY
Ohio State University

CAMBRIDGE UNIVERSITY PRESS

Cambridge
London New York New Rochelle
Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1983

Printed in the United States of America

Library of Congress Cataloging in Publication Data

Townsend, James T.

The stochastic modeling of elementary psychological processes.

Includes bibliographical references and index.

1. Psychology – Mathematical models.
2. Cognition – Mathematical models.
3. Stochastic processes. I. Ashby, F. Gregory.
II. Title.
BF39.T63 1983 150'.724 82-9613
ISBN 0 521 24181 2 hard covers
ISBN 0 521 27433 8 paperback

To our parents

Contents

Preface

It takes only a little reading of late nineteenth- and early twentieth-century philosophical literature to arrive at the conclusion that many of the existing ideas about human cognition were present even before much rigorous experimentation had been accomplished. Ideas about memory, perception, discrimination, motivation, will (conation), attention, intelligence, and courage, to name a few, were discussed by people like James, Sully, Hamilton, Ladd, Wundt, Cattell, Fechner, Stumpf, and Helmholtz.

The data varied from nil to impressive, and hardly a year goes by without the design of some famous recent experiment being discovered buried away in the older literature. Nevertheless, the concepts were not always cleanly defined, particularly in what we would now call operational terms; the theories were typically amorphous and qualitatively drawn (with the notable exception of some parts of psychophysics and, to a reasonable extent, physiological psychology); introspection was sometimes relied on too extensively; and the experiments did not always allow one to confidently embrace a "winning" theory or hypothesis while disposing of the alternatives.

It is our opinion that the most significant contribution of twentieth-century psychology has been in developing and refining the theories, concepts, and methodologies concerned with explaining in an elegant and nontrivial fashion how reasonably complex organisms go about their business. Our own bias, which is apparent on simply leafing through the pages of the book, is that mathematical work, when directed toward specific psychological questions or problems, has led to and will continue to lead to some of the most critical advances in psychology. It may be contested that there exist theoretical efforts of a mathematical nature that are little more than variations on an esoteric glass bead game (Hesse 1969). On the other hand, the frequency of productions of little quality or ultimate impact is perhaps not more profuse than those of an almost purely empirical or entirely verbal substance.

Quite a lot of theoretical work, even in the reasonably rigorous avenues of experimental psychology, is still qualitative. Although in some cases the nature of the material or the stage of the research would render mathematical theorizing futile, many others, sometimes even the best, might be improved in clarity and testability by expressing the main ideas in mathematical form. One manner of accomplishing this is, of course, to write the axioms, derivations, and theorems in analytic form (that is, in closed mathematical expressions), whereas another increasingly popular strategy is computer simulation of psychological processes. The latter is especially helpful with very complex cognitive processes.

In the disciplines that are still amenable to some analytic modeling, typically dealing with fairly elementary perceptual, memorial, or decision behavior, we are even now seeing a move into ever more intricate models, often before understanding the limits of testability of far simpler concepts. This may represent the natural order of scientific evolution, but we would hope that attention to such questions would follow more or less inexorably and with not too long a delay.

An allied difficulty in some cases is that formal mathematical modeling has recently been carried out on complex or higher-level mental activities, but at the same time the applications have retreated to a qualitative account of the pertinent behavior. The latter trend is unfortunate because one of the most admirable properties of mathematical theorization has been the amenability to rigorous quantitative testing.

A less serious problem is that the investigator is often satisfied to show that his or her model is sufficient to explain the data within some criterion (that is sometimes not even given beforehand). Much is often learned in this way, but a difficulty is that when taking a model as the null hypothesis, running a very clean experiment for sufficiently many trials is almost certain to "falsify" the model, whereas running a sloppier experiment for fewer trials is more likely to statistically "verify" the model. Perhaps the best way we know of how to ameliorate this dilemma, albeit a far from perfect way, is to test two or more models against one another. The models should probably encompass the broadest possible opposing psychological concepts at first so that the widest regions can be falsified in principle. If one model is tentatively accepted and the other falsified, then further efforts can be undertaken to narrow the latitude of the "correct" model – for instance, deciding among interesting subcases of the model. It is for this reason that a fair amount of the labor in the chapters to follow is geared toward fashioning general classes of models to represent interesting and opposing psychological processing concepts with a view to ascertaining where they can and where they cannot be tested.

Before giving a brief overview of the contents of the book, we remark on what it is not.

It is not an introduction to mathematical psychology (as in Coombs, Dawes, & Tversky 1970) because it does not survey the field. Mathematical learning theory (e.g., as in Atkinson, Bower, & Crothers 1965 or Restle & Greeno 1970) is absent, as is material on the foundations of measurement (as in Krantz, Luce, Suppes, & Tversky 1971) except for a little discussion in Chapter 12. No physiological models are present. The only decision theory (as traditionally defined) is that based on random walk models in Chapter 10.

The major goals of the present work are twofold. The first is to offer an introduction to the fundamentals of probabilistic modeling of certain simple cognitive processes, most of it done in the context of continuous time-dependent phenomena.[1] This aim is in the spirit of McGill (1963), in the sense of constructing models of time-dependent psychological functions based on reasonably simple stochastic processes. The second goal, not mutually exclusive of the first, is the development of mathematical theories capable of capturing broad classes of models that represent psychological mechanisms differing on one or more important dimensions or qualities. These theories aim, in particular, to establish the regions of the models wherein little or no hope of rigorous testability lies and those regions and consequent experimental paradigms where theoretical issues can be tested.

One important focal point of the developments resides within the confines of reaction time methodology, where the response latency is employed to aid in mapping out various cognitive operations. The employment of reaction time as an experimental dependent variable in conjunction with the manipulation of various independent variables has long been a valuable technique in psychology. A terse summary of its historical development is offered in Chapter 1.

A word is in order concerning the part that accuracy plays in theories and experiments based on reaction time. Accuracy can be strategic in interpreting reaction time effects in some experimental circumstances because the two may covary as the experimental variables are manipulated. First of all it is important to note that if the accuracy (given in probability correct etc.) is a well-specified function of the completion times of the objects being processed (as well as, of course, other contextual and experimental variables), then any two models that predict the same probability distributions on completion times must perforce also make identical predictions on accuracy as well as the relation between speed of processing (i.e., completions) and accuracy. Many natural models of a wide variety of psychological situations fall into this camp. This is important because it means that two such models that are equivalent on completion times will also produce identical speed–accuracy relations, without the necessity of working out the latter predictions. Secondly, the response errors that appear in some experimental procedures may be entirely unrelated to the psychological process under examination.

On the other hand, there are important situations where two models differ fundamentally in the way in which errors are produced or in which accuracy predictions form a fundamental aspect of a model's application to a specific empirical situation. Such considerations particularly arise in Chapters 5, 9, 10, 11, and 15, although several of the other chapters have occasion to remark on accuracy effects.

The book can be read with a background of some calculus and a little ele-

[1] The term *process* will be used to denote any cognitive operation. A process typically will reside in a single functionally defined subsystem of the overall system complex. Often in the present work, *processing* will refer to the identification of objects, the latter usually being referred to as *elements*.

mentary continuous probability theory, although a smattering of experience with stochastic processes might make the going a bit smoother in places. Some transform theory is employed, but the major facts needed are covered early in the text. A modicum of exposure to experimental psychology, particularly information-processing concepts, would of course also be helpful. We have tried to avoid overformalism in an effort to make the development more readable, especially for the nonspecialist, and to emphasize psychological underpinnings. Nevertheless, important results are often given in proposition–proof form for clarity and "bookkeeping" processes.

The first four chapters set the stage and introduce the reader to the primary tools, concepts, and issues used throughout the book. They are requisite reading preparatory to the rest of the book except, perhaps, for the expert. Those who have a little modeling background may wish to skip to Chapter 3. Some of the major issues introduced in Chapters 1–4 are parallel vs. serial (roughly, one-at-a-time vs. simultaneous) processing, self-terminating vs. exhaustive processing (see below) and limited vs. unlimited capacity (i.e., the question of whether an increased processing load produces slower reaction times and/or lower accuracy).

Chapter 5 discusses the more complex compound processing models and may be omitted without harm to the remaining portions on a first reading. Paradigms built around search for a target in a short list stored in memory or displayed visually have received an enormous amount of attention, yet significant issues remain unresolved. Chapter 6 examines some of the critical problems and potential solutions. Chapter 7 takes up the self-terminating vs. exhaustive processing issue, long popular in cognitive psychology, concerned as it is with partial as opposed to complete processing of all items, when some are sufficient to determine the response.

Chapter 8 focuses on reaction time distributions and how they might be used to aid system identification – that is, to discriminate parallel from serial, limited from unlimited capacity, and self-terminating from exhaustive processing. Chapters 9 and 10 discuss models that make specific predictions about response accuracy as well as latency. Chapter 9 examines counter models and models that we have called varied-state models – that is, models proposing that on each trial the observer is in one of a number of states, each of which is associated with its own accuracy and latency distribution. Chapter 10 then discusses random walk models, emphasizing the work of Link and Heath (1975) and Laming (1968) but also including some previously unpublished results of our own.

Chapters 11 and 12 consider two content areas of substantial contemporary interest in information processing in light of the present modeling and methodology. Questions about how a person perceives, retains, and reports as many unrelated letters as possible are investigated in Chapter 11 on the so-called whole report procedure. The basic concept of additivity of latencies associated with distinct mechanisms or processing subsystems is reviewed in

Chapter 12 with regard to parallel and serial processing, measurement theory, and temporally overlapping processing operations.

The last three chapters concentrate on the parallel–serial question, which indeed serves as a hub for a not insignificant portion of the book's issues. Chapter 13 introduces the parallel–serial testing paradigm and generalizes results found earlier. The most general (to date) aspects of parallel–serial equivalence, mathematical discriminability and empirical testability, are tackled in Chapters 14 and 15. Chapter 15 is in a somewhat less formalized and more discursive style because several of the notions are still in embryonic form.

We should point out that little or no attention is devoted to statistical questions per se here. It would be of interest to possess, in addition to knowledge concerning, say, how far apart the expected mean reaction times of two opposed models are, the probabilities of Type I and Type II errors under various conditions. This will be easy for certain simple models, but it may be more difficult, even when straightforward in principle, in some of the more recondite cases.

Two additional principles served as guides in these investigations. The first might be referred to as a "minimal systems strategy." That approach seeks to ascertain the broadest classes of model explanation residing in a single type of process or system, as opposed to complex multisystem interactions. For instance, certain types of predictions might be made by a sophisticated parallel model of a single system that are ordinarily associated with a more baroque model based on two or more interacting subsystems. In this way, we emphasize parsimony in theorization. Second, we attempted to restrain potential empirical testing paradigms to as few conditions or phases as possible. The reason is that as the number of experimental conditions grows, so grows the complexity of the participating psychological system. Thus, the "true" theoretical alternative explanations in the presence of a more complicated paradigm may no longer be the relatively simple models with which one began.

Most of the detailed theoretical results in the following chapters are based on our own work, and we have attempted to provide accurate and fairly extensive citations to the large germane theoretical and experimental literature; also a number of interesting theoretical investigations of others are pointed out and discussed in appropriate sections. In some contexts, our results are clearly not the most general of which one could conceive, and it seems likely (and we hope) that more elegant and encompassing representations will be found. Many open questions also remain for future research, and we try to point these out when they arise, if they are not already quite evident.

Although we often take issue with postulates, approaches, or conclusions drawn in past investigations, we wish to acknowledge a deep indebtedness to these authors. Several of the cited scientists were prime motivators of the rich epistemology inherent in the information processing approach to cognition that forms a backdrop to this volume. This book is, in fact, a tribute to their work.

Finally, there are a couple of points pertinent to the future of modeling in psychology. First, the question might be posed as to how the more complicated analytic models that seem to arise each month are going to be investigated for equivalence and distinguishability relations, not to mention the simulation models of almost byzantine complexity that are rapidly proliferating. It is probably safe to say the field is not going to be overrun with candidate approaches, but hopefully there will be some. One very speculative possibility might incorporate the power of computers to explore model spaces in ways that are somehow analogous to their employment in proving mathematical theorems that involve a horrendous number of alternatives, as did the famous four-color map problem.

It is probably fair to say that whereas the number of psychologists who are willing to be called "mathematical psychologists" to their face has not grown appreciably over the past 10 or 15 years, the amount of modeling showing up in the experimental journals appears to have increased substantially. We think this, if true, is very encouraging. To take physics as an example, all physicists learn and employ a great deal of sophisticated mathematics, although there are still certain individuals who stress the theoretical side of matters. Even in the "softer" biological sciences a student typically must take at least some calculus, physics, and chemistry, whereas these courses are probably still rarely required in undergraduate psychology programs.

We will hazard a final bias as concerns the use of scaling techniques in experimental psychology, which has been making inroads in recent years, as in many other disciplines. We believe that scaling can be effective in yielding important geometric information about psychological functions, but is at this point basically a static conceptualization that tends to ignore the rich underlying processing dynamics that must be involved in these functions. Further, scaling, like other measurement-oriented approaches, by its nature more or less lumps all the different internal structures together – for example, the sensory as well as the decision phases in perceptual contexts. This can perhaps be partially ameliorated by first separating out these aspects with a process model and then scaling one of these phases (e.g., the sensory, to yield a set of perceptual dimensions and distances). Nevertheless, the scaling techniques are fraught with pitfalls and, we feel, should be used sparingly and with a concerted effort to incorporate the findings into process-oriented structures.

On a related theme, although accuracy and reaction time can reveal much about a person's internal processing structure, it would be nice if many more dependent variables could be brought into the picture. Factor analysis (or its close and less problematic sister, component analysis; see, e.g., Schönemann 1976) has this property of considering a number of distinct response dimensions. However, it shares with the scaling methodology some of the unfortunate aspects mentioned above. Of considerable import to the information-

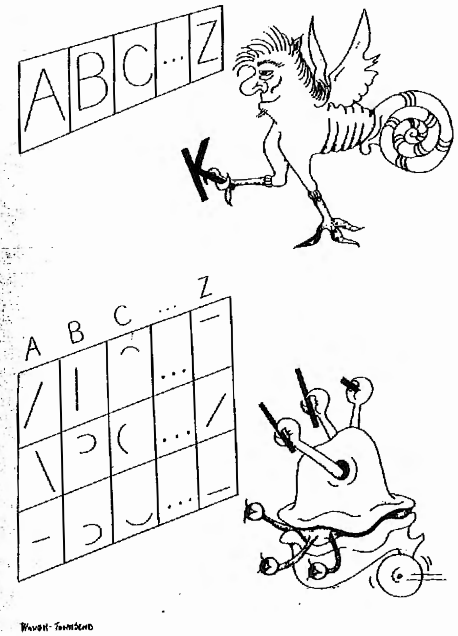
processing field would be the development of a methodology analogous to components analysis or, even better, systems identification as in electrical engineering, designed for the explication of *dynamic* psychological systems that interact and function in real time. Although certain of the developments in the following hark to such lofty goals, the latter presently reside beyond the range of most present-day psychological methodologies.

The altogether singular creatures that inhabit the illustrations at the beginning of each chapter were given life in the fantasy universe of artist Leslie Waugh. We hope they add a touch of humor to the serious business of psychological theorizing; if they also help to clarify a point or two, so much the better.

Acknowledgments

A number of scientists greatly aided the final stages of preparation of this manuscript by providing helpful comments and suggestions, some of them quite detailed. They are Drs. Donald Bamber, William Batchelder, William K. Estes, Jean C. Falmagne, Ulrich Glowalla, David Green, James Juola, Lester Krueger, Steven Link, R. Duncan Luce, Cristof Micko, Roger Ratcliff, and Richard Schweickert. We thank them all.

Several students in the Mathematical Psychology Program at Purdue and at Ohio State assiduously proofed the final manuscript for mathematical errors and math-related typographical mistakes: Ronald Evans, Gary Hu, Douglas Landon, Nancy Perrin, Susan Piotrowski, and Sheue-Ling Hwang. Ronald Evans also helped to prepare those graphs that were done by computer. As mentioned in the Preface, Leslie Waugh kindly supplied the humorous illustrations at the beginning of each chapter. We are indebted to Jan Krizan and Julie McKinzie for the long hours devoted to many aspects of the book's preparation. The staff at Cambridge University Press were eminently helpful throughout the preparation of the book, with special thanks to Susan Milmoe, Rhona Johnson, and Bill Green. Finally, several sources of research funding supported both the creative as well as the more prosaic aspects of the work. The project was begun while the first author (J.T.T.) held a visiting professorship at Technische Universität Braunschweig that was funded by the Deutsche Forschungsgemeinschaft. The second author (F.G.A.) was also a visitor at Braunschweig, supported in part by Purdue Research Foundation Grant #XR0104. More recently the project has been funded by Purdue Research Foundation Grant #XR0422 and NSF Grants #7920298 and #7684053 to Townsend. The penultimate stages of book preparation occurred while J.T.T. was Visiting Scholar at the School of Social Sciences, University of California at Irvine, and F.G.A. held a National Science Foundation Postdoctoral Fellowship at the Department of Psychology and Social Relations, Harvard University.



The issues underlying search or recognition experiments, in which symbols or characters are matched against a memory set of stored items, play an important role in this book. The creature at the top is comparing a letter composed in its entirety (or as a template) against his memory list. The bottom creature, on the other hand, is matching the *K* as a set of features against her memory items – also viewed as a set of features. Later we shall discuss in more detail how such searches might be carried out.

1 Modeling elementary processes: reaction time and a little history

The ambitious goal of providing an elegant and meaningful explanation of mammalian behavior that provides for prediction of future behavior is perhaps the most formidable version of the general black box problem: finding out what goes on in a system whose precise internal construction is unknown, but that interacts with its environment in a more or less observable fashion. In this chapter we shall outline a few concepts relevant to modeling of the nature treated in the remainder of the book.

Figure 1.1 illustrates the basic problem in psychological modeling, at least of any modeling that hopes to do more than simply catalogue stimulus–response correspondences. It can be seen that the organism is conceived of as acting upon the stimulus input from the environment and producing some response. The response can be written as a function $f_{\text{ob}}(S, O)$ of the stimulus and the state of the organism, which we just express as O. The subscript ob stands for observed. The aim of the theorist is to provide a model for the behavior, which we can give as $f_{\text{th}}(S, O)$, where the subscript th stands for theoretical. The model should give sufficient description to f as a function of

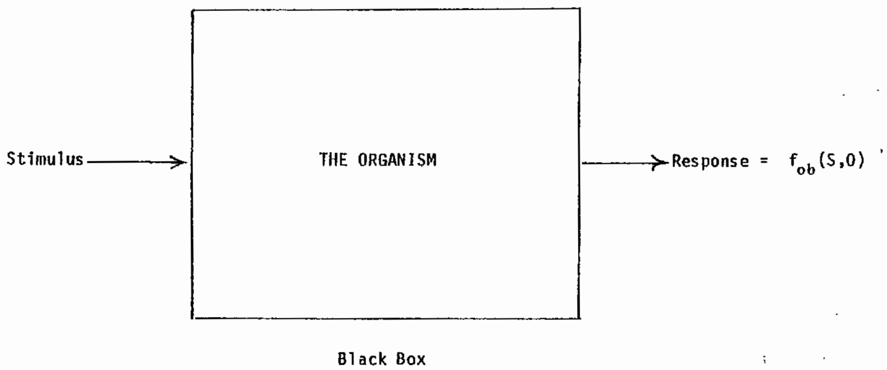


Figure 1.1. Schematic showing organism as "black box" with behavior described by a function of the stimulus and organism, $f_{\text{ob}}(S, O)$.

the stimulus and the state of the organism so that $f_{\text{th}}(S, O)$ will be as close to $f_{\text{ob}}(S, O)$ as possible.

Usually, f will be expressed as a function that depends on parameters (typically real numbers) associated with the stimulus and the hypothetical mechanisms in the organism and associated with its state. Some means of estimating the unknown parameters (some associated with the stimulus may be given by the experimental procedure) must be present. Once these are estimated in such a way as to make the theoretical predictions as close as possible to the data, then some measure of how good the prediction or "fit" is, is desirable. For instance, one such measure is the chi-square (χ^2) statistic,

$$\frac{[f_{\text{th}}(S, O) - f_{\text{ob}}(S, O)]^2}{f_{\text{th}}(S, O)}$$

summed over the conditions of the experiment. It can be employed both in the context of parameter estimation and also as a test of fit. That is, one can estimate the parameters in such a way as to minimize χ^2 and then test if the resulting statistic is statistically significant. If not, the fit is deemed acceptable. More detail on estimation in psychological modeling contexts can be found in Atkinson, Bower, and Crothers (1965), Restle and Greeno (1970), and Bush (1963).

Any number of dependent variables can be associated with the response, of course, just as the range of stimulus possibilities is almost infinitely large and must be tailored to the particular research goals of the investigator.

A natural question arises as to the value of models that are not directly based on the detailed anatomy and physiology of the organism's contributive systems, for example, its nervous and endocrine systems. Our feeling is that natural phenomena can be described at many different levels, with each description being appropriate to some particular level. Thus, at a given level of description of the stimulus and behavior, there will exist a model that describes the data as well as possible and more economically than is feasible at any other level. Actually, a set of models will exist that cannot be tested against one another because for a given set of behaviors, they give equivalent predictions. Part of this book is occupied with such problems. Furthermore, if we were to wait until all the physiology were worked out, even if that would immediately yield appropriate models at the macroscopic level in which we are interested (which is itself problematic), our great-great-grandchildren still might not be in a position to deal with human behavior. A classic case is that of memory; despite all the impressive advances in neurophysiology, the actual mechanism and locations of memory storage are not yet agreed upon.

We should make clear that the question of whether one should attempt construction of models based on neurological concepts is entirely different from that posed in the preceding paragraph. Such modeling employs what *is* known of such processes (frequently together with a liberal amount of speculation) to provide for more or less macroscopic prediction or description, depending on the level of behavior aimed for (e.g., description of interneural spike activity vs. a person trying to detect weak radar signals). Such theorization can sometimes help constrain the mathematical formalisms used in the modeling to a reasonable subset of the vast repository of mathematical tools and structures initially available to the theorist. A problem for the theorist is not too few mathematical possibilities, but often too many to know where to start; alternatively, the latitude in modeling is so great that it is virtually certain that, for instance, an elementary probability model will fit it, a neurological model will fit it, a geometrically motivated model will fit it, and so on, at least given sufficient theoretic assiduity. On the other hand, the present age of psychological theorizing is perhaps akin to the pre-Archimedean age of physics: We know too little to rule out any promising approach.

A significant part of our efforts in this book are directed to aspects and issues of processing associated with the amounts of time taken by various parts of the internal structure to do their job. *Latency* or *latency period* is a term that means an interval of dormancy or unobservability by the usual dictionary definition. In psychology, it attained a more specialized meaning some time ago. It is often used to denote the *reaction time* (RT) or the duration between some specified signal or designated point in time and an experimental observer's response (be it human or animal). It is also used to denote some part of the overall RT, frequently a part that is supposed to represent the duration that some internal process consumes in performing a psychological task of some type (perceptual, cognitive, etc.). We shall use *RT* and *latency* interchangeably, although when writing theoretical expressions involving time, *RT* will be the favored term. In the remainder of this chapter we shall take a short excursion through some of the history of experimental psychology that relates to RT analysis.

Reaction time in the history of experimental psychology

We begin with the name of a man who, although he personally did little explicitly with RT or latency mechanisms, was an important precursor in the employment of mathematics in psychology.[1] Johann Friedrich Herbart (1776–1841) was a philosopher at a time when psychology was still firmly attached to the tree of philosophy, though it was beginning to thrust off fertile independent seeds in America as well as in Europe.

Herbart thought psychology should be (a) empirical; (b) dynamic, in the sense that ideas can vary and interact over time; and (c) mathematical – but not experimental. Thus, one could presumably come to a mathematical understanding of the way the mind functions through observation and informal introspection, but not through controlled experimentation. The latter predisposition did not enamor him to the soon-to-follow experimentally inclined psychologists, and is thought to be largely responsible for his contribution being propagated through later experimentalists rather than through his own apostles. He was instrumental in elucidating and giving mathematical expression to the concept of a mental limen or threshold separating subconscious from conscious ideas and of how conflicting or compatible ideas might interact below or above the threshold of consciousness. In this, he foreran Freud in an obvious way and also Fechner, who attained fame in developing the classical methods of psychophysics – techniques designed to investigate the psychological influences of external stimuli, particularly weak-magnitude stimuli around the threshold of sensation. He also established the term *complication,* which referred to a mental state resulting from stimulation of two or more sensory systems. Of this, we shall see more below.

Curiously, an important contribution to the use of RT in psychology was born out of the firing of an experimental assistant (named Kinnebrook) of the eighteenth-century astronomer Maskelyne in 1796. At that time, the accepted method of measuring the relative movement of heavenly bodies was the so-called eye-and-ear method of Bradley. The telescopic sighting field was divided by a grid of equally spaced lines and the goal was to observe, to within one-tenth of a second, the time at which a given star crossed a given line. After noting the present time within a second's accuracy, the observer began counting seconds along with the beats of a clock. As the star approached the next line, the observer pinpointed the number of seconds that passed just before the star crossed the line, as well as the proportion of the intersecond interval consumed between the last sounded beat and the next clock beat. This proportion was also figured in tenths, so that the overall transit time was roughly calculated, within the accuracy of the human measuring instrument, to a tenth of a second.

Kinnebrook had the misfortune of computing transit times about a second later than Maskelyne did and so was fired for incompetency. A report of this incident was noticed almost 20 years later by the famous astronomer and applied mathematician Friedrich Wilhelm Bessel, who began to make systematic comparisons of the transit times of various astronomers. The difference between two observers came to be called "the personal equation," referring as it did to the emergence of one of the first experimentally examined individual differences on record. These fairly stable personal-equation differences could then be used to correct recorded transit times for separate observers.

[1] Specific references will not be given to the literature by historic figures. However, a good place to begin for readers desiring more detail is E. G. Boring's classic *History of Experimental Psychology* (1957).

Around 1863, Wilhelm Wundt was to combine the ideas of Herbart concerning "complication," which involved information coming over more than one sensory modality, and the empirical data of Bessel and other astronomers involving the personal RT differences. Wundt, with a background in physiology and medicine, is usually accorded the status of being the first experimental psychologist. Later, he and von Tchisch employed a pendulum that swung across a scale and sounded a click at a given point in its excursion, which was to become known as the *complication clock.* Experiments were performed on the "simultaneous" perception of sight and sound. It was found, for example, that that which is perceived first depends on whether attention is placed primarily on the auditory or visual modality. This topic is still being investigated, now with modern experimental and mathematical techniques (see Sternberg & Knoll 1973).

Another branch that led off the personal-equation findings was the so-called reaction experiment in which observers simply responded as rapidly as they could to some predesignated stimulus. The most elementary of these was the *simple reaction time* experiment in which a single predetermined stimulus was followed by a single predetermined response as fast as possible. Even this almost ridiculously simple type of experiment unearthed a good deal of psychological paydirt concerning individual differences and the effects of various types of experimental variables. For instance, it was found that naive observers exhibited longer RTs when asked to focus attention on the stimulus to be presented rather than on the response to be made. It was further discovered that simple RTs were affected by the quality and intensity of the stimuli. The amount of information that an observer had about when the stimulus would be produced was also important and has led to greater knowledge concerning the effects of preparation in recent years (see Thomas 1974).

It was natural enough to next consider the possibility that the durations consumed by various internal psychological processes (e.g., discrimination, judgment, decision) might be wrought observable by subtracting an observer's simple RT from that involving higher mental processes. That is just what F. C. Donders, a Dutch physiologist, set out to do by way of what later came to be called the *method of subtraction.* If various psychological subprocesses are carried out in a serial fashion, that is, in a series with no overlap in processing time, then the method is bound to succeed, as long as the experimental tasks really do eliminate certain of the subprocesses. Such serially arranged psychological systems will be referred to as *Donderian systems* in the present book (see especially Chapters 6 and 12). Of course, a single system or subsystem may be engaged in processing a set of objects (to be called elements) serially, and much of our later discussion will be oriented around such simple systems. If the underlying subprocesses are carried out with an overlap in processing time (that is, if they are active during part or all of the same duration, as in parallel processing), then the method of subtraction will introduce significant error and will typically underestimate the contributions of the subprocess to the overall RT. Attempts to generalize this method to techniques requiring less stringent assumptions will be considered in other chapters, particularly Chapter 12.

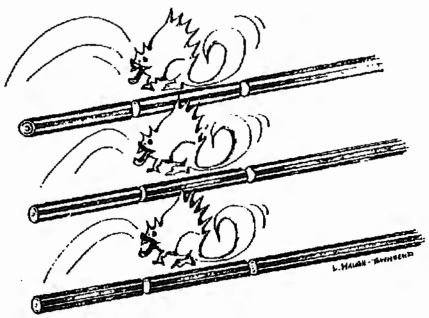
James McKeen Cattell (1860–1944), in addition to innovative work in other areas, employed RT in the study of visual attention, including its span or capacity for simultaneously handling various numbers of objects, for instance, forms, letters, and colors. This type of research carried on a tradition set by the philosopher William Hamilton when, around 1860, he surmised through personal observation that the limits on visual attentional span are about 6 items. The concept of capacity will play an important role in our developments in this book. Cattell's work remains influential in modern researches on human information processing (e.g., reading).

We would be remiss if we did not mention in passing the impact made by the great Hermann von Helmholtz (1821–1894) by his measurement of the conduction time of neural impulses in actual nerves of frog in vitro and in man by stimulating the toe and thigh and noting the difference in simple RT. This was accomplished about 1849 and put the lie to the more preposterous estimated conduction velocities (one physiologist remarked it to be 60 times the speed of light!) as well as gave physiological credence to the use of RT in physiological and psychological contexts.

Although RT and concurrent speculation concerning perception and thought continued to see some use during the time between the early days of experimental psychology (roughly 1860 to 1913) and the 1950s, the more or less cognitively sterile behaviorism was to dominate psychology for almost half a century, except for certain areas of psychophysics and psychophysiology. Then, during the 1950s, undoubtedly influenced by the intriguing lines of development in automata theory (e.g., John von Neumann), cybernetics and communications theory (e.g., Norbert Wiener), and decision theory (e.g., von Neumann and O. Morgenstern), psychologists began to invest the human (and more recently even the animal) organism with powers of thought once again. Thus, centralistic ideas began to appear in the work by quantitatively oriented psychologists in applications of information theory (e.g., Garner 1962; Attneave 1959; Luce 1960), in mathematical learning theory (e.g., Estes 1950; Bush & Mosteller 1955), and in human signal detection (e.g., Tanner & Swets 1954; Green 1960).

By the mid-1960s, a general area known as the *information-processing approach* had become established. It could be broadly defined as a set of experimental techniques and theoretical concepts addressed to the goal of discerning the underlying subprocesses, and their interactions, that are activated in various psychological tasks. One of the foremost methodologies involved in this approach has been that of various forms of RT analysis, including

suggested ways of delineating the latency contributions of these subprocesses. One of the reasons for the embracing of behaviorism and its lengthy ascendancy is thought to be its justification for eschewing the quandaries and paradoxes connected with our philosophical heritage. As fears of the old dilemmas waned and general expertise in modeling with the digital computers continued to wax, the domain of cognitive investigations grew ever broader into formerly relatively unassayed areas of problem solving, psycholinguistics, reading and text comprehension, and memory. The information-processing approach continues to be of use within this broadened spectrum, but it retains its greatest power, so far, in a more elementary cognitive milieu. This book is intended to reside within the general domain of the information-processing approach.



The Saw Dragons participate in a favorite activity: rapidly cutting off sections of boards. Saw Dragons are genetically predestined to perform in a *deterministic* manner. That is, every board is always cut exactly the same length, with absolutely no chance or probabilistic factor involved. A simple way to introduce a chance factor would be to give each Saw Dragon a coin to flip before each new cut. Heads, the length would be 10 meters, and tails, the length would be 5 meters. This would introduce a probabilistic or stochastic factor across time as well as across different Saw Dragons operating at the same time.

2 *Some basic issues and deterministic models of processing*

Deterministic models are the poor but stolid and productive workhorses of theory building. For some purposes, they may be all that is required. For others, they can often be employed to offer early intuition or knowledge about some process. They may also serve as a skeleton structure upon which nondeterministic flesh is draped at a subsequent stage of theorizing.

It is possible to discuss for hours the various meanings, nuances, and philosophical ramifications of the term *deterministic.* We shall avoid these problems by roughly defining this word, for our purposes, as meaning "always giving a fixed result" and "if more than one observation is made on the event in question, that observation will always be precisely the same." These properties are meant to imply that there is no aspect of chance or randomness concerning the particular phenomenon at issue. Put another way, the phenomenon has no variability; or again, it possesses no aspect of probabilism or stochastic quality.

The main "event" to which we apply this term is that of time. For example, if the train operating between Brussels and Frankfurt always takes exactly the same amount of time for the trip, that travel time would be said to be deter-

ministic. Similarly, if the time for a person to make a response indicating a choice between two alternative wines is without fail exactly 3.5 minutes, that time is represented by a deterministic quantity or a "variable" having no variability. It might be that the internal processes leading to the final response are of a stochastic nature, and therefore vary each time the person makes the choice. But if the final resultant time is always 3.5 minutes, this latter quantity will still be deterministic.

We now turn without further ado to an elementary consideration of the most central processing issues with which this book is concerned. It will also be appropriate to introduce the types of time events of special interest here.

Serial vs. parallel processing

When there are several elements to be processed, a number of important issues arise connected with how they are processed. The first is whether they are worked on one at a time (i.e., *serially*), all together (i.e., in *parallel,* simultaneously), or in some other manner. More precisely, parallel processing occurs when processing begins on all elements simultaneously and proceeds until each element is completed (or until all processing is terminated for some reason). Serial processing occurs when processing takes place on one element at a time, each element being completed before the next is begun.

In order to make matters more concrete and to provide the basis for some other concepts, consider the following experiment (as in Atkinson, Holmgren & Juola 1969 or Townsend & Roos 1973 and considered in detail in Chapter 6). Suppose a target symbol of some kind (e.g., a letter) has been shown to an observer so that he or she sees it with perfect accuracy. Suppose next that a multisymbol display is then viewed by the observer for a brief time relative to the duration necessary for an eye movement but long enough to disallow sensory or perceptual errors. This second display is one of two types, a so-called *positive* vs. a *negative* trial. The first contains a copy of the target (sometimes again called the *target* or sometimes the *probe*), whereas the second contains only other symbols, often referred to as *distractors, noise elements,* or *nontargets.* Usually several noise elements accompany the target on a positive trial. The observer makes a positive or "yes" response if the target was present and a negative or "no" response if not. The responses are usually pressing one of two buttons but might be a vocal response timed by a voice-coil timer apparatus. The experimenter records the reaction time (RT) on each type of trial and can then plot a frequency function on RT as well as compute the sample mean, variance, and so on. If, as will usually be the case, we are interested in the latency characteristics of the symbols themselves, they will be our *elements,* usually designated by their position in the display but sometimes by their identity. *Processing* will refer to the comparison of the target element with the second, multielement set.

In much of this book, this and similar paradigms will be used to exemplify our development, although the theory and methodology will usually be applicable to almost any situation where (1) a system or subsystem is processing a finite number of elements, or (2) two or more subsystems are engaged in a similar operation. In most cases, the focus will be on the characteristics of processing taking place on the elements themselves, but sometimes we will emphasize the separate subsystems (as in Chapter 12). For instance, we may be interested in whether comparison of the target element with the displayed multielement set is parallel or serial. More rigorous definitions of these concepts will be given later in Chapter 4. There the focus is on the processing of the elements themselves and not on whether, say, parallel processing is carried out by a single subsystem or by separate subsystems. However, in other cases, particularly when it may be that different functions are carried out by separate subsystems, the investigation will concentrate on those, rather than on the particular elements being processed. Thus, suppose that one subsystem processes the intensity of any visual input while another processes color; then it may be asked whether these two subsystems operate in parallel or serially on a given input (see Chapter 12).

In the remainder of this chapter, we employ the above paradigm and the elements themselves to illustrate some principles that will be employed throughout the book. Accuracy issues will be neglected here for the sake of simplicity, but see Chapters 5, 6, and 9–12.

Suppose for illustration that the second display contains exactly two elements and consider a trial where the element occupying position a finishes processing first, followed by the element in position b (hereafter simply called elements a and b unless otherwise noted). Figure 2.1 illustrates the main temporal concepts required throughout the book. The letter t denotes *intercompletion time,* which is defined as the interval between successive completions, and, as can be seen from the figure, is also an *actual processing time* (duration spent by the system on an element) for some element in a serial system. The intercompletion time will *not* be an actual processing time in parallel processing except on the first element completed. Conversely, the *total completion time* on a particular element is represented by the symbol τ and refers to the duration from $t=0$ until the element is completed, whatever its actual processing time. Thus, as is apparent in Fig. 2.1, the total completion time of an element in a parallel system is equal to its actual processing time, but in a serial system the total completion time of an element is equal to the sum of the actual processing times up to and including that of the considered element. The total completion time is *always* the sum of the number of intercompletion times preceding the completion of the considered element. Note that total completion time does *not* refer to the time necessary to finish all the elements. Finally, a concept that will be required below is that of *stage.* "Stage k" refers to the interval occupied by the kth intercompletion time, that is, by the duration between the $(k-1)$th and kth completion. This is somewhat redundant with "intercompletion time" but will aid in describing the state of the system during a stage when the focus is on the system itself rather than the particular value of time that is the intercompletion time.

A slightly different perspective may help to elucidate these concepts and

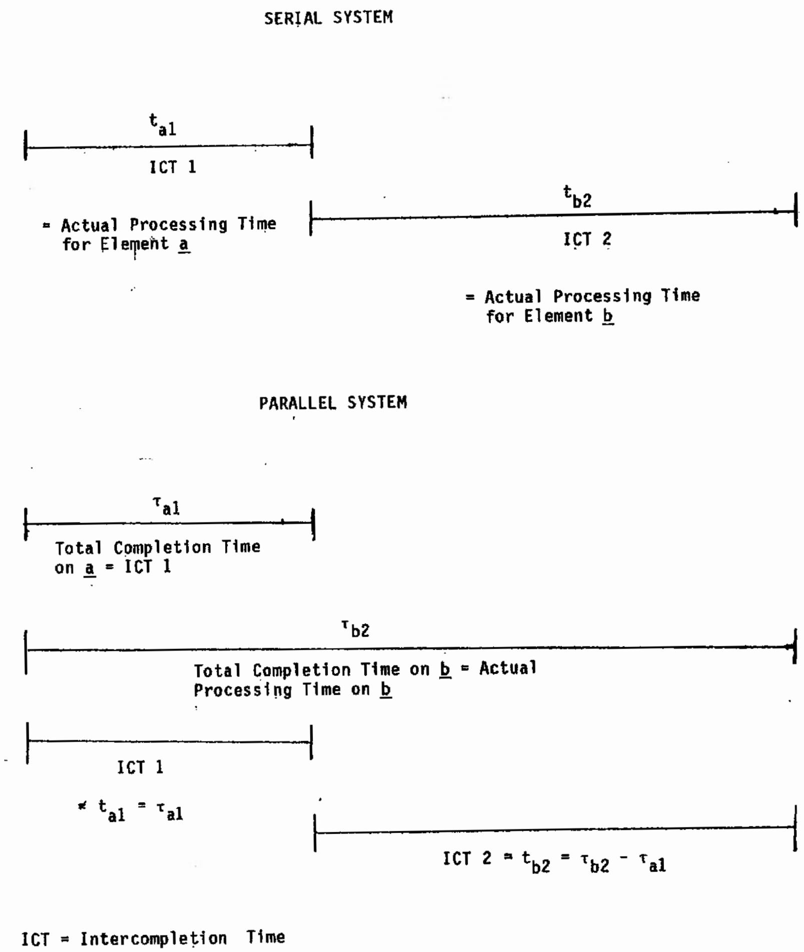


Figure 2.1. Illustration of *intercompletion times, actual processing times,* and *total completion times.*

prepare the reader for the subsequent stochastic developments in Chapters 3 and 4.

When processing on a pair of elements is parallel and deterministic, then, if element a takes duration τ_a to be completed and element b takes time τ_b, then these are always the actual processing times on every trial. Now if (without loss of generality) $\tau_a < \tau_b$, then

$$\min(\tau_a, \tau_b) = \tau_{a1}, \quad \max(\tau_a, \tau_b) = \tau_{b2}$$

as in Fig. 2.1. Note that

$$\tau_{b2} = \tau_{a1} + t_{b2} = t_{a1} + t_{b2}$$

where in this case, t_{a1} and t_{b2} are the first and second intercompletion times, respectively.

If processing is serial and deterministic and a is always processed first, then the same equations result, only the interpretations change! For example, $\max(\tau_a, \tau_b) = \tau_{b2}$, with τ_{b2} being the total completion time. However, now the *actual processing time* of element b is $\tau_{b2} - \tau_{a1} = t_{b2}$ rather than τ_{b2} as it was in the parallel system. Therefore, if we assign a new symbol to represent *actual* processing time, say z (as in Townsend 1974b: 140), then for the serial system

$$z_{a1} = t_{a1}, \quad z_{b2} = t_{b2}$$

and for the parallel system

$$z_{a1} = \tau_{a1}, \quad z_{b2} = \tau_{b2}$$

Unfortunately, of course, z is ordinarily unobservable. The most we can usually hope to record in RT experiments is the intercompletion times. Most often, even these are not available for inspection, so that overall RT is a composite of intercompletion times.

Our reason for introducing our notation in terms of position index instead of the actual elements themselves is that it has been theoretically useful to assume some kind of constancy or invariance of certain aspects of processing with respect to distinct elements. For instance, consider the above visual search paradigm. The practicing experimental psychologist may have some hope of testing models that assume some difference of processing rate according to the various stimulus input positions, but models that assume that processing rate can differ for each and every possible stimulus *letter* will often have little chance of being testable within this paradigm. This does not mean that no experiment can be done involving, indeed based upon, the possibility of different processing rates for distinct elements. It is simply that most theorists and experimentalists have taken the aforementioned approach. One manner in which identity can affect systems engaged in comparing pairs of elements is that the speed of the comparison process might be different on matching elements than on elements that mismatch. On the one hand, this greater generality permits more flexibility of serial and parallel systems, sometimes in ways that place in question conclusions reached with simpler systems (e.g., Townsend & Roos 1973). On the other hand, natural serial and parallel systems can differ in how they are affected by match vs. mismatch comparisons in ways that may be helpful in testing parallel vs. serial models (Townsend 1976a). These ideas will be developed further in later chapters.

Self-terminating vs. exhaustive processing

There are many cognitive tasks involving the processing of more than one thing where the completion of some subset of the things may provide sufficient information for the observer to make the correct response. If the mechanism responsible for the processing is able to stop when this sufficient subset of elements is finished, it is said to be a *self-terminating* processor. This implies, of course, that on the majority of trials processing will be terminated before all of the stimulus pattern is processed. If the system is incapable of halting and must always process the entire stimulus pattern, then it is said to be an *exhaustive* processor. Of course, in many cases the task may force exhaustive processing.

For instance, in a visual search experiment of the type discussed above, suppose the target letter, presented first, is A and the second display is

Position:	a	b
Second stimulus:	A	B

Then, if processing is serial and self-terminating with a processed first, the overall time before the system stops is just t_{a1}, the time to match the target A against the stimulus A in position a. Obviously, with b processed first the time to cessation of processing would be $t_{b1} + t_{a2}$, where the subscripts denote the element position and the stage of processing. Here both elements had to be processed to determine whether the target was present and thus required a positive response.

Self-termination might just as well occur in parallel processing, in which case the time to stoppage is always τ_a. When processing is exhaustive, the time until processing ceases is always $\max(\tau_a, \tau_b)$.

The capacity issue

Capacity refers to how a system reacts with regard to speed and accuracy when its processing load is varied. Most of our emphasis will be on load as represented by n, the number of elements to be processed. The question of capacity may be raised on different levels, even within the same processing context or experimental paradigm. A low or fairly "micro" level of processing might be the finest-grained element or feature assumed in the stimulus (or channel, etc.), and the highest or most "macro" level might be the exhaustive processing of all the grossest-grained elements to be processed. For example, if letters are assumed to be made up of features which serve as "atoms," then the lowest level in a visual search paradigm would be at the level of a single feature and the highest level would be the processing of the entire set of letters in memory. An intermediate level would be self-terminating processing, processing that ceases when the target letter is discovered.

Roughly speaking, if the processing time of an element increases or if accuracy falls when the number of elements (at any particular defined level) is increased, then the system is said to be limited capacity at that particular level. If performance on both these dimensions remains unchanged, we say it is unlimited capacity, and if it should actually improve, it attains the distinction of being supercapacity.

From these informal definitions, we can see that in a serial system in which the values of t_a and t_b do not depend on how many elements are processed, the system is unlimited capacity at the level of the individual element. On the other hand, if t_a or t_b should increase when the system has to process both elements, it is limited capacity at the individual element level.

It should be observed that at the exhaustive or self-terminating level, serial systems will evidence limited capacity effects when two elements rather than one are processed, even though capacity is unlimited at the level of the individual element, since the overall processing time for the two elements is greater than that for only one. The individual element capacity would have to increase in order to predict, for example, unlimited capacity at the exhaustive level.

In deterministic parallel processing, if capacity at the individual level is unlimited, so will it be unlimited at the level of self-terminating or exhaustive processing. This is because the actual processing times do not get added together in parallel processing. On the other hand, it seems more likely, on an intuitive level, that processing will be at a more limited capacity *at the individual element level* in parallel mechanisms than in serial systems. This is because there may be a limited source of processing capacity or energy that must be spread out over the elements or channels to be covered simultaneously. If processing time is inversely related to the processing capacity allocated to an element, the time will obviously increase as more elements are added to be processed. Thus, the amount that can be allocated to any one element will decrease as the total number of elements to be handled gets larger. Clearly, when this occurs, self-terminating or exhaustive processing will also be of limited capacity.

It should be mentioned that changes in capacity at the individual element level can occur across stages. That is, as successive elements are completed, the speeds on the remaining elements may be altered. For instance, we can conceive of a serial mechanism that gets tired as it performs on more and more elements, thus slowing down across stages. Similarly, there may be some experimental circumstances where a priming effect occurs and processing speed actually increases as more elements are completed.

Although a "tiring" or "priming" effect might also occur with parallel processing, an especially important change that could take place in parallel systems is capacity reallocation. Suppose that when an element is completed, the capacity that was devoted to it is now reallocated to the remaining elements. This will cause the remaining intercompletion times to be shortened compared with what they would be without the reallocation property. Reallocation is a concept that has been significant in producing stochastic parallel models that are equivalent to standard types of serial models (Townsend 1969; Atkinson et al. 1969; Townsend 1974b; and see also Chapters 4 and 6).

We should point out finally that the time behavior of any deterministic model can be mimicked by the mean-time properties of a stochastic (nondeterministic) model. The section "Limited vs. unlimited capacity issue" in Townsend (1974b) contains a slightly more advanced statement regarding deterministic parallel models, but one that is completely compatible with these remarks.

Although one occasionally comes across deterministic models in the

Figure 2.2. A set of mental processes in series.

psychological literature, they have not been in the forefront in recent years due to the pronounced random component associated with most behavior. Even so, they may sometimes be of use to describe overall average behavioral characteristics.

Moreover, when a new theoretical enterprise is initiated that goes beyond current theory or methodology in depth and complexity, then a deterministic formulation may be the only feasible path in its early development.

The following section briefly outlines a deterministic theory–methodology recently put forth by Schweickert (1978). It composes an interrelated set of mathematical and experimental techniques designed to uncover the underlying functional processes that take place during performance on a cognitive task. Although there are certain limitations in the theory as it stands, it goes well beyond most approaches in its power to explicate a variety of subprocess interactions. Further, it seems certain to receive further generalization, particularly in imbuing it with probabilistic structure. The basic concepts arise in areas of applied mathematics (e.g., operations research) concerned with optimal scheduling problems, but the major theorems of the theory are due to Schweickert.

Latent network theory

Suppose that to perform a certain task, a set of mental processes such as perceiving and deciding must be executed serially, as illustrated in Fig. 2.2. Also suppose there are manipulations that can be made experimentally, each of which prolongs a single process, while leaving everything else unchanged. A visual perception process may be prolonged, for example, by making the stimulus dimmer. According to Sternberg's (1969a, b) additive factor method (see Chapter 12), the combined effect of prolonging two of the processes will be the sum of the effects of prolonging them individually.

What happens if some processes are executed sequentially and others concurrently (e.g., in parallel), as in Fig. 2.3? Suppose that to perform the task in Fig. 2.3 all the processes illustrated must be completed. Processes, such as A and D, that are not joined by a path can be executed at the same time. But processes such as A and C that are joined by a path must be executed in the order in which they occur on the path. Process C in Fig. 2.3, for example, cannot start until both A and B are finished. We assume, in short, that the processes are partially ordered.[1]

[1] A set S is partially ordered by a relation $\lesssim$ if for all $x, y, z, \ldots$ in S: (i) $x \lesssim x$; (ii) $x \lesssim y$ and $y \lesssim x$ imply $x \approx y$; and (iii) $x \lesssim y$ and $y \lesssim z$ imply $x \lesssim z$. In short, the relation $\lesssim$ is reflexive, antisymmetric, and transitive.

If P is the set of all processes in a certain task, and $X, Y \in P$, we say that X precedes

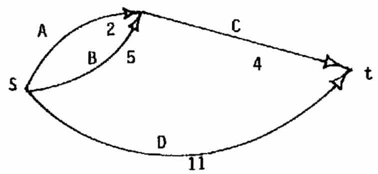


Fig. 2.3. A task network. Each arrow represents a process, and the number on the arrow is the duration of the corresponding process. The stimulus is presented at S and the response is made at t.

The number on each arc in Fig. 2.3 indicates the duration of the corresponding process. The duration of a path is the sum of the durations of all the processes on it. The path going from the start of the task to the end that has the longest duration is called the *critical path.* Since all the processes on the critical path must be executed for the task to be completed, the duration of the critical path is equal to the reaction time. In Fig. 2.3, the critical path consists only of D, which has duration 11; so the reaction time is 11.

Schweickert typically assumes that the durations of the processes are fixed quantities, not random variables. This is, of course, a drastic oversimplification. A stochastic model would be more realistic, but the predictions of such a model are difficult to derive (see Christie & Luce 1956; Fulkerson 1962; Sigal, Pritsker, & Solberg 1980). Until a tractable stochastic model is available, we will have to be content with the approximate predictions of a deterministic model.

Inverting the critical path method

The problem of finding an optimal schedule for a set of processes arises in factories and construction projects, and recently has become an important issue in the design of computer systems. One of the first modern procedures for solving scheduling problems was the critical path method (Kelley 1961; Malcolm, Roseboom, Clark, & Fazar 1959; Modor & Phillips 1970; Wiest & Levy 1977). There are now many scheduling techniques available, and the study of scheduling algorithms is of considerable interest because of the complexity of the problems (Coffman 1976; Conway, Maxwell, & Miller 1967).

As the critical path method is ordinarily used, the partial ordering of the processes and the time required for each process are known, but the time

Y, $X < Y$, if there is a directed path in which X comes before Y. We say that $X \approx Y$ if X and Y are the same process. Finally, we say $X \lesssim Y$ if either $X < Y$ or $X \approx Y$. Then $\lesssim$ is a partial order on P, and the precedence order can be represented by a directed, acyclic graph (Harary, Norman, & Cartwright 1965; Roberts 1976). In Fig. 2.3, for example, $A < C$ and $B < C$; furthermore, $A \lesssim C$ and $B \lesssim C$. It is neither the case that $C \lesssim D$ nor that $D \lesssim C$. The foregoing formulation implies that such networks can be represented ... acyclic graph (Roberts 1976; Harary et al. 1965).

required to complete the entire task is unknown and must be computed. The psychologist has the opposite problem. He or she knows how long it takes to complete the task under various conditions and would like to reconstruct the unknown network as far as possible. This is the goal of latent network theory. A surprising amount of information about the task network can be obtained by observing the effects of prolonging processes.

Slack

Notice that in Fig. 2.3 if the duration of A were increased to 3, process C would not start any later, because C must wait for both A and B to finish. The largest amount of time by which a process X can be prolonged without making process Y start late is called the *slack* from X to Y, written $S(XY)$. The slack from A to C in Fig. 2.3 is 3. The largest amount of time by which a process X can be prolonged without increasing the reaction time is called the *total slack* for X, written $S(Xt)$. In Fig. 2.3, $S(At)=5$.

Suppose process X is prolonged by an amount ΔX. Let $\Delta T(\Delta X)$ be the corresponding increase in reaction time. If ΔX is less than $S(Xt)$, there is no increase in reaction time. If ΔX is greater than $S(Xt)$, then an amount $S(Xt)$ of the prolongation is used up in overcoming the slack, and the reaction time is increased by whatever is left over. If a is a real number, let

$$[a]^+ = 0 \quad \text{if } a \leqslant 0$$
$$\quad = a \quad \text{if } a > 0$$

Then $\Delta T(\Delta X) = [\Delta X - S(Xt)]^+$.

Prolonging two processes

The results in this section are derived in Schweickert (1978) and will be presented here without proof.

If two processes X and Y are not joined by a path, then the effect of prolonging both will be the maximum of the effects of prolonging them individually,

$$\Delta T(\Delta X, \Delta Y) = \max\{\Delta T(\Delta X, 0), \Delta T(0, \Delta Y)\} \tag{2.1}$$

In Fig. 2.3, if A were prolonged by 10, the reaction time would increase by $\Delta T(\Delta A, 0) = 5$. If D were prolonged by 2, the reaction time would increase by $\Delta T(0, \Delta D) = 2$. Now, if A were prolonged by 10 and at the same time D were prolonged by 2, the reaction time would increase by $\Delta T(\Delta A, \Delta D) = 5 = \max\{5, 2\}$.

If two processes, X and Z, are joined by a path, the situation is more complicated. If the prolongations ΔX and ΔZ are not too small, then the combined effect of prolonging both X and Z is

$$\Delta T(\Delta X, \Delta Z) = \Delta T(\Delta X, 0) + \Delta T(0, \Delta Z) + K(XZ) \tag{2.2}$$

where $K(XZ) = S(Xt) - S(XZ)$.

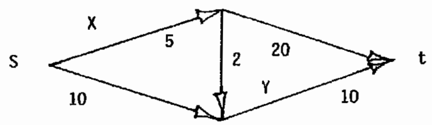


Fig. 2.4. If the coupled slack between X and Y is negative, then they are in a Wheatstone bridge.

The term $K(XZ)$ is called the *coupled* slack for X and Z. Note that the value of $K(XZ)$ does not depend on the values of the prolongations ΔX and ΔZ. This important fact provides a way to test whether a network model applies to a given set of data. For all values of ΔX and ΔZ large enough for Eq. 2.2 to hold, the observed value of $K(XZ)$ should be constant.

In Fig. 2.3, for example, if A is prolonged by $\Delta A = 10$, the increase in reaction time is $\Delta T(\Delta A, 0) = 5$. If C is prolonged by $\Delta C = 20$, the reaction time is increased by $\Delta T(0, \Delta C) = 18$. Now if A is prolonged by 10 and C by 20, the increase in reaction time is $\Delta T(\Delta A, \Delta C) = 25$. Therefore,

$$\Delta T(\Delta A, \Delta C) - \Delta T(\Delta A, 0) - \Delta T(0, \Delta C) = 2 = K(AC)$$

In this case, the difference between the combined effect of prolonging A and C and the sum of the effects of prolonging A and C individually is 2. The reader can check that using larger values for the prolongations ΔA and ΔC will still yield a value of 2 for $K(AC)$. This is because the value $K(AC) = S(At) - S(AC) = 5 - 3 = 2$ is related to certain slacks in the network, but not to the prolongations.

The Wheatstone bridge

The coupled slack $K(XY)$ can sometimes be negative. In Fig. 2.4, if X is prolonged by 10 and Y by 20, the combined effect of both prolongations is less than the sum of the individual effects by 3, that is,

$$\Delta T(\Delta X, \Delta Y) - \Delta T(\Delta X, 0) - \Delta T(0, \Delta Y) = -3 = K(XY)$$

An experimental measurement of a negative coupled slack reveals a great amount of information about the task network. The following result, due to Schweickert (1978), shows that a negative value of coupled slack occurs if and only if (a) the task network contains a subnetwork of the shape illustrated in Figs. 2.4 and 2.5, called a *Wheatstone bridge;* and (b) certain relationships hold among the path durations. The statement of the result involves several details about the constraints on the path durations, which the reader may want to skip for now. The important point for our present purposes is that a negative coupled slack always implies the presence of a Wheatstone bridge structure. A Wheatstone bridge by itself will not necessarily yield a negative coupled slack, as the reader can verify by using 20 as the baseline duration of X, instead of 5, in Fig. 2.4. It is the Wheatstone bridge together with the path duration relationships that lead to a negative coupled slack.

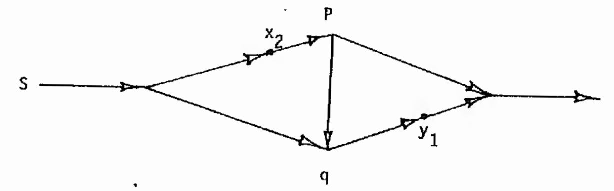


Fig. 2.5. A Wheatstone bridge. If $K(XY) < 0$, y_1 is not on the longest path from s to y_1. x_2 is the terminal of process X and y_1 is the starting point of process Y.

The reader is encouraged to work out a few examples to get a feel for the behavior of the network.

If X is a process, we will denote the starting point of X by x_1 and the terminating point by x_2. If u and v are points, we will denote the duration of the path between u and v that has longest duration as $\delta(uv)$. Proposition 2.1 is illustrated in Fig. 2.5.

Proposition 2.1: Suppose X precedes Y. Then $K(XY) < 0$ if and only if all the following conditions hold:

(i) The longest path from x_2 to y_1 is not contained in the longest path from x_2 to t; let p be the last point preceding y_1 to be on both paths.

(ii) The longest path from x_2 to y_1 is not contained in the longest path from s to y_1; let q be the first point following x_2 to be on both paths.

(iii) $p \neq q$ and $\delta(st) - \delta(sq) - \delta(pt) + \delta(pq) < 0$.

Effects of small prolongations

Equation 2.2 describes the effects of prolonging two processes X and Y joined by a path only when the prolongations are not too small. The following equation applies to processes X and Y joined by a path, for *all* values of prolongation ΔX and ΔY:

$$\Delta T(\Delta X, \Delta Y) = \max\{\Delta T(\Delta X, 0), [\Delta Y - S(Yt) + [\Delta X - S(XY)]^+]^+\} \quad (2.3)$$

Equation 2.2 is a special case of the above equation, for if $\Delta X \geqslant \max[S(Xt), S(XY)]$ and if $\Delta Y \geqslant \max[S(Yt), S(Yt) - S(Xt) + S(XY)]$, then Eq. 2.3 can be shown to reduce to Eq. 2.2.

A peculiar situation can occur if X precedes Y in a Wheatstone bridge. Suppose $K(XY) = S(Xt) - S(XY)$ is negative. Then if $\Delta X < S(XY)$, Eq. 2.3 becomes

$$\Delta T(\Delta X, \Delta Y) = \max\{\Delta T(\Delta X, 0), [\Delta Y - S(Yt)]^+\}$$

that is,

$$\Delta T(\Delta X, \Delta Y) = \max\{\Delta T(\Delta X, 0), \Delta T(0, \Delta Y)\}$$

But this equation is the same as Eq. 2.1, for processes not joined by a path.

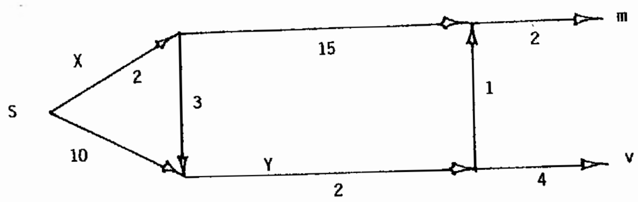


Fig. 2.6. A task in which the observer makes one response at m and another at v. The fact that X precedes Y can be deduced by observing the changes in each response time when X and Y are prolonged.

Equation 2.3 will also have the above form if $\Delta Y < S(Yt) - S(Xt) + S(XY)$. Under certain circumstances, then, sequential processes in a Wheatstone bridge behave like processes not joined by a path. This mimicking of nonsequential processes by sequential processes, when processes are prolonged, is analogous to the mimicking of the parallel processes by serial processes discussed throughout this book. (See especially Chapters 4 and 12–15.)

Determining execution order

Suppose X and Y are two processes joined by a path. If the observer makes two responses, say one verbal and one manual, then we can find out which process, X or Y, comes first. Suppose X precedes Y, which precedes both responses (see Fig. 2.6). If the prolongations of X and Y are not too small, then

$$\Delta T_m(\Delta X, \Delta Y) - \Delta T_m(0, \Delta Y) = \Delta T_v(\Delta X, \Delta Y) - \Delta T_v(0, \Delta Y) \quad (2.4)$$

where the subscripts m and v represent the manual and verbal reaction times, respectively.

The order of X and Y can be determined, because Eq. 2.4 is not symmetrical in X and Y. If Y preceded X, the appropriate equation would be

$$\Delta T_m(\Delta X, \Delta Y) - \Delta T_m(\Delta X, 0) = \Delta T_v(\Delta X, \Delta Y) - \Delta T_v(\Delta X, 0) \quad (2.5)$$

It is possible that the above equation and Eq. 2.4 will both hold. This happens if, for example $K_m(XY) = K_v(XY) = 0$. But if one version holds and not the other, then the execution order of X and Y is revealed. If neither version holds, then a critical path model is invalid for the data.

In the network in Fig. 2.6, if X is prolonged by 10 and Y by 15, Eq. 2.4 holds, while 2.5 does not. This indicates that X precedes Y.

An example

Suppose an observer in a hypothetical experiment is presented with two digits side by side. If the number on the left divides the number on the

Table 2.1. *Reaction times and changes in reaction times in a hypothetical experiment*

			Reaction times		Changes in reaction times	
Clarity	Divisor	Prolongations	T_m	T_v	ΔT_m	ΔT_v
Clear	Yes	Baseline	260	420	0	0
Blurred	Yes	$\Delta_1 X$	320	620	60	200
Clear	No	ΔY	660	720	400	300
Blurred	No	$\Delta_1 X\ \Delta Y$	720	780	460	360
Very blurred	Yes	$\Delta_2 X$	520	820	260	400
Very blurred	No	$\Delta_2 X\ \Delta Y$	920	980	660	560

right without leaving a remainder, he or she is to press a button with the right index finger. Otherwise, instructions are to press a different button with the right middle finger. Meanwhile, the observer is to say "odd" or "even" depending on whether the left-hand digit is odd or even. For simplicity, we will assume that the observer is just as fast to respond "odd" as "even."

Suppose two factors in the experiment affect the reaction times. The first factor is the clarity of the digit on the left. It is clear sometimes, blurred sometimes, and very blurred the rest of the time. Let X be the mental process prolonged when the digit is blurred. Let $\Delta_1 X$ be the amount by which X is prolonged when the digit is blurred, and let $\Delta_2 X$ be the prolongation when the digit is very blurred. The second factor is whether the digit on the left divides the one on the right without a remainder or with a remainder. Suppose a process Y is prolonged when the left-hand digit is not a divisor of the right-hand digit, and let ΔY be the amount of the prolongation when this occurs.

The results of this hypothetical experiment are given in Table 2.1. The manual and verbal response times are denoted T_m and T_v, respectively.

Consider the manual reaction times first. Equation 2.2 describes the combined effects of blurring the left-hand digit and having it be a nondivisor of the right-hand digit,

$$\Delta T_m(\Delta_1 X, \Delta Y) = \Delta T_m(\Delta_1 X, 0) + \Delta T_m(0, \Delta Y) + K_m(XY)$$

i.e., $460 = 60 + 400 + 0$. Since Eq. 2.2 holds, we conclude that X and Y are on a path together. A way to check this idea is to consider the effects of blurring the left-hand digit even more, and we find

$$\Delta T_m(\Delta_2 X, \Delta Y) = \Delta T_m(\Delta_2 X, 0) + \Delta T_m(0, \Delta Y) + K_m(XY)$$

i.e., $660 = 260 + 400 + 0$. As we would expect, the calculated value of $K_m(XY)$ has the same value, 0, for both levels of prolongation, $\Delta_1 X$ and $\Delta_2 X$. If this

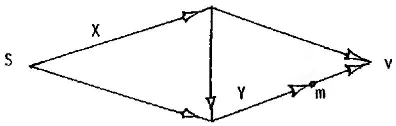


Fig. 2.7. A network for a hypothetical task. Process X is prolonged when the left-hand stimulus digit is blurred, and process Y is prolonged when the digit on the left divides the digit on the right leaving a remainder.

equality fails to hold, we would conclude either that the prolongations were not long enough for Eq. 2.2 to hold, so that Eq. 2.3 should hold, or that a critical path model is not valid for these data.

Now consider the verbal reaction times. We concluded from the manual reaction times that X and Y are on a path together, so we would expect Eq. 2.2 to hold for the verbal reaction times. For $i=1,2$ we expect

$$\Delta T_v(\Delta_i X, \Delta Y) = \Delta T_v(\Delta_i X, 0) + \Delta T_v(0, \Delta Y) + K_v(XY)$$

For $\Delta_1 X$, $360=200+300-140$, and for $\Delta_2 X$, $560=400+300-140$. Equation 2.2 holds, and the calculated value of $K_v(XY)$ is the same for both $\Delta_1 X$ and $\Delta_2 X$, as the theory requires, so there is good evidence that X and Y are on a path together.

Since $K_v(XY)=-140$ is negative, we know that X and Y are in a Wheatstone bridge (see Fig. 2.7). Which comes first, X or Y? Suppose Y precedes X. Then the manual and verbal reaction times should be related through Eq. 2.5: for $i=1,2$,

$$\Delta T_m(\Delta_i X, \Delta Y) - \Delta T_m(\Delta_i X, 0) = \Delta T_v(\Delta_i X, \Delta Y) - \Delta T_v(\Delta_i X, 0)$$

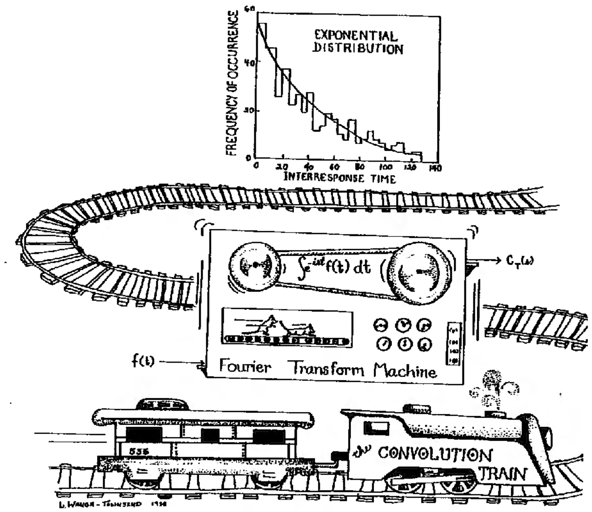
But for $\Delta_1 X$, $460-60 \neq 360-200$, and for $\Delta_2 X$, $660-260 \neq 560-400$. This indicates that Y does not precede X.

If X precedes Y, we expect Eq. 2.4 to hold; for $i=1,2$,

$$\Delta T_m(\Delta_i X, \Delta Y) - \Delta T_m(0, \Delta Y) = \Delta T_v(\Delta_i X, \Delta Y) - \Delta T_v(0, \Delta Y)$$

And indeed, for $\Delta_1 X$, $460-400=360-300$, and for $\Delta_2 X$, $660-400=560-300$. Since Eq. 2.4 holds, we conclude that X precedes Y. If neither Eq. 2.4 nor Eq. 2.5 held, we would have concluded that a network model is not valid for these data.

A network representing this hypothetical task is given in Fig. 2.7. The stimuli are presented at point s, the manual response is made at point m, and the verbal response is made at point v.



Three concepts that will be dealt with in this chapter: The smooth curve in the graph at the top is of the important *exponential* density function (which establishes uniquely the exponential distribution or probability law). The bar graph shows data from neural interresponse times (from Hunt & Kuno 1959), which obviously seem to follow the exponential curve. The *Fourier transform* is a method of significant usefulness in (among others) probability theory and systems theory. We will focus on its use in probability theory in this chapter. Finally, a *convolution* is what you get when you figure out the probability density for the sum of two independent probabilistic variables. It also has a different interpretation in systems theory that will show up in Chapter 12.

3 *Mathematical tools for stochastic modeling*

The assumption of absolute knowledge of any physical system reflects a naive optimism about man's comprehension of nature. Any model based on such an assumption, that is, any completely deterministic model, is bound to mispredict the true behavior of the system to some extent. Generally the magnitude of this misprediction increases directly with the complexity of the system. Even the most elementary cognitive activities are exceedingly complex events, and thus even the most naive cognitive models must include some aspect of indeterminism. Rather than develop models that attempt on every occasion to predict completion and intercompletion times exactly, we will henceforth concentrate on models that attempt to determine probabilities on the possible range of completion and intercompletion times one might expect on a given trial.

Developing such models will require some mathematical machinery, machinery which might not be equally familiar to all readers. In this chapter we shall review some selected topics from probability theory that will come up repeatedly throughout the rest of the book. The reader well versed in probability theory may wish to skip to the next chapter or might just move to the summary at the end of this chapter in order to obtain an idea as to the kind of tools our development requires. Slightly less experienced readers may wish to skim the bulk of the chapter, carefully reading only those sections containing unfamiliar material. Finally, the reader unfamiliar with probability theory will do well to spend some time with this chapter, perhaps supplementing it with an outside source such as Parzen (1960), because a ready familiarity with the definitions and results of the chapter will make the reading of the rest of this book a much more rewarding and less frustrating experience.

With the obligatory recommendations and cautions out of the way, we are ready to begin our review.

Density and distribution functions

We begin with the task of determining a convenient means of expressing our probabilistic intentions. We are specifically interested in determining the probability that the system completes processing on some input during some interval of time. Therefore, unless explicitly stated, we will assume that the variable $t \geqslant 0$. However, in some cases we will want to keep the discussion as general as possible and allow t to vary over the entire real line $-\infty < t < +\infty$, and later specialize it to the positive real line if we wish. To keep the nomenclature simple, we shall nevertheless refer to $t \in (-\infty, +\infty)$ as the *time domain*. With regard to t as latency or time, we are interested in determining

$$P(\mathbf{T} \leqslant t) = F(t), \quad t \geqslant 0$$

where P is a probability measure and $F(\cdot)$ is a cumulative distribution function, *distribution function* for short. $\mathbf{T}$ is a random variable representing the random time until the completion of processing. $F(t)$ then gives the probability that completion occurs at a time less than or equal to t. Next we define the *density function*. In most cases of interest to us, $F(t)$ is differentiable everywhere on the positive real line $t \in (0, +\infty)$. In such cases,

$$f(t) = \frac{dF(t)}{dt} = \lim_{\Delta t \to 0} \frac{P(t < \mathbf{T} \leqslant t + \Delta t)}{\Delta t} \tag{3.1}$$

and $f(t)$ is called the probability density function.[1]

[1] Strictly, there is a whole class of densities f that are equivalent except on a set of measure 0. These densities are all Lebesgue-integrable and yield the same $F(t)$ for any $t \geqslant 0$. In our investigations, we can select one $f(t)$ that is continuous over a desired interval, and it can be shown that no other Lebesgue-integrable f that produces the same F is continuous. Finally, the notion of a "random variable" is briefly presented in the context of measure theory in Chapter 14. Here it is assumed the reader has at least been introduced to the idea as it appears in elementary probability texts.

The density function $f(t)$ tells us how the completion probabilities change over time. For instance, to compute the probability that the system completes processing somewhere between the times t_1 and t_2, we only have to integrate the density function $f(t)$ over this same interval

$$P(t_1 < \mathbf{T} \leqslant t_2) = F(t_2) - F(t_1) = \int_{t_1}^{t_2} f(t)\, dt \tag{3.2}$$

It follows that

$$P(0 < \mathbf{T} < +\infty) = \lim_{t \to +\infty} F(t) = \int_0^{\infty} f(t)\, dt = 1$$

Often Eq. 3.1 will be the easiest way to determine the density function $f(t)$ of $\mathbf{T}$. That is, it may frequently turn out that the distribution function $F(t)$ is much easier to calculate directly than is the density function $f(t)$. Once the distribution function is known, the density function can easily be derived through Eq. 3.1.

An example depicting some of these relationships is shown in Fig. 3.1. Figure 3.1a shows one possible example of a density function of $\mathbf{T}$, whereas Fig. 3.1b depicts the associated distribution function. Note that the area of the shaded region in Fig. 3.1a is $F(t_1)$ and that therefore the area between t_1 and t_2 is equal to $F(t_2) - F(t_1)$, corresponding to Eq. 3.2.

Thus, using only the distribution function, we can compute the probability that processing is completed before any time t or that completion occurs within the interval (t_1, t_2) for any t and any $t_1 \leqslant t_2$. It is also easy to imagine instances where we might like to compute the probability that processing has *not* been completed be some time t, that is, that completion occurs after time t. This can be done by setting $t_1 = t$ and $t_2 = \infty$ in Eq. 3.2. Thus

$$\begin{aligned} P(\mathbf{T} > t) &= P(t < \mathbf{T} < \infty) = \int_t^{\infty} f(t')\, dt' \\ &= \int_0^{\infty} f(t)\, dt - \int_0^{t} f(t')\, dt' \\ &= 1 - F(t) = \bar{F}(t) \end{aligned} \tag{3.3}$$

The function $\bar{F}(t)$ is often referred to as the survivor function in applied probability and reliability theory. Thus one might say that $\bar{F}(t)$ is the probability that a component has survived (i.e., has not failed) to time t. We shall retain this standard terminology even though "survivor" does not really relate to anything in the present field of investigation. In our scheme, the survivor function indicates the probability that completion has not yet occurred.

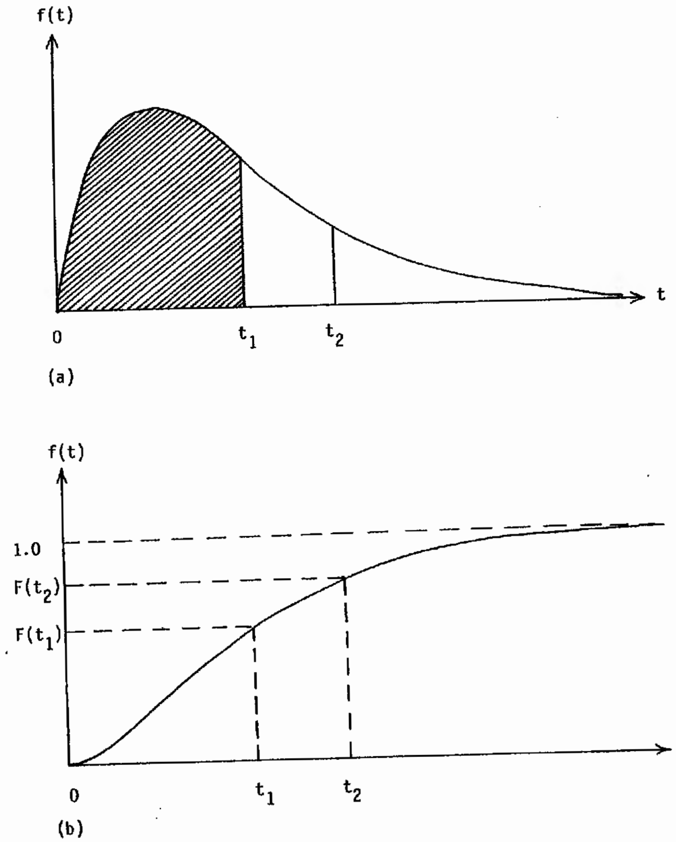


Fig. 3.1. Example of a density function (a) and its corresponding distribution function (b).

Another function which we will find extremely useful, especially in our discussions of capacity, is the conditional probability that processing will be completed in the next instant given it has not yet been completed,

$$\lim_{\Delta t \to 0} \frac{P(t<\mathbf{T}\leqslant t+\Delta t \mid \mathbf{T}>t)}{\Delta t}$$

Using the definition of conditional probability, it can be rewritten as

$$\begin{aligned}\lim_{\Delta t \to 0} \frac{P(t<\mathbf{T}\leqslant t+\Delta t \mid \mathbf{T}>t)}{\Delta t} &= \lim_{\Delta t \to 0} \frac{P(t<\mathbf{T}\leqslant t+\Delta t \cap \mathbf{T}>t)}{P(\mathbf{T}>t)\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{P(t<\mathbf{T}\leqslant t+\Delta t)}{[1-F(t)]\Delta t} \\ &= \frac{f(t)}{\bar{F}(t)} = H(t) \end{aligned} \tag{3.4}$$

The function $H(t)$ is utilized extensively in reliability and renewal theory and is usually called the hazard function or intensity function but is sometimes referred to as the conditional rate of failure function or the age-specific failure rate (e.g., Cox 1962; Gumbel 1958). For us, it is the conditional rate of completion or age-specific completion rate, but we shall continue the traditional name *hazard function*.

Note that if we integrate both sides of Eq. 3.4, we obtain

$$-\ln[\bar{F}(t)] = \int_0^t H(t')\,dt'$$

The integration constant is zero in this case, since we assume $F(0)=0$. Thus, given the hazard function we can always obtain the survivor function from

$$\bar{F}(t) = \exp\left[-\int_0^t H(t')\,dt'\right] \tag{3.5}$$

and therefore the distribution function can be written as

$$F(t) = 1-\exp\left[-\int_0^t H(t')\,dt'\right] \tag{3.6}$$

Finally, by differentiation of $F(t)$, the density is found to be

$$f(t) = H(t)\exp\left[-\int_0^t H(t')\,dt'\right] \tag{3.7}$$

Equations 3.5, 3.6, and 3.7 look suspiciously like the survivor, distribution, and density functions of an exponentially distributed random variable (which we will discuss shortly). They are not, though, unless the hazard function is a constant (i.e., does not vary with time). Nevertheless, this indicates that we can conceive of *any* stochastic process on a random event as a sort of time-varying exponential process.

Bivariate distributions

More often than not, the random variable $\mathbf{T}$, representing the completion time of some input, will be considered as some function of other random times. For instance, suppose we are interested in the maximum of two random times $\mathbf{T}_1$ and $\mathbf{T}_2$ (this concept will be related to exhaustive parallel processing in the next chapter). Then the random variable $\mathbf{T}$ is equal to the larger of $\mathbf{T}_1$ and $\mathbf{T}_2$, and in this case $F(t)=P(\mathbf{T}\leqslant t)=P(\mathbf{T}_1\leqslant t \cap \mathbf{T}_2\leqslant t)$, that is, the probability that $\mathbf{T}$ is less than or equal to t equals the probability that $\mathbf{T}_1$ is less than t *and* $\mathbf{T}_2$ is also less than t. This joint probability is an example of what is known in probability theory as a joint distribution function on the random times $\mathbf{T}_1$ and $\mathbf{T}_2$, and is formally defined by

$$F(t_1, t_2) = P(\mathbf{T}_1 \leqslant t_1 \cap \mathbf{T}_2 \leqslant t_2) \tag{3.8}$$

where in our cases, $0 \leqslant t_1, t_2 < \infty$.

Joint distribution functions have many properties similar to those of the univariate distribution function. For instance, the probability that $\mathbf{T}_1$ is greater than t_a but less than or equal to t_b, whereas $\mathbf{T}_2$ is less than or equal to t_c may be expressed as

$$P(t_a < \mathbf{T}_1 \leqslant t_b \cap \mathbf{T}_2 \leqslant t_c) = P(\mathbf{T}_1 \leqslant t_b \cap \mathbf{T}_2 \leqslant t_c) - P(\mathbf{T}_1 \leqslant t_a \cap \mathbf{T}_2 \leqslant t_c)$$
$$= F(t_b, t_c) - F(t_a, t_c)$$

The joint density function $f(t_1, t_2)$ can be found from the joint distribution function by differentiating $F(t_1, t_2)$ successively with respect to t_1 and t_2:

$$f(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} F(t_1, t_2) \tag{3.9}$$

Alternatively, if we have the joint density and we wish the corresponding distribution function we can integrate:

$$F(t_1, t_2) = \int_0^{t_2} \int_0^{t_1} f(t_1', t_2')\, dt_1'\, dt_2'$$

The one-dimensional density functions $f(t_1)$ and $f(t_2)$ are called the *marginal density functions* of $\mathbf{T}_1$ and $\mathbf{T}_2$. They can be derived from the joint density by integrating out the unwanted variable. Thus,

$$f(t_1) = \int_0^{\infty} f(t_1, t_2)\, dt_2 \quad \text{and} \quad f(t_2) = \int_0^{\infty} f(t_1, t_2)\, dt_1 \tag{3.10}$$

These definitions of marginal densities also illustrate the important property that the area under any joint density function integrates to 1. To see this, we need only integrate both sides of Eq. 3.10 over all values of t_1. Thus

$$1 = \int_0^{\infty} f(t_1)\, dt_1 = \int_0^{\infty} \int_0^{\infty} f(t_1, t_2)\, dt_2\, dt_1$$
$$= \int_0^{\infty} \int_0^{\infty} f(t_1, t_2)\, dt_1\, dt_2$$

A very important concept, especially with parallel processing systems, is the idea of independence. As we shall see later, there are many different kinds of independence, but what we have in mind here is the question of whether or not the completion times $\mathbf{T}_1$ and $\mathbf{T}_2$ are independent random variables. This would be the case, for instance, if knowledge of the completion time $\mathbf{T}_1$ carries with it no information about the completion time $\mathbf{T}_2$. According to the theory of probability the random times $\mathbf{T}_1$ and $\mathbf{T}_2$ are independent if and only if

$$f(t_1, t_2) = f(t_1) f(t_2 \mid t_1) = f(t_1) f(t_2), \quad 0 \leqslant t_1, t_2 < \infty$$

This result gives us a fairly straightforward way to check the independence of any two random times.

If the random variable $\mathbf{T}$ that motivated this section is the maximum of more than two random times, then the joint density and distribution functions will be more than two-dimensional. Even so, all the properties stated above will still hold. For instance,

$$F(t_1, t_2, \ldots, t_n) = P(\mathbf{T}_1 \leqslant t_1 \cap \mathbf{T}_2 \leqslant t_2 \cap \cdots \cap \mathbf{T}_n \leqslant t_n)$$

and

$$f(t_1, t_2, \ldots, t_n) = \frac{\partial^n}{\partial t_1 \partial t_2 \cdots \partial t_n} F(t_1, t_2, \ldots, t_n)$$

whereas the marginals can now be found by

$$f(t_1) = \int_0^{\infty} \int_0^{\infty} \cdots \int_0^{\infty} f(t_1, t_2, \ldots, t_n)\, dt_2\, dt_3 \cdots dt_n$$

and so on.

Mathematical expectations

There will be many times in the chapters ahead when we will not be specifically interested in examining the RT density function predicted by some model. Instead we may wish to study mean RT or RT variance predictions. Suppose we conduct an experiment where we repeatedly record the completion time of processing on an input and that when the experiment is over we calculate the mean of all these observations. In an ideal experiment, with an infinite number of trials, this sample mean would equal the population mean. In probability theory, this population mean is also called the expected value of $\mathbf{T}$, denoted by $E(\mathbf{T})$, because it is one way of trying to summarize the probability density function $f(t)$ by a single number representing a typical value of the random completion time $\mathbf{T}$.

The sample mean is just a weighted sum of the observed completion times, where the coefficients are the relative frequencies of occurrence. That is,

$$\bar{T} = \sum_{i=1}^{k} \frac{n_i}{n} t_i = \sum_{i=1}^{k} P_i t_i$$

where t_i is the center of the ith time "bin" and k = total number of "bins"; n_i = number of observations in the ith bin; n = total number of observations; and P_i = proportion of observations in the ith bin. The population mean, or expected value of $\mathbf{T}$, is basically the same thing, although in this case the relative frequencies are specified by the density function $f(t)$. Thus, given the density we can compute the expected value of $\mathbf{T}$ from

$$E(\mathbf{T}) = \int_0^{\infty} t f(t)\, dt \tag{3.11}$$

Often, we will want to compute the expectation of a slightly more complicated random variable such as $\mathbf{T} = a\mathbf{T}_1 + b$ or $\mathbf{T} = \mathbf{T}_1^2$. Problems like this are easily solved. In fact, in a straightforward fashion we can find the expectation of any function of the random time $\mathbf{T}_1$ (or of $\mathbf{T}$). Suppose $A(\mathbf{T})$ is some arbitrary function of the completion time $\mathbf{T}$. Then

$$E[A(\mathbf{T})]=\int_0^\infty A(t)f(t)\,dt \tag{3.12}$$

This last definition leads us to an algebra of expectations. We can use the facts we know about density functions and about integration to derive some very useful properties of expectations. In the derivations below assume c is any arbitrary constant.

Proposition 3.1: $E(c)=c$.

Proof: $E(c)=\int_0^\infty cf(t)\,dt=c\int_0^\infty f(t)\,dt=c$. □

Proposition 3.2: $E(c\mathbf{T})=cE(\mathbf{T})$.

Proof: $E(c\mathbf{T})=\int_0^\infty ctf(t)\,dt=c\int_0^\infty tf(t)\,dt=cE(\mathbf{T})$. □

Proposition 3.3: Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be two random times with joint density function $f(t_1,t_2)$; then $E(\mathbf{T}_1+\mathbf{T}_2)=E(\mathbf{T}_1)+E(\mathbf{T}_2)$.

Proof:

$$\begin{aligned}E(\mathbf{T}_1+\mathbf{T}_2)&=\int_0^\infty\int_0^\infty (t_1+t_2)f(t_1,t_2)\,dt_1\,dt_2\\&=\int_0^\infty\int_0^\infty t_1f(t_1,t_2)\,dt_2\,dt_1+\int_0^\infty\int_0^\infty t_2f(t_1,t_2)\,dt_1\,dt_2\\&=\int_0^\infty t_1f(t_1)\,dt_1+\int_0^\infty t_2f(t_2)\,dt_2=E(\mathbf{T}_1)+E(\mathbf{T}_2)\quad\square\end{aligned}$$

Other properties could be derived, but in many cases they would be simple derivatives of one or more of these three. For instance, using all three properties we can easily show that $E(a\mathbf{T}+b)=aE(\mathbf{T})+b$, where a and b are constants.

The convolution integral and transform methods

There will be many times in the chapters to follow when we shall be interested in the sum of two or more *random* times. For instance, suppose a serial system processes two elements, taking $\mathbf{T}_1$ time units for the first and $\mathbf{T}_2$ time units for the second, where $\mathbf{T}_1$ and $\mathbf{T}_2$ are independent. Now the total time the system is in operation is the random time $\mathbf{T}=\mathbf{T}_1+\mathbf{T}_2$. We have already seen how to calculate the expectation of $\mathbf{T}$, but now we are interested in finding its density function in situations where we know the densities of $\mathbf{T}_1$ and $\mathbf{T}_2$. This problem has been thoroughly studied, and it is well known that $f(t)$, the density function of $\mathbf{T}$, is given by the so-called convolution of $f_1(t)$ and $f_2(t)$:[2]

$$\begin{aligned}f(t)&=\int_{-\infty}^\infty f_1(t_0)f_2(t-t_0)\,dt_0\\&=\int_{-\infty}^\infty f_1(t-t_0)f_2(t_0)\,dt_0=f_1(t)*f_2(t)\end{aligned}\tag{3.13}$$

The asterisk is a common abbreviation for the convolution operation.

The convolution integral can be simplified somewhat when the random variables $\mathbf{T}_1$ and $\mathbf{T}_2$ represent processing times, since these must always be greater than or equal to zero. In this case

$$\begin{aligned}f_1(t)*f_2(t)&=\int_0^\infty f_1(t_0)f_2(t-t_0)\,dt_0\\&=\int_0^t f_1(t_0)f_2(t-t_0)\,dt_0\end{aligned}$$

The first equality follows since $f_1(t_0)=0$ for $t_0<0$, and the second equality follows since $f(t-t_0)=0$ for $t_0>t$.

Equation 3.13 often is very difficult to evaluate. Fortunately, however, there exist several transformations of the density function that conveniently

[2] To derive the density function $f(t)$ of the random variable $\mathbf{T}=\mathbf{T}_1+\mathbf{T}_2$, first note that for a fixed t the event $\{\mathbf{T}\leq t\}$ is equivalent to the event $\{(\mathbf{T}_1,\mathbf{T}_2)\in A_t\}$, where $A_t=\{(t_1,t_2)\mid t_1+t_2\leq t\}$. Thus

$$\begin{aligned}F(t)&=P(\mathbf{T}\leq t)=P[(\mathbf{T}_1,\mathbf{T}_2)\in A_t]\\&=\iint_{A_t} f_{1,2}(t_1,t_2)\,dt_1\,dt_2\\&=\int_{-\infty}^\infty\int_{-\infty}^{t-t_1} f_{1,2}(t_1,t_2)\,dt_2\,dt_1\end{aligned}$$

Making the change of variables $t_2=t'-t_1$ reduces the expression to

$$F(t)=\int_{-\infty}^\infty\int_{-\infty}^{t} f_{1,2}(t_1,t'-t_1)\,dt'\,dt_1$$

If we now interchange the order of integration, then

$$F(t)=\int_{-\infty}^{t}\int_{-\infty}^\infty f_{1,2}(t_1,t'-t_1)\,dt_1\,dt'$$

The density function $f(t)$ is now easily found by differentiating with respect to t:

$$f(t)=\int_{-\infty}^\infty f_{1,2}(t_1,t-t_1)\,dt_1$$

If $\mathbf{T}_1$ and $\mathbf{T}_2$ are independent, then this integral simplifies to Eq. 3.13.

convert the operation of convolution into one of multiplication. The simplest of these is the so-called *moment-generating function* (mgf) of the random time **T**. Our development of mgf material will be for general densities defined on $-\infty<t<+\infty$ but can trivially be restricted to $0\leqslant t<+\infty$. The mgf is defined as[3]

$$M_T(\theta)=E(e^{-\theta\mathbf{T}})=\int_{-\infty}^{\infty} e^{-\theta t} f(t)\, dt \tag{3.14}$$

where θ is a real constant up until the integration is performed, at which point it becomes the variable in the "transform space." When t is interpreted as time and thus is always nonnegative, $M_T(\theta)=\int_0^\infty e^{-\theta t} f(t)\, dt$.

Proposition 3.4: If $f(t)=f_1(t)*f_2(t)$, where $f_1(t)$ and $f_2(t)$ are density functions, then $M_T(\theta)=M_1(\theta)M_2(\theta)$.

Proof: Note first that

$$M_T(\theta)=\int_{-\infty}^{\infty} e^{-\theta t}\left[\int_{-\infty}^{\infty} f_1(t_0)f_2(t-t_0)\, dt_0\right]dt$$

Interchanging the order of integration yields

$$M_T(\theta)=\int_{-\infty}^{\infty} f_1(t_0)\left[\int_{-\infty}^{\infty} e^{-\theta t}f_2(t-t_0)\, dt\right]dt_0$$

Now making the change of variables $t'=t-t_0$ reduces the expression to

$$\begin{aligned}M_T(\theta)&=\int_{-\infty}^{\infty} f_1(t_0)\left[\int_{-\infty}^{\infty} \exp[-\theta(t'+t_0)]f_2(t')\, dt'\right]dt_0\\ &=\int_{-\infty}^{\infty} f_1(t_0)\left[\exp(-\theta t_0)\int_{-\infty}^{\infty} \exp(-\theta t')f_2(t')\, dt'\right]dt_0\\ &=\left[\int_{-\infty}^{\infty} \exp(-\theta t_0)f_1(t_0)\, dt_0\right]\left[\int_{-\infty}^{\infty} \exp(-\theta t')f_2(t')\, dt'\right]\\ &=M_1(\theta)M_2(\theta) \quad \square\end{aligned}$$

Thus, mgfs convert convolution in the time domain into multiplication in the θ domain.

Moment-generating functions also have other useful properties. For instance, one important property of the mgf is the one to which it owes its name. The mgf allows one to easily derive all of the raw moments (the nth raw moment is $E(\mathbf{T}^n)$) of the associated density.

[3] Some authors define the mgf as $M_T(\theta)=E(e^{\theta\mathbf{T}})$. The two functions have identical properties. However, one trivial difference is that if $M_T(\theta)=E(e^{\theta\mathbf{T}})$, then the -1 in the statement of Proposition 3.5 is replaced by $+1$. We chose the Eq. 3.14 definition because of Cox and Miller (1965) and their important contributions to the theory of random walks, which we depend on heavily in Chapter 10.

Proposition 3.5:

$$E(\mathbf{T}^n)=(-1)^n \left.\frac{d^n M_T(\theta)}{d\theta^n}\right|_{\theta=0}$$

Proof:

$$\begin{aligned}\left.\frac{d^n}{d\theta^n} M_T(\theta)\right|_{\theta=0}&=\left.\frac{d^n}{d\theta^n}\int_{-\infty}^{\infty} e^{-\theta t}f(t)\, dt\right|_{\theta=0}\\ &=\left.\int_{-\infty}^{\infty} \frac{d^n e^{-\theta t}}{d\theta^n} f(t)\, dt\right|_{\theta=0}\\ &=\left.\int_{-\infty}^{\infty} (-t)^n e^{-\theta t} f(t)\, dt\right|_{\theta=0}\\ &=\int_{-\infty}^{\infty} (-t)^n f(t)\, dt\\ &=(-1)^n\int_{-\infty}^{\infty} t^n f(t)\, dt=(-1)^n E(\mathbf{T}^n) \quad \square\end{aligned}$$

An obvious corollary of Proposition 3.5 is that the mean of a random variable **T** can be found from

$$E(\mathbf{T})=-\left.\frac{d}{d\theta} M_T(\theta)\right|_{\theta=0} \tag{3.15}$$

Differentiation is usually easier to perform than integration, and thus if the mgf is known, it will usually be easier to calculate $E(\mathbf{T})$ from Eq. 3.15 than from Eq. 3.11.

There are many important properties of the mgf we will not take time to develop here. Many of these will be derived in Chapter 10, where the mgf will play a key role in our theoretical developments. The interested reader is referred to McGill (1963) for a more detailed discussion of psychological applications and to Parzen (1960) or Cox and Miller (1965) for a more complete mathematical development.

Unfortunately, as it happens, not all density functions possess mgfs, although by far and away most of the well-known densities do. An example of a density without an mgf is given by the Cauchy distribution, which is defined as

$$f(t)=\frac{a}{\pi(a^2+t^2)}, \quad -\infty<t<+\infty, \quad 0<a$$

The Cauchy density function looks much like a normal distribution centered at zero, except that its tails are much higher. It is a rather strange distribution since its mean is infinite:

$$E(\mathbf{T}) = \int_{-\infty}^{\infty} tf(t)\,dt = \int_{-\infty}^{\infty} t\frac{a}{\pi(a^2+t^2)}\,dt$$
$$= \frac{a}{2\pi}\left[\log(a^2+t^2)\Big|_{-\infty}^{\infty}\right] = \infty$$

There is, however, a simple modification which can be made to the definition of the mgf to ensure that the resulting transformation exists for *all* density functions. This new transformation, called the characteristic function of the random time $\mathbf{T}$, is defined as[4]

$$C_T(s) = E(e^{-is\mathbf{T}}) = \int_{-\infty}^{\infty} e^{-ist}f(t)\,dt$$

where $i=\sqrt{-1}$. In the engineering literature this function is known as the Fourier transformation. The reason that every random variable has a characteristic function but not an mgf is that $E(e^{-is\mathbf{T}})$ is always finite (i.e., bounded) for all real values of s but that $E(e^{-\theta\mathbf{T}})$ is not always bounded for the necessary range of $\theta \in (-\infty, +\infty)$. For example, although we noted that the Cauchy distribution has no mgf, its characteristic function turns out to be the very simple $\exp(-a|\theta|)$.

The characteristic function has many of the same properties as the mgf. For instance, it converts convolution in the time domain into multiplication in the s domain. It is a unique transformation, and from it the raw moments can be calculated in a manner almost identical to the way they are from the mgf. The characteristic function is a much more powerful transformation. Even so, the Fourier transform, of which it is a special case (in which the transformed function is a probability density), is not sufficiently general to handle functions such as $F(t)$, the cumulative distribution function. There will be few instances in the work that follows where this limitation will cause us any bother. Thus, a generalization of the characteristic function that has weaker existence conditions will seldom have to be employed. Nevertheless, there is one such generalization (which has been much studied) that will not only exist for virtually all functions we could ever be interested in, but in addition is easily manipulated and thus can quickly and painlessly derive us properties of the convolution operation that we shall occasionally find useful in our future work. The generalization is called the Laplace transformation. Given a function $A(t)$, its Laplace transform is defined as

$$L\{A(t)\} = \int_{-\infty}^{\infty} \exp[-(r+is)t]A(t)\,dt$$

where again $i=\sqrt{-1}$. Note that we can rewrite $L\{A(t)\}$ as follows:

[4] The characteristic function is frequently defined as $E(e^{is\mathbf{T}})$, in which case it is not identical to the Fourier transform. As in the case of the mgf, however, the differences are trivial.

$$L\{A(t)\} = \int_{-\infty}^{\infty} \exp(-ist)[\exp(-rt)A(t)]\,dt$$

which is the Fourier transform of $\exp(-rt)A(t)$, where r is some real constant.

It is now easy to see the increased generality of the Laplace transform. Notice that the Fourier transform is just the special case in which $r=0$. Also, if $A(t)=f(t)$ is a density function, then the mgf is just a special case in which $s=0$ (and $r=\theta$ to make our notation consistent). While functions can be found for which the Laplace transform does not exist, we shall never encounter any in our work.

For convenience it is usually agreed to set $q=r+is$ so that the definition of the Laplace transform can be written as

$$L\{A(t)\} = \int_{-\infty}^{\infty} \exp(-qt)A(t)\,dt$$

The Laplace transform is characterized by many very useful properties. We shall not try to develop most of these. The interested reader is referred to Churchill (1958) for a fuller development. Some of these properties, however, will be of vital interest to us. For instance, as one might by now expect, the Laplace transform converts convolution in the time domain into multiplication in the q domain.

A second useful property is that the Laplace transform converts integration in the time domain into division by q in the q domain. Thus,

$$L\left\{\int_{-\infty}^{t} A(t')\,dt'\right\} = \frac{1}{q}L\{A(t)\}$$

Now if $A(t)=f(t)$ is a probability density function, then

$$\int_{-\infty}^{t} f(t')\,dt' = F(t)$$

is its associated cumulative distribution function. Thus the Laplace transforms of densities and distribution functions are related by

$$L\{F(t)\} = \frac{1}{q}L\{f(t)\}$$

As an example of how this property might be applied, consider the case in which we are interested in the convolution of a distribution function and a density function. For instance, assume we are given two density functions $f_1(t)$ and $f_2(t)$ and we know that $f_1(t)*f_2(t)=f_3(t)$. Now suppose we are interested in evaluating $F_1(t)*f_2(t)$, where $F_1(t)$ is, of course, the distribution function associated with $f_1(t)$. The simplest way to evaluate this expression is to transform into the q domain. If we do this, we find

$$L\{F_1(t)*f_2(t)\} = L\{F_1(t)\}L\{f_2(t)\}$$
$$= \frac{1}{q}L\{f_1(t)\}L\{f_2(t)\}$$

$$= \frac{1}{q} L\{f_1(t) * f_2(t)\}$$

$$= \frac{1}{q} L\{f_3(t)\} = L\{F_3(t)\}$$

When we invert back to the time domain, we find that $F_1(t) * f_2(t) = F_3(t)$. The convolution of a distribution function and a density function is a distribution function.

The exponential distribution

Many of the results that we will later present will be developed and illustrated with exponentially distributed completion and intercompletion times. There are several reasons for this special interest in exponential processing. The mathematical properties of this distribution are simple so that most manipulations are easily analyzed and yield intuitive results. There is also an important historical interest; the exponential distribution has been an integral part of reaction time theorizing for quite some time (Christie & Luce 1956; Restle 1961; McGill 1963; McGill & Gibbon 1965; Luce & Green 1972; Townsend 1972, 1974b, 1976a). In addition, under certain circumstances, it is now possible to test the assumption that the duration of at least some RT components are exponentially distributed (Ashby & Townsend 1980; Ashby 1982a). This means, of course, that our theoretical investigations need never be divorced from empirical application. In light of these facts, this section will contain a description of the exponential distribution and some of its properties.

The exponential density with parameter w is given by

$$\begin{aligned} f(t) &= we^{-wt}, & t \geqslant 0 \\ &= 0, & t < 0 \end{aligned} \tag{3.16}$$

and the corresponding distribution function is

$$\begin{aligned} F(t) &= 1 - e^{-wt}, & t \geqslant 0 \\ &= 0, & t < 0 \end{aligned} \tag{3.17}$$

Examples of these functions are illustrated in Fig. 3.2.

Before calculating moments we will derive the mgf.

Proposition 3.6: If $\mathbf{T}$ is exponentially distributed with parameter w, then $M_T(\theta) = w/(w+\theta)$.

Proof:

$$M_T(\theta) = \int_0^\infty \exp(-\theta t) we^{-wt}\, dt = w \int_0^\infty \exp[-(w+\theta)t]\, dt$$

$$= w\left[-\frac{\exp[-(w+\theta)t]}{w+\theta} \bigg|_0^\infty \right] = \frac{w}{w+\theta} \quad \square$$

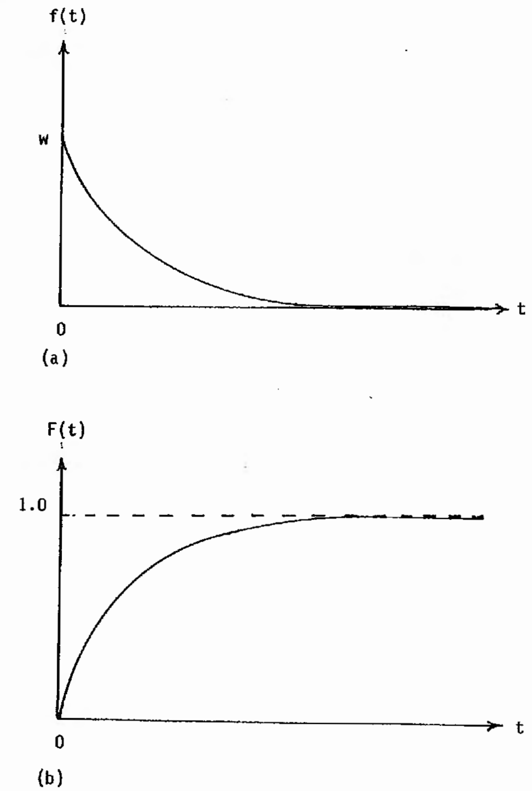


Fig. 3.2. The density (a) and distribution function (b) of an exponential distribution with parameter w.

We can now use this result to derive the raw moments of the distribution.

Proposition 3.7: If $\mathbf{T}$ is exponentially distributed with parameter w, then $E(\mathbf{T}^n) = n!/w^n$.

Proof: From Proposition 3.5,

$$E(\mathbf{T}^n) = (-1)^n \frac{d^n}{d\theta^n} M_T(\theta)\bigg|_{\theta=0} = (-1)^n \frac{d^n}{d\theta^n}\left(\frac{w}{w+\theta}\right)\bigg|_{\theta=0}$$

$$= w(-1)^n \frac{d^n (w+\theta)^{-1}}{d\theta^n}\bigg|_{\theta=0} = w(-1)^n \frac{n!(-1)^n}{(w+\theta)^{n+1}}\bigg|_{\theta=0}$$

$$= \frac{w(-1)^n n!(-1)^n}{w^{n+1}} = \frac{n!}{w^n} \quad \square$$

This result implies immediately that the exponential mean $E(\mathbf{T})$ is $1/w$; that is, the mean is the reciprocal of the exponential parameter. Thus we may view the parameter w as the *rate* of processing, in the sense that an increase in processing rate (i.e., in w) implies faster processing and therefore a decrease in the average processing time $E(\mathbf{T})$. This conceptualization of the exponential parameter will often improve our insight into the nature of processing, since it firmly ties an abstract mathematical parameter to a physical property of the system (rate of processing), and is just one more advantage of the exponential distribution as a model of processing times.

The variance of the exponential distribution can be easily derived from the raw moments:

$$\begin{aligned}\operatorname{var}(\mathbf{T}) &= E[\mathbf{T}-E(\mathbf{T})]^2 = E(\mathbf{T}^2)-[E(\mathbf{T})]^2\\ &= \frac{2}{w^2}-\frac{1}{w^2}=\frac{1}{w^2}\end{aligned}$$

The memoryless property and the gamma and Poisson processes

Earlier we remarked that when an arbitrary density or distribution function is written in terms of its hazard function, the result looks exponential when $H(t)$ is constant (i.e., compare Eqs. 3.6 and 3.7 with Eqs. 3.16 and 3.17). This can be easily verified by calculating the hazard function of the general exponential distribution:

$$H(t)=\frac{f(t)}{\bar{F}(t)}=\frac{f(t)}{1-F(t)}=\frac{we^{-wt}}{1-(1-e^{-wt})}=w \tag{3.18}$$

Note that it is time-invariant. In fact, $H(t)$ is constant if and only if $\mathbf{T}$ is exponentially distributed. The time-invariant hazard function of the exponential distribution illustrates its memoryless or ageless property. Equation 3.18 states that if processing has not been completed, then the probability it will be completed during the next instant of time is a constant and therefore does not depend on how long the system has been processing the element.

Another way of viewing this property is by way of the functional equation

$$P(\mathbf{T}<t_1+t_2 \mid \mathbf{T}>t_1)=P(\mathbf{T}<t_2) \tag{3.19}$$

Equation 3.19 states that if processing has not been completed by time t_1, then the probability that it is completed in the next t_2 time units is equal to the unconditional probability that it was completed during the *first* t_2 time units.

Although we will not do so here, it can be shown that this functional equation is true if and only if $\mathbf{T}$ is exponentially distributed (see Feller 1957: 413). Thus, the memoryless property is unique to this family of distributions. At first exposure such a property may seem very uncharacteristic of nature, but it has been shown to hold, or at least approximately hold, for a large number of physical phenomena. For instance, the exponential distribution has been used to model the time between incoming telephone calls (Jensen 1948), neuron pulses (Luce & Green 1972), servicing of machines (Palm 1947), disintegrations of radioactive material (Bateman 1910), and many other events (see Bharucha-Reid 1960; McGill 1963).

Once we know that the probability that a completion will occur in the next instant is a constant and independent of processing time, or in other words, once we know that the times between events are exponentially distributed, it is a simple matter to calculate the probability $P_t(k)$ that k completions will occur during any time interval of length t, for it turns out that this probability has the well known Poisson distribution (see Feller 1957)

$$P_t(k)=\frac{(wt)^k}{k!}e^{-wt}, \quad k=0,1,2,\ldots$$

In fact, such a stochastic process, where the times between events are independent and exponentially distributed, is known as a *Poisson process* with parameter w. The mean number of events occurring in any time interval of length t is wt and the variance is also wt.

A Poisson process might make a good model of a serial processing system in which the intercompletion times are all independent and identically distributed. Suppose the serial system has n stages and that $\mathbf{T}_i$ is the ith random intercompletion time. Then the total completion time $\mathbf{T}$ of the system equals the sum of the n intercompletion times, $\mathbf{T}=\sum_{i=1}^{n}\mathbf{T}_i$. In a Poisson process with parameter w, the $\mathbf{T}_i$ are all independent and exponentially distributed with rate w.

We can use many of the results of this chapter to study the behavior of the Poisson random completion time $\mathbf{T}$. For instance,

$$E(\mathbf{T})=\sum_{i=1}^{n}E(\mathbf{T}_i)=\sum_{i=1}^{n}\frac{1}{w}=\frac{n}{w}$$

Similarly, because of independence

$$\operatorname{var}(\mathbf{T})=\sum_{i=1}^{n}\operatorname{var}(\mathbf{T}_i)=\sum_{i=1}^{n}\frac{1}{w^2}=\frac{n}{w^2}$$

The mgf of the total completion time $\mathbf{T}$ can be found from Propositions 3.4 and 3.6 to be

$$M_T(\theta)=\prod_{i=1}^{n}M_i(\theta)=\prod_{i=1}^{n}\frac{w}{w+\theta}=\left(\frac{w}{w+\theta}\right)^n \tag{3.20}$$

It turns out that this is the mgf of the gamma distribution with n stages and rate equal to w. The density function associated with this gamma distribution is found by inversion to be

$$f(t)=\frac{(wt)^{n-1}}{(n-1)!}we^{-wt}, \quad t>0 \tag{3.21}$$

Note that when $n=1$, Eq. 3.21 is equivalent to Eq. 3.16 and thus the family of exponential distributions is contained within the family of gamma distributions.

The gamma and Poisson distributions thus form a sort of duality. Given a train of independent and exponentially distributed intercompletion times, there are two different random variables we could concentrate on: the time $\mathbf{T}$ until the kth completion and the number $\mathbf{K}$ of completions occurring by time t. The distribution of the first is gamma and the distribution of the second is Poisson. Given one the other can always be found, since the only way exactly k completions can occur before time t is if the time of the kth completion is less than t and the time of the $(k+1)$th completion is greater than t.

The Poisson process has been a very popular model of reaction time (Christie & Luce 1956; Restle 1961; McGill 1963), but the situation it describes is not the most general. It could be the case that the $\mathbf{T}_i$, although all exponentially distributed and independent, do not all have the same rate. In this more general model, it would be assumed that

$$f_i(t) = w_i e^{-w_i t}, \quad t \geqslant 0$$

Now we ask the obvious question: What is the distribution of $\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_n$? The mgf of $\mathbf{T}$ is still the product of all the component mgfs, and thus

$$M_T(\theta) = \prod_{i=1}^{n} \left(\frac{w_i}{w_i + \theta} \right) \tag{3.22}$$

The distribution with this mgf is known as a general gamma or Erlang distribution (see McGill & Gibbon 1965), and its density function is just a weighted sum of the component exponential densities. It is easy to see that with $w_i = w_j = w$ for all i and j, Eq. 3.22 reduces to Eq. 3.20 and therefore the family of simple gamma distributions is contained within the family of general gammas. Thus, given any number of exponentially distributed intercompletion times, we always know the distribution of their sum (i.e., the total completion time).

We shall run into the general gamma distribution as a model of human RT many times in the course of this book. Frequently, however, only part of the RT process is modeled by this distribution. There is often a random residual base time term $\mathbf{T}_B$ added on to account for such things as the time it takes to physically execute the response. Now $\mathbf{T}_B$ is usually not assumed to have an exponential distribution, and therefore the random reaction time when there are n independent and exponentially distributed intercompletion times, $\mathbf{RT}_n = \mathbf{T}_B + \mathbf{T}_1 + \cdots + \mathbf{T}_n$ will not have a general gamma distribution. In fact, without knowing something about the distribution of $\mathbf{T}_B$, very little can be said about the RT distribution itself. Nevertheless, Ashby and Townsend (1980) and Ashby (1982a) developed a means of simultaneously testing the assumptions of this model and estimating the exponential rate w_n, in the case when both $\mathbf{RT}_n$ and $\mathbf{RT}_{n-1}$ are available.

Proposition 3.8: Suppose $\mathbf{RT}_n = \mathbf{T}_B + \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_n$ and $\mathbf{RT}_{n-1} = \mathbf{T}_B + \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_{n-1}$, where all components are mutually independent and $\mathbf{T}_n$ has an exponential distribution with parameter w_n. Let $f_n(t)$ be the density function of $\mathbf{RT}_n$ and $F_n(t)$ the associated distribution function. Then

$$w_n = \frac{f_n(t)}{F_{n-1}(t) - F_n(t)} \quad \text{for all } t > 0$$

In addition, $f_n(t)$ and $f_{n-1}(t)$ must intersect at the mode of $f_n(t)$ and only there.

Proof: The conditions of the proposition imply

$$f_n(t) = f_{n-1}(t) * w_n \exp(-w_n t)$$

Taking Laplace transforms of both sides results in

$$L\{f_n(t)\} = L\{f_{n-1}(t)\} \frac{w_n}{w_n + q}$$

Algebraic manipulation leads to

$$L\{f_n(t)\} = \frac{w_n}{q} [L\{f_{n-1}(t)\} - L\{f_n(t)\}]$$

$$= w_n [L\{F_{n-1}(t)\} - L\{F_n(t)\}]$$

Inverting back to the time domain produces the first result.

To prove the second, note that the first result implies

$$f_n(t) = w_n [F_{n-1}(t) - F_n(t)] \quad \text{for all } t > 0$$

Differentiating both sides with respect to t yields

$$\frac{d}{dt} f_n(t) = w_n [f_{n-1}(t) - f_n(t)]$$

The result follows because of the equality. □

Thus this proposition provides two methods of testing this class of models. One can simply plot the two RT density functions and check their point of intersection. If it is not at the mode of $f_n(t)$, the models can be ruled out. If it is at the mode, an additional test can be performed by plotting $f_n(t)/[F_{n-1}(t) - F_n(t)]$ against RT. If the plot is flat, the RT model of the proposition is supported and the rate w_n can be estimated. The reader interested in more detailed discussions of these results, as well as some applications, is referred to Ashby (1982a) and Ashby and Townsend (1980).

Simultaneous Poisson processes

In addition to standard or generalized Poisson processes there will often be instances, particularly when we are investigating parallel models, when we shall be interested in the behavior of several independent Poisson processes operating simultaneously. Using the tools we have already developed, it is fairly easy to predict the behavior of such a system.

To simplify things, suppose there are two independent Poisson processes with k_1 and k_2 stages operating in parallel with respective rates w_1 and w_2 and suppose we are interested in the time $\mathbf{T}$ of the first-stage completion, not caring on which process it occurs. Call the time between completions $\mathbf{T}_1$ on the first process and $\mathbf{T}_2$ on the second. Then the density function of $\mathbf{T}=\min(\mathbf{T}_1, \mathbf{T}_2)$ is given by

$$f(t)=f_1(t)\bar{F}_2(t)+f_2(t)\bar{F}_1(t)$$

where $f_i(t)$ is the density function of $\mathbf{T}_i$. The first term on the right is the probability (density) that the first-stage completion occurs on the first process (i.e., the probability density that a completion occurs on process 1 at time t times the probability that a completion has not occurred on process 2 by this time), and the second term is the probability that this completion occurs on the second process.

Both $\mathbf{T}_1$ and $\mathbf{T}_2$ are exponentially distributed, and so

$$\begin{aligned} f(t) &= w_1\exp(-w_1 t)\exp(-w_2 t)+w_2\exp(-w_2 t)\exp(-w_1 t) \\ &= w_1\exp[-(w_1+w_2)t]+w_2\exp[-(w_1+w_2)t] \\ &= (w_1+w_2)\exp[-(w_1+w_2)t] \end{aligned}$$

Thus the time until the first-stage (not element or process) completion is itself exponentially distributed with rate w_1+w_2. What about the time between the first- and second-stage completions? If k_1 and k_2 (the number of stages) are greater than 1, then at the instant of the first completion one process is reset and so is in exactly the same state it was when the whole thing started. By the memoryless property of the exponential distribution, the other process, although not reset, is also in exactly the same state it was when processing began, so the time between the first and second completions must have exactly the same distribution as the time until the first completion. We have therefore expressed (for the special $n=2$ case) the following result.

Proposition 3.9: Until one of the processes is completed, the stage completion times of a system composed of n independent Poisson processes with rates $w_1, w_2, \ldots, w_n$ and all operating simultaneously themselves form a Poisson process with rate $w_1+w_2+\cdots+w_n$. After a process, say the ith, has completed its k_i stages, the new rate will be $\sum_{\substack{j=1\\ \neq i}}^{n} w_j$. □

Typical parallel models of the sort we are interested in do not keep operating indefinitely. In one important class of models we shall encounter, each individual process terminates after it is responsible for exactly one completion ($k_i=k_j=1$ for $1\leqslant i,\ j\leqslant n$). Proposition 3.9 allows us to predict the behavior of such systems. For instance, suppose the n processes of such a system all have rate w and we are interested in computing the mean total completion time.

We know the time until the first completion is exponentially distributed with rate nw (the sum of the n rates) and mean $1/nw$. At this point one of the processes terminates and so by the memoryless property the system is equivalent to one with $n-1$ Poisson processes, all with rate w, operating in parallel. The time between the first and second completions in this new system is therefore exponentially distributed with rate $(n-1)w$ and mean $1/(n-1)w$. Continuing with this logic leads to an expression for the mean total completion time for all n processes $\mathbf{T}$,

$$\begin{aligned} E(\mathbf{T}) &= \frac{1}{nw}+\frac{1}{(n-1)w}+\frac{1}{(n-2)w}+\cdots+\frac{1}{w} \\ &= \frac{1}{w}\sum_{i=1}^{n}\frac{1}{i} \end{aligned} \tag{3.23}$$

Relationship between discrete and continuous variables

It is of interest to take a moment to develop the discrete-time distributions that are analogs to the exponential, the gamma, and the Poisson distributions. This may help to sharpen the intuition for these latter distributions, particularly if the reader is more familiar with discrete probability theory. Also, although we shall not need them in the present work, situations could potentially arise where the discrete cases could be used, perhaps as approximations to the continuous case.

The geometric probability density function with parameter p is given by

$$P(k)=p(1-p)^k, \quad k=0,1,2,\ldots \tag{3.24}$$

The geometric distribution is used to model the number of independent Bernoulli trials until the first success occurs. In the probability literature, trials that can result in either a success or a failure are called *Bernoulli trials*. Let p be the probability of a success on any one trial, then the probability that the first success occurs on the $(k+1)$th trial is equal to the probability that the first k trials are all failures, $(1-p)^k$, times the probability that the $(k+1)$th trial is a success, p. This exactly yields Eq. 3.24.

A common example of a Bernoulli trial is a coin toss, where, say, a success is defined as the event that the coin comes up heads. If the coin is fair, the probability of a success, p, equals $\frac{1}{2}$. Thus, the geometric distribution with $p=\frac{1}{2}$ should be a good model of the number of coin tosses required before the first head appears.

Now suppose we define a trial as a small interval of time of length Δt. Then it can be shown that if we force Δt to be arbitrarily small by taking the limit as Δt approaches zero, and simultaneously keep $p/\Delta t$ constant, the geometric distribution begins to look more and more like the exponential distribution, and in the limit they are equivalent. (For the proof, see Bush & Mosteller 1955: 315–316; McGill 1963 gives a somewhat less rigorous argument). Thus the geometric density is the discrete analog of the exponential.

The geometric distribution, as one might suspect, has the same lack-of-memory property as the exponential. This can be easily seen by examining the definition of p. Since it is the probability of a success on any trial, it is necessarily time-invariant and thus the process is ageless. The probability that the next event is a success is the same on the first trial as it is after 100 successive failures. As is the case with the exponential distribution in continuous time, the geometric is the only discrete distribution characterized by this memoryless property.

Earlier we saw that when we add times that are each distributed exponentially, the sum has a gamma distribution and the family of exponentials is contained within the family of gammas. We might ask the same question of the discrete analog. What is the consequence of adding discrete times (i.e., the **K** values, the number of trials until the first success) that are all distributed geometrically?

The mgf of the geometric distribution is given by

$$M_K(\theta)=\frac{p}{1-(1-p)e^{-\theta}} \tag{3.25}$$

Thus when we add n identically distributed discrete times, the mgf is

$$M_K(\theta)=\left[\frac{p}{1-(1-p)e^{-\theta}}\right]^n \tag{3.26}$$

It turns out (by inversion of (3.26)) that this is the mgf of the negative binomial distribution, given by

$$P(k)=\binom{n+k-1}{k}p^n(1-p)^k, \quad k=0,1,2,\ldots \tag{3.27}$$

where $P(k)$ is the probability that the nth success occurs on trial $n+k$ (see Feller 1957). Note that, as we intuited, the geometric family is contained within the family of negative binomials as the special case when $n=1$. Further, in a proof that is almost a trivial generalization of the one relating geometric and exponential densities, it can be shown that the negative binomial is the discrete analog of the gamma distribution.

There is one other discrete-continuous analogy we would like to draw, and that is between the binomial and the Poisson distributions. Imagine a time interval of length t, where a random number of events (e.g., completions) are distributed uniformly. Now let us divide this interval into n equal length subintervals, with n large enough such that the probability that two or more events occur in the same interval is negligible. Since in every interval the probability that an event occurs is the same, say equal to p (i.e., therefore, $1-p$ is the probability of no event occurring in the interval), it follows that the probability of k events occurring in the n subintervals has the familiar binomial distribution with parameters n and p:

$$P(k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad 0\leqslant k\leqslant n \tag{3.28}$$

Assume now that we increase n. This will cause the length of the subintervals to shrink and at the same time the value of p to decrease. If the increase in n is proportional to the decrease in p, as one would expect, such that $np\simeq wt$, where w is some constant, and if we take the limit as n approaches infinity, then Eq. 3.28 becomes equivalent to the Poisson distribution

$$P_t(k)=\frac{(wt)^k}{k!}e^{-wt}, \quad k=0,1,2,\ldots$$

(See Feller 1957 for a proof of this result.) Now of course, $P_t(k)$ gives the probability that k events will occur somewhere in the interval $(0,t)$. Thus, the binomial distribution is the discrete analog of the Poisson.

Summary

In this chapter we introduced the basic mathematical tools that we will make use of in our subsequent developments. For a given nonnegative random time **T** we defined:

1. The distribution function of **T** as $F(t)=P(\mathbf{T}\leqslant t)$
2. The density function of **T** as $f(t)=dF(t)/dt$
3. The survivor function of **T** as $\bar{F}(t)=1-F(t)$
4. The hazard function of **T** as $H(t)=f(t)/\bar{F}(t)$
5. The moment-generating function (mgf) of **T** as $M_T(\theta)=\int_{-\infty}^{\infty}e^{-\theta t}f(t)\,dt$

In the case of n random times $\mathbf{T}_1,\mathbf{T}_2,\ldots,\mathbf{T}_n$ we defined:

1. The joint distribution function of the $\mathbf{T}_i$ as

$$F(t_1,t_2,\ldots,t_n)=P(\mathbf{T}_1\leqslant t_1,\mathbf{T}_2\leqslant t_2,\ldots,\mathbf{T}_n\leqslant t_n)$$

2. The joint density function as

$$f(t_1,t_2,\ldots,t_n)=\frac{\partial^n}{\partial t_1\,\partial t_2\cdots\partial t_n}F(t_1,t_2,\ldots,t_n)$$

3. The random times to be mutually independent if and only if

$$f(t_1,t_2,\ldots,t_n)=f(t_1)f(t_2)\cdots f(t_n)$$

We then derived the following properties of expectations:

1. $E(c)=c$
2. $E(c\mathbf{T})=cE(\mathbf{T})$
3. $E(\mathbf{T}_1+\mathbf{T}_2)=E(\mathbf{T}_1)+E(\mathbf{T}_2)$

Next we introduced the exponential distribution and developed some of its properties. We saw that if **T** is distributed exponentially, then for $t\geqslant 0$:

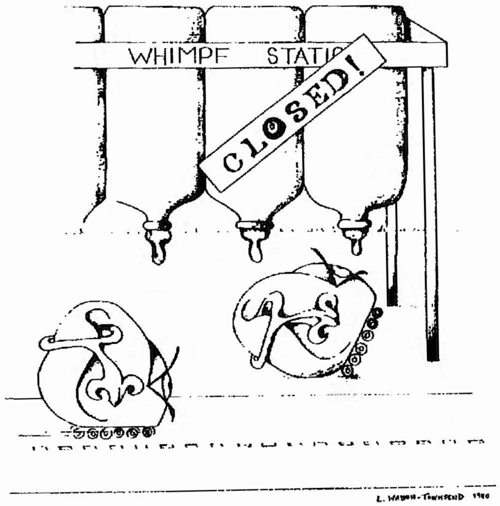
1. $f(t)=we^{-wt}$ and $F(t)=1-e^{-wt}$.
2. The moments are given by $E(\mathbf{T})=1/w$, $\text{var}(\mathbf{T})=1/w^2$, and $E(\mathbf{T}^n)=n!/w^n$.
3. The distribution is memoryless or ageless, i.e., $H(t)=w$ and

$$P(\mathbf{T}<t_1+t_2\,|\,\mathbf{T}>t_1)=P(\mathbf{T}<t_2)$$

4. The mgf is given by $M_T(\theta)=w/(w+\theta)$.
5. The sum of n such random variables is gamma-distributed.

Finally, we pointed out that the continuous densities that will come up time and again in our later developments have discrete analogs. Specifically,

Discrete		*Continuous*
geometric	$\longleftrightarrow$	exponential
negative binomial	$\longleftrightarrow$	gamma
binomial	$\longleftrightarrow$	Poisson



These two Wogols have an insatiable appetite for Whimpf and must be refilled at frequent intervals depending on their "mileage." The concept of *limited capacity* is well demonstrated by the empty Whimpf tanks following the possibly cataclysmic Whimpf famine on the neighboring plant Mut, which has always before seemed to possess an inexhaustible and perpetually available supply.

4 Stochastic models and cognitive processing issues

We are now ready to begin our development of stochastic models of processing. As noted, these are models that do not attempt to predict events exactly; instead, they predict probabilities that the event will occur during any specified time interval. Thus, in one sense this chapter will be the probabilistic analog of Chapter 2. However, as evidence of our bias that stochastic models are potentially much more powerful tools than are deterministic models, the present development will be of substantially greater depth (and breadth) than that of Chapter 2. In fact, for the most part, this development will continue throughout the remainder of the book.

In this chapter we plan to develop reasonably general yet precise definitions of stochastic parallel and serial processes. Then armed with these definitions we will investigate four critical issues or theoretical dimensions that serve to help determine the nature of the processing system. These issues include (1) parallel-serial equivalence – under what conditions is it possible (or impossible) for serial and parallel systems to mimic each other? (2) self-terminating vs. exhaustive processing – does processing terminate when all pertinent information has been extracted or does the system always process all of the "stimulus" pattern regardless of when the critical information is discovered? (3) independence vs. dependence – is the processing of the individual elements statistically independent or dependent? and (4) capacity –

how is the behavior of the system affected by changes in the processing load? It is to be emphasized that these issues are all logically independent of one another (Townsend 1974b), in the sense that any combination of "values" of the above four dimensions might be found in physically realizable systems. For instance, knowledge that a system is parallel can, ultimately, tell us nothing about whether processing is self-terminating or exhaustive, independent or dependent, or of limited or unlimited capacity. Of course, certain combinations may be more intuitively or psychologically acceptable than others.

In a sense, these issues could all be trivially resolved if the processing system were completely observable. For instance, to determine whether processing is serial or parallel we would only have to look in and count the number of operative channels. It seems extremely unlikely, however, that such a high degree of observability will be possible in the forseeable future. We shall therefore assume, throughout the book, that the most one could ever hope to observe are the exact times when each element is completed (i.e., the intercompletion times) and their order of completion.

Many of the results in this chapter will be derived fully only for the case when there are two elements to be processed. In most cases the added insights to be gained from deriving the results for $n>2$ are more than offset by the notational complexities induced by larger n. Within some reasonable bounds of simplicity, the models should be as general as possible. When we define parallel and serial processes, and when we derive equivalence mappings between the two, we necessarily wish these results to hold for as many specific models as possible. For instance, a system wherein processing rate differs for each element and for each element location within the display can be just as serial (or parallel) as a system that always processes all elements in the same time and in the same order, so we want to be sure to include these more general models in our discussions.

With this problem in mind we have tried to construct the definitions that follow so that they impose only those restrictions that we consider the essence of "seriality" and "parallelity" and yet still are specific enough to allow reasonable analytic investigations into their behavior. The serial definition that follows includes all of those models (and only those) that possess what we feel to be the essence of seriality. Unfortunately, this is not quite the case with our parallel definition. It assumes within-stage independence, a condition we feel is not necessary in a parallel process.

Within-stage independence states that *during any single stage* (i.e., the time between the completion of two successive elements) the processing of all unfinished elements is independent.[1] This does not rule out a possible across-stage dependency as found, say, in capacity reallocation models. These

[1] Vorberg (1977) showed that a large class (but not all) of parallel models may be given a within-stage independent representation.

models, which we closely examine in Chapter 6, postulate that the processing power or capacity that is freed when an element is completed is redistributed or reallocated to aid the processing of uncompleted elements. Such a reallocation causes a dependency to occur *between* stages, but as long as reallocation only occurs immediately after an element is completed, within-stage independence is still possible. Thus, although the assumption of within-stage independence rules out certain interesting parallel models, it still allows many different kinds of dependency to occur.

Why have we included this unnecessary restriction in our parallel definition? Within-stage independence guarantees that the intercompletion time density function can be written as the product of the separate densities of the uncompleted elements for the stage in question (see Chapter 3). This property enormously simplifies our analytic calculations and allows us to easily and intuitively derive many important results. At the time of this writing, the severity of this assumption is empirically undetermined. It may well turn out that within-stage independence adequately describes the relevant behavior of many natural parallel processing systems. There is certainly no question that such models yield a great diversity of processing behaviors. Chapters 14 and 15, which address general issues of parallel and serial systems and models, will consider some of the implications of relaxing this assumption.

Models describing the time course of the completion of a set of elements in a system can be written in terms of total completion times on the individual elements, on the intercompletion times, or on actual processing times. To facilitate comparison of parallel and serial processes, the primary definitions will be given in terms of intercompletion times. Chapter 14 and certain other developments take the different approach of defining serial models in terms of intercompletion times (designated by t) but parallel models in terms of total completion times (designated by τ) or *completion times* for brevity. Statements about time durations in general will usually employ t as the time symbol.

Before we begin, we need a more precise definition of *stochastic model*. Basically, we shall identify a model with a probability distribution on the time events in question, here the intercompletion times. Strictly speaking, however, a model is given by a *numerically* specific probability distribution (or density). A *class of models* will then be defined by a *set* of probability distributions on the events to be described. Ordinarily we write a model via a function of a set of parameters. When such a function is interpreted as if its parameters were specific values, it determines a model. If, on the other hand, it is interpreted as a function of variables, the set of potential values of the variables specify a class of models because they yield a set of probability distributions on the event set. In most cases, informal use of *model* and *class of models* should cause no confusion with regard to a particular probability distribution or set of distributions. Where potential confusion may occur, we shall attempt to be precise.

Parallel and serial definitions

Suppose a and b are the two positions of the elements to be processed and search is *serial*; then let $f_{ai}(t_{ai})$, for $i=1$ or 2, be defined as the probability density that the random time, $\mathbf{T}_{ai}$, for a to be processed when it is the ith position completed, actually turns out to be t_{ai}. In other words, $f_{ai}(t_{ai})$ is the density function describing the ith intercompletion time when the element in position a is completed ith and processing is serial. Analogous expressions hold for position b. Notice that here we are specifically indexing according to the *positions* a and b and not to the elements located in those positions. A processing system may favor one position over another regardless of the elements therein (or vice versa), even though it is the elements themselves that are actually processed. We will retain this focus on positions rather than element identity throughout most of this chapter. However, the question of element identity is very important, especially in certain experimental paradigms, and will come up on several occasions in the book (e.g., in Chapters 13 and 15).[2]

To emphasize the potential difference between the probability densities associated with serial and parallel models, we similarly define $g_{ai}(t_{ai})$ for $i=1$ or 2 as the density function describing the ith intercompletion time when a is completed ith and processing is *parallel*. Let $G_{ai}(t_{ai})$ be the parallel distribution function and $\bar{G}_{ai}(t_{ai})=1-G_{ai}(t_{ai})$ the survivor function for $i=1,2$ on position a.

We now state our definitions:

Definition 4.1: A model of a system for processing positions a and b is *serial* if and only if

$$f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)=pf_{a1}(t_{a1})f_{b2}(t_{b2}\mid t_{a1}) \tag{4.1}$$

and

$$f_{b1,a2}(t_{b1},t_{a2};\langle b,a\rangle)=(1-p)f_{b1}(t_{b1})f_{a2}(t_{a2}\mid t_{b1}) \quad \square \tag{4.2}$$

The quantity p is the probability that a is processed first. Also, $f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)$ is an expression of the probability density of the joint occurrence of the completion order $\langle a,b\rangle$, that a consumes t_{ai} time units of processing (i.e., $\mathbf{T}_{ai}=t_{ai}$) and that b completes processing t_{b2} time units after a.

Thus, $f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)$ is a joint density function of three events – two continuous and one discrete; so if we sum over all possible completion orders

[2] Paradigms that come immediately to mind here are those where the observer searches through some stimulus list for the presence of a critical or target element (e.g., Sternberg 1966; Atkinson, Holmgren, & Juola 1969). It is very natural to imagine in these cases that the system might process target elements and nontarget elements differently.

and then integrate over all possible processing times, the result should equal 1. In other words,

$$\int_0^\infty\int_0^\infty f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)\,dt_{a1}\,dt_{b2}+\int_0^\infty\int_0^\infty f_{b1,a2}(t_{b1},t_{a2};\langle b,a\rangle)\,dt_{b1}\,dt_{a2}=1$$

Equations 4.1 and 4.2 are intuitive descriptors of the behavior of serial models within each stage. For example, Eq. 4.1 states that with probability p, position a is processed first, in which case the first-stage density is $f_{a1}(t_{a1})$. Now the second-stage density must allow for the length of time taken for a to affect the processing time of b. Thus the second-stage density is the conditional density function $f_{b2}(t_{b2}\mid t_{a1})$. Equation 4.2 can be analyzed in an analogous fashion.

In the preceding chapter we indicated our interest in the exponential density function as a model for the intercompletion time. We saw that it is an extremely popular assumption, probably the most popular distributional assumption made by RT theorists. In addition, it is not without empirical support (e.g., McGill 1963; Hohle 1965; McGill & Gibbon 1965; Green & Luce 1967, 1971; Snodgrass, Luce, & Galanter 1967; Luce & Green 1970; Ratcliff & Murdock 1976; Ratcliff 1978; Ashby & Townsend 1980; Ashby, 1982a). Our investigations into the four critical "issues" in this chapter will depend heavily upon exponential models. In anticipation of this we now restate Definition 4.1 for the special case when the intercompletion times are distributed exponentially.

Definition 4.1A: A model of a system for processing positions a and b is *serial, across-stage independent, and exponential* if and only if

$$f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)=[pu_{a1}\exp(-u_{a1}t_{a1})][u_{b2}\exp(-u_{b2}t_{b2})] \tag{4.1A}$$

$$f_{b1,a2}(t_{b1},t_{a2};\langle b,a\rangle)$$
$$=[(1-p)u_{b1}\exp(-u_{b1}t_{b1})][u_{a2}\exp(-u_{a2}t_{a2})] \quad \square \tag{4.2A}$$

The brackets in these expressions illustrate how they are composed of terms describing the behavior of the system in each of the two stages of processing. Note that the second-stage term is just an unconditional density function that does not depend on the first-stage completion time. This illustrates a property we saw in the preceding chapter, namely *across-stage independence,* which implies, for example, that $f_{b2}(t_{b2}\mid t_{a1})=f_{b2}(t_{b2})$. This property will hold for both the serial and parallel exponential models that we will focus on.

We should be aware, however, that members of a set of distributions on the intercompletion times can be exponential and yet be dependent across stages. For example, if $u_{b2}=t_{a1}$, then

$$f_{b2}(t_{b2}\mid t_{a1})=t_{a1}\exp(-t_{a1}t_{b2})$$

That is, we simply let the rate of the second stage be equal to the duration of the first stage. This will produce a negative correlation across stages in the sense that long durations in stage 1 will tend to be followed by short durations in stage 2 and short durations in the first stage by long ones in stage 2.

We next define within-stage independent parallel models.

Definition 4.2: A model of a system for processing positions a and b is *within-stage independent and parallel* if and only if

$$g_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)=g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})g_{b2}(t_{b2}\mid t_{a1}) \tag{4.3}$$

and

$$g_{b1,a2}(t_{b1},t_{a2};\langle b,a\rangle)=g_{b1}(t_{b1})\bar{G}_{a1}(t_{b1})g_{a2}(t_{a2}\mid t_{b1})\ \square \tag{4.4}$$

Here, of course, $g_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)$ is completely analogous to $f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)$ of the serial model. As noted, $\bar{G}_{b1}(t_{a1})$ is the survivor function of $\mathbf{T}_{b1}$, that is, $\bar{G}_{b1}(t_{a1})=P\{\mathbf{T}_{b1}>t_{a1}\}$. This survivor function $\bar{G}_{b1}(t_{a1})$ along with $g_{a1}(t_{a1})$ describes the first stage of processing when the completion order is $\langle a,b\rangle$. The product gives the probability density that a is completed first at time t_{a1} and that b is not yet completed by this time. This component of parallel models represents the fact that a and b begin processing simultaneously. That it can be written as a product of functions of the individual elements is a result of our assumption of within-stage independence. The corresponding exponential definition follows easily.

Definition 4.2A: A model of a system for processing positions a and b is *parallel within-stage-independent, across-stage-independent, and exponential* if and only if

$$g_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)$$
$$=\{v_{a1}\exp[-(v_{a1}+v_{b1})t_{a1}]\}[v_{b2}\exp(-v_{b2}t_{b2})] \tag{4.3A}$$

$$g_{b1,a2}(t_{b1},t_{a2};\langle b,a\rangle)$$
$$=\{v_{b1}\exp[-(v_{b1}+v_{a1})t_{b1}]\}[v_{a2}\exp(-v_{a2}t_{a2})]\ \square \tag{4.4A}$$

Notice that we have added all exponents from the first-stage terms; that is, since $\bar{G}_{b1}(t_{a1})=\exp(-v_{b1}t_{a1})$ is the survivor function of the random time $\mathbf{T}_{b1}$ evaluated at time t_{a1}, then the first stage (when the order is $\langle a,b\rangle$) becomes

$$g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})=[v_{a1}\exp(-v_{a1}t_{a1})][\exp(-v_{b1}t_{a1})]$$
$$=v_{a1}\exp[-(v_{a1}+v_{b1})t_{a1}]$$

A casual inspection of these definitions yields several interesting observations about parallel and serial processes. First is the similarity between the second-stage functions of the two models. In fact, the components of Eqs. 4.1 and 4.3 (also for Eqs. 4.2 and 4.4) that describe the second stage of processing are structurally identical. This represents the fact that on the second stage

(i.e., during the second intercompletion time), whether processing is parallel or serial, only one element remains to be completed. It is obvious that with only one element left to be processed and a common processing history, serial and parallel models must be equivalent.

The second point to be noted is how the two processes differ in determining processing order. If the system is parallel, then the determination of processing order inherently depends on the rates with which a and b are processed. For instance, if a is processed with greater rate than b, then the completion order $\langle a,b\rangle$ will be more likely than the order $\langle b,a\rangle$. If the system operates serially, then the decision as to whether the processing order will be $\langle a,b\rangle$ or $\langle b,a\rangle$ is made, in fact must be made, a priori. The instant the system begins processing the first element, the completion order is known. Thus the probability that a is completed first can in no way depend on the processing rates of the elements in positions a or b.[3] This difference in how the systems select processing order is a fundamental difference between parallel and serial processing models.

The differences between the parallel and serial structure can be more easily seen if we compare the serial and parallel joint intercompletion time densities after they have been conditioned on processing order. First, in the serial case,

$$f_{a1,b2}(t_{a1},t_{b2}\mid\langle a,b\rangle)=\frac{f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)}{P^s(\langle a,b\rangle)}$$

We saw earlier that $P^s(\langle a,b\rangle)=p$, where the superscript s denotes serial processing, and therefore from Eq. 4.1,

$$f_{a1,b2}(t_{a1},t_{b2}\mid\langle a,b\rangle)=\frac{pf_{a1}(t_{a1})f_{b2}(t_{b2}\mid t_{a1})}{p}$$
$$=f_{a1}(t_{a1})f_{b2}(t_{b2}\mid t_{a1}) \tag{4.5}$$

In the parallel model

$$g_{a1,b2}(t_{a1},t_{b2}\mid\langle a,b\rangle)=\frac{g_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)}{P^p(\langle a,b\rangle)}$$

As we have already noted, when processing is parallel, $P^p(\langle a,b\rangle)$ depends on the rates of a and b. Specifically, for any completion time of a the order $\langle a,b\rangle$ will occur only if b is not yet complete. Thus, given any specific

[3] Of course, if there are more than two elements to be processed, a serial system need not decide the entire processing order before beginning, although it may. For instance, with three elements in positions a, b, and c, it might decide to process b first and then after b's completion choose between a and c. If this second choice (between a and c) is made independently of b's processing time, then this system is behaviorally equivalent to the system that selects processing order a priori. However, assume the system chooses a second when b is processed slowly, so that the secondary choice between a and c depends on the processing time of position b. It is easily seen that this system is more general than either of the first two. The serial models under study in the present work do not permit the latter type of dependence.

$\mathbf{T}_{a1}=t_{a1}$, the probability density on the order $\langle a,b\rangle$ is $g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})$. The probability $P^p(\langle a,b\rangle)$ can now be obtained by summing (i.e., integrating) over all possible processing times of a:

$$P^p(\langle a,b\rangle)=\int_0^\infty g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})\,dt_{a1} \tag{4.6}$$

From this expression we see that

$$g_{a1,b2}(t_{a1},t_{b2}\mid\langle a,b\rangle)=\frac{g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})g_{b2}(t_{b2}\mid t_{a1})}{\int_0^\infty g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})\,dt_{a1}} \tag{4.7}$$

Now the differences between parallel and serial processes can be more clearly seen. Equations 4.5 and 4.7 are structurally very different. In the serial models $P^s(\langle a,b\rangle)$ cancels out of the numerator and denominator of Eq. 4.5, reflecting the fact that the same structure does not determine both processing order and rate (although the two structures need not be independent). In parallel models the same structure performs both functions and thus Eq. 4.7 does not ordinarily simplify in the same way as the serial expression, Eq. 4.5.

Generalization to n elements

We now consider briefly how these definitions can be generalized to n elements. First observe that each definition must contain an expression analogous to Eqs. 4.1–4.4 for every possible processing order. With n elements there are $n!$ possible completion orders and thus the definitions will contain $n!$ equations. Let $\langle a_1,a_2,\ldots,a_n\rangle$ denote the completion order in terms of serial position with a_i denoting the serial position of the element finished ith. That is, $\langle a_1,a_2,\ldots,a_n\rangle$ is one of the $n!$ possible permutations of the n serial positions.

If processing is serial and the completion order is $\langle a_1,a_2,\ldots,a_n\rangle$, then the appropriate equation is

$$\begin{aligned}
&f_{a_1,a_2,\ldots,a_n}(t_{a_1},t_{a_2},\ldots,t_{a_n};\langle a_1,a_2,\ldots,a_n\rangle)\\
&=P(\langle a_1,a_2,\ldots,a_n\rangle)f_{a_1}(t_{a_1})f_{a_2}(t_{a_2}\mid t_{a_1})\cdots f_{a_n}(t_{a_n}\mid t_{a_1},t_{a_2},\ldots,t_{a_{n-1}})\\
&=[P(a_1)f_{a_1}(t_{a_1})][P(a_2\mid a_1)f_{a_2}(t_{a_2}\mid t_{a_1})]\cdots[P(a_n\mid a_1,\ldots,a_{n-1})\\
&\quad\times f_{a_n}(t_{a_n}\mid t_{a_1},t_{a_2},\ldots,t_{a_{n-1}})]
\end{aligned} \tag{4.8}$$

where, of course, $P(a_n\mid a_1,\ldots,a_{n-1})=1$.

Observe that this notation is slightly different from the notation we employed when $n=2$. For instance, when $n=2$, a_1 can equal $a1$ or $b1$ depending on whether a or b is processed first. Similarly, $a_2=a2$ or $b2$ depending on whether a or b is second.

Note that Eq. 4.8 is completely homologous to Eq. 4.1. The first-stage component $P^s(a_1)f_{a_1}(t_{a_1})$ contains the probability that a is selected first for processing and the probability density that $\mathbf{T}_{a_1}=t_{a_1}$. The second-stage component consists of the probability that b is selected second when a was first and the probability density that $\mathbf{T}_{b_2}=t_{b_2}$ given $\mathbf{T}_{a_1}=t_{a_1}$. Analogous interpretations can be given to all stages.

The n-element counterpart to the parallel Eq. 4.3 is

$$\begin{aligned}
&g_{a_1,a_2,\ldots,a_n}(t_{a_1},t_{a_2},\ldots,t_{a_n};\langle a_1,a_2,\ldots,a_n\rangle)\\
&=[g_{a_1}(t_{a_1})\bar{G}_{a_2,\ldots,a_n}(t_{a_1},t_{a_1},\ldots,t_{a_1})]\\
&\quad\times[g_{a_2}(t_{a_2}\mid t_{a_1})\bar{G}_{a_3,\ldots,a_n}(t_{a_2},t_{a_2},\ldots,t_{a_2})]\times\cdots\\
&\quad\times[g_{a_n}(t_{a_n}\mid t_{a_1},t_{a_2},\ldots,t_{a_{n-1}})]
\end{aligned} \tag{4.9}$$

The survivor function $\bar{G}_{a_2,\ldots,a_n}(t_{a_1},t_{a_1},\ldots,t_{a_1})$ gives the probability that none of the elements other than the one in position a_1 is completed by time t_{a_1}. Because of our assumptions of within-stage independence, these terms could have been written as products of the survivor functions of the individual positions. We wrote them instead in the form of Eq. 4.9 primarily to facilitate comparison with the earlier $n=2$ definition.

Note that in Eq. 4.9 the first-stage term gives the probability density that a is completed at time $\mathbf{T}_{a_1}=t_{a_1}$ and that none of the other elements is completed by that time. Similarly, the second-stage term gives the density that the a_2 element is completed second t_{a_2} time units after a_1 (i.e., at $t_{a_1}+t_{a_2}$) and that positions $a_3,\ldots,a_n$ have not completed processing by this time. Other stages have similar interpretations.

As can be seen, the serial and parallel definitions of Eqs. 4.1 to 4.4 for the $n=2$ case do generalize, in a very intuitive manner, to the general n case. However, equations such as 4.8 and 4.9 are cumbersome to manipulate and therefore, for the time being at least, we will continue to concentrate on the situation where the processing load on the system is two elements.

Parallel–serial equivalence

We are now in a position to investigate the conditions under which parallel and serial processes are equivalent. We shall begin our investigations by studying exponential models and then generalize our results. For convenience we restate the following equations. For serial exponential processing on two elements in positions a and b,

$$f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)=pu_{a1}\exp(-u_{a1}t_{a1})u_{b2}\exp(-u_{b2}t_{b2}) \tag{4.1A}$$

$$f_{b1,a2}(t_{b1},t_{a2};\langle b,a\rangle)=(1-p)u_{b1}\exp(-u_{b1}t_{b1})u_{a2}\exp(-u_{a2}t_{a2}) \tag{4.2A}$$

and for parallel exponential processing

$$g_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)=v_{a1}\exp[-(v_{a1}+v_{b1})t_{a1}]v_{b2}\exp(-v_{b2}t_{b2}) \tag{4.3A}$$

$$g_{b1,a2}(t_{b1},t_{a2};\langle b,a\rangle)=v_{b1}\exp[-(v_{a1}+v_{b1})t_{b1}]v_{a2}\exp(-v_{a2}t_{a2}) \tag{4.4A}$$

Our task is to derive parameter mappings [e.g., $p=f(v_{a1},v_{b1},v_{a2},v_{b2})$] that leave Eqs. 4.1A and 4.3A and Eqs. 4.2A and 4.4A totally equivalent, s guaranteeing that the parallel and serial *processes* are equivalent. To in, we can see immediately that with respect to the second stages this is a ial problem since whatever are the rates v_{a2} and v_{b2} (or u_{a2} and u_{b2}), we immediately set $v_{a2}=u_{a2}$ and $v_{b2}=u_{b2}$ so that during the second stage it l be impossible to discriminate between parallel and serial processing.

On the other hand, we can also observe that the *overall* rate of processing ring stage 1 is always $v_{a1}+v_{b1}$ in the parallel model but in the serial model n be either u_{a1} or u_{b1} depending on the order of processing. In fact, as one ght expect, when $u_{a1} \neq u_{b1}$, the parallel model *cannot* exactly mimic the ial model. A possible (although not necessarily observable) statistic pable of detecting this nonequivalence is the expected first-stage processing ne, conditioned on completion order. For instance, suppose we have a way recording the time $\mathbf{T}_1$ at which the first element is completed and that we ow what that element is. With serial processing, the expected value of $\mathbf{T}_1$ ien the completion order is $\langle a,b\rangle$ is

$$E^s(\mathbf{T}_1 \mid \langle a,b\rangle)=\frac{E^s(\mathbf{T}_1;\langle a,b\rangle)}{P^s(\langle a,b\rangle)}$$

$$=\frac{\int_0^\infty tpu_{a1}\exp(-u_{a1}t)\,dt}{p}=\frac{1}{u_{a1}}$$

here again, the superscript (s) refers to serial. When the completion order is $b,a\rangle$,

$$E^s(\mathbf{T}_1 \mid \langle b,a\rangle)=\frac{\int_0^\infty t(1-p)u_{b1}\exp(-u_{b1}t)\,dt}{1-p}=\frac{1}{u_{b1}}$$

nd thus, in general, $E^s(\mathbf{T}_1 \mid \langle a,b\rangle) \neq E^s(\mathbf{T}_1 \mid \langle b,a\rangle)$.

On the other hand, with parallel processing,

$$E^p(\mathbf{T}_1 \mid \langle a,b\rangle)=\frac{\int_0^\infty tv_{a1}\exp[-(v_{a1}+v_{b1})t]\,dt}{\int_0^\infty g_{a1}(t)\bar{G}_{b1}(t)\,dt}$$

$$=\frac{\int_0^\infty tv_{a1}\exp[-(v_{a1}+v_{b1})t]\,dt}{\int_0^\infty v_{a1}\exp[-(v_{a1}+v_{b1})t]\,dt}$$

$$=\frac{v_{a1}/(v_{a1}+v_{b1})^2}{v_{a1}/(v_{a1}+v_{b1})}$$

$$=\frac{1}{v_{a1}+v_{b1}}=E^p(\mathbf{T}_1 \mid \langle b,a\rangle)$$

Thus, in the parallel model, the two conditional means are the same, regardless of completion order.

Obviously, if $u_{a1} \neq u_{b1}$, then there is no way the parallel prediction can equal both of the statistics predicted by the serial model. Of course, the experimental value of this statistic (or any other) depends on its being observable, and unfortunately, this particular one is most often not. However,

suppose the observable response follows the first element to be completed, resulting in the minimum completion time plus whatever other residual times are included in the overall RT. This situation occurs when, for example, either element determines a unique response so that the first one completed leads to that response. If the mean of the added residual time is invariant over the two responses and does not depend on the identity of the element undergoing processing, then the observed mean RTs for the two responses provide a test of the proposition that

$$E^p(\mathbf{T}_1 \mid \langle a,b\rangle)=E^p(\mathbf{T}_1 \mid \langle b,a\rangle)$$

In fact, it can be shown that these exponential models even predict equivalent distributions, so that

$$g(t_1 \mid \langle a,b\rangle)=g(t_1 \mid \langle b,a\rangle)$$

However, this is such a strong prediction – one not implied by many (nonexponential) parallel models – that it does not appear to have ever been tested.

Proceeding with the question of sheer mathematical equivalence (and thus a fortiori experimental equality) it is obvious that the present parallel model can mimic this serial model only when $u_{a1}=u_{b1}=v_{a1}+v_{b1}$, so that the averaged (conditional) minimum processing time above is the same for parallel and serial models. Even assuming that $u_{a2}=v_{a2}$, $u_{b2}=v_{b2}$, and $u_{a1}=u_{b1}=v_{a1}+v_{b1}$, however, does not guarantee that the parallel and serial models will be fully equivalent. For we have as yet failed to translate into parallel terms the serial probability that position a is processed first, namely p. Now recall that $P^s(\langle a,b\rangle)=p$ and that

$$P^p(\langle a,b\rangle)=\int_0^\infty g_{a1}(t)\bar{G}_{b1}(t)\,dt=\int_0^\infty v_{a1}\exp[-(v_{a1}+v_{b1})t]\,dt$$

Thus to make the models fully equivalent and thus definitely indistinguishable, we must set

$$p=\int_0^\infty v_{a1}\exp[-(v_{a1}+v_{b1})t]\,dt=\frac{v_{a1}}{v_{a1}+v_{b1}}$$

We have now proved the following two results.

Proposition 4.1: Given any parallel exponential model in the form of Eqs. 4.3A and 4.4A, we can *always* construct a serial exponential model in the form of Eqs. 4.1A and 4.2A that is completely equivalent to it by setting $u_{a1}=u_{b1}=v_{a1}+v_{b1}$, $p=v_{a1}/(v_{a1}+v_{b1})$, $u_{a2}=v_{a2}$, and $u_{b2}=v_{b2}$.

Proposition 4.2: Given any serial exponential model (Eqs. 4.1A and 4.2A) where $u_{a1} \neq u_{b1}$, there exists *no* parallel exponential model (Eqs. 4.3A and 4.4A) that is equivalent to it. If $u_{a1}=u_{b1}=u_1$, we can construct a parallel exponential model that is completely equivalent to the serial model by setting $v_{a1}=pu_1$, $v_{b1}=(1-p)u_1$, $v_{a2}=u_{a2}$, and $v_{b2}=u_{b2}$.

Proof: The solutions are obtained by solving the equations of Proposition 4.1 for the parallel parameters. □

These two results indicate that, for this class of models, serial processes are more general than parallel processes. From a naive point of view this result is not unexpected since the serial exponential model has five parameters to the four for the parallel model.

Before proceeding to other issues we briefly discuss parallel–serial equivalence when no distributional assumptions are made about intercompletion times. In this case it is apparent that the two models are equivalent if and only if Eqs. 4.1 and 4.3 are equivalent and Eqs. 4.2 and 4.4 are equivalent; that is, it must be true that

$$pf_{a1}(t_{a1})f_{b2}(t_{b2} \mid t_{a1}) \equiv g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})g_{b2}(t_{b2} \mid t_{a1}) \tag{4.10}$$

and that

$$(1-p)f_{b1}(t_{b1})f_{a2}(t_{a2} \mid t_{b1}) \equiv g_{b1}(t_{b1})\bar{G}_{a1}(t_{b1})g_{a2}(t_{a2} \mid t_{b1}) \tag{4.11}$$

To obtain the parallel–serial equivalence mappings we need to alternatively solve these equations for the serial and parallel functions. The solutions, provided by Townsend (1976b), are given in the next two results.

Proposition 4.3: Given any within-stage independent parallel model (i.e., Eqs. 4.3 and 4.4) we can *always* construct a serial model (i.e., Eqs. 4.1 and 4.2) that is completely equivalent to it by setting

$$p = \int_0^\infty g_{a1}(t)\bar{G}_{b1}(t)\,dt$$

$$f_{a1}(t_{a1}) = \frac{1}{p}\, g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})$$

$$f_{b1}(t_{b1}) = \frac{1}{1-p}\, g_{b1}(t_{b1})\bar{G}_{a1}(t_{b1})$$

$$f_{b2}(t_{b2} \mid t_{a1}) = g_{b2}(t_{b2} \mid t_{a1})$$

and

$$f_{a2}(t_{a2} \mid t_{b1}) = g_{a2}(t_{a2} \mid t_{b1})$$

Proof: First it is obvious that, as in the case of the exponential models, we can immediately set the second-stage densities equal:

$$f_{b2}(t_{b2} \mid t_{a1}) = g_{b2}(t_{b2} \mid t_{a1}), \quad f_{a2}(t_{a2} \mid t_{b1}) = g_{a2}(t_{a2} \mid t_{b1})$$

so that our equivalence conditions reduce to

$$pf_{a1}(t_{a1}) = g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1}) \tag{4.12}$$

and

$$(1-p)f_{b1}(t_{b1}) = g_{b1}(t_{b1})\bar{G}_{a1}(t_{b1}) \tag{4.13}$$

We can solve for p by integrating Eq. 4.12 from zero to infinity, yielding

$$p = \int_0^\infty g_{a1}(t)\bar{G}_{b1}(t)\,dt$$

The full set of solutions is obtained by dividing both sides of Eq. 4.12 by p and both sides of Eq. 4.13 by $1-p$. □

An examination of these solutions reveals the very natural result that equivalence requires that the serial parameter p be set equal to $P^p(\langle a, b \rangle)$, the parallel probability that a finishes before b (see Eq. 4.6). It also reveals that no matter what the density functions $g_{a1}(t_{a1})$ and $g_{b1}(t_{b1})$ look like, as long as they are well defined, $f_{a1}(t_{a1})$ and $f_{b1}(t_{b1})$ will also be well-defined densities. This is a very important point, for it implies that given *any* within-stage independent parallel model, we can *always* construct a serial model that is completely equivalent to it by using the Proposition 4.3 solutions. To see that $f_{a1}(t_{a1})$ is, say, always well defined, note what happens when we integrate it over all time:

$$\int_0^\infty f_{a1}(t_{a1})\,dt_{a1} = \frac{1}{p}\int_0^\infty g_{a1}(t_{a1})\bar{G}_{b1}(t_{a1})\,dt_{a1}$$
$$= \frac{1}{p}(p) = 1$$

Solving Eqs. 4.10 and 4.11 for the parallel g terms, though not quite so simple, is nevertheless straightforward.

Proposition 4.4: Given a serial model in the form of Eqs. 4.1 and 4.2, then *if* there exists a within-stage independent parallel model (in the form of Eqs. 4.3 and 4.4) that is completely equivalent to it, it can be found by setting

$$\bar{G}_{a1}(t) = \exp\left[-\int_0^t \frac{pf_{a1}(t')}{p\bar{F}_{a1}(t') + (1-p)\bar{F}_{b1}(t')}\,dt'\right]$$

$$\bar{G}_{b1}(t) = \exp\left[-\int_0^t \frac{(1-p)f_{b1}(t')}{p\bar{F}_{a1}(t') + (1-p)\bar{F}_{b1}(t')}\,dt'\right]$$

$$g_{b2}(t_{b2} \mid t_{a1}) = f_{b2}(t_{b2} \mid t_{a1})$$

and

$$g_{a2}(t_{a2} \mid t_{b1}) = f_{a2}(t_{a2} \mid t_{b1})$$

Proof: The second-stage solutions are obvious. Thus we turn to Eqs. 4.12

and 4.13. Now adding these two and integrating from t to infinity converts densities into survivor functions and yields

$$\begin{aligned} p\bar{F}_{a1}(t)+(1-p)\bar{F}_{b1}(t) &= \int_t^\infty [g_{a1}(t')\bar{G}_{b1}(t')+g_{b1}(t')\bar{G}_{a1}(t')]\,dt' \\ &= \int_t^\infty -\left[\frac{d}{dt'}\bar{G}_{a1}(t')\bar{G}_{b1}(t')\right]dt' \\ &= \bar{G}_{a1}(\infty)\bar{G}_{b1}(\infty)+\bar{G}_{a1}(t)\bar{G}_{b1}(t) \\ &= \bar{G}_{a1}(t)\bar{G}_{b1}(t) \end{aligned}$$

Next we divide Eqs. 4.12 and 4.13 by this expression, giving

$$\frac{pf_{a1}(t)}{p\bar{F}_{a1}(t)+(1-p)\bar{F}_{b1}(t)}=\frac{g_{a1}(t)}{\bar{G}_{a1}(t)}$$

and

$$\frac{(1-p)f_{b1}(t)}{p\bar{F}_{a1}(t)+(1-p)\bar{F}_{b1}(t)}=\frac{g_{b1}(t)}{\bar{G}_{b1}(t)}$$

Now integrating both sides of each expression from zero to t yields

$$\begin{aligned} \int_0^t\left[\frac{pf_{a1}(t')\,dt'}{p\bar{F}_{a1}(t')+(1-p)\bar{F}_{b1}(t')}\right] &= \int_0^t\frac{g_{a1}(t')}{\bar{G}_{a1}(t')}\,dt' \\ &= -\ln\bar{G}_{a1}(t) \end{aligned}$$

and

$$\begin{aligned} \int_0^t\left[\frac{(1-p)f_b(t')\,dt'}{p\bar{F}_{a1}(t')+(1-p)\bar{F}_{b1}(t')}\right] &= \int_0^t\frac{g_{b1}(t')}{\bar{G}_{b1}(t')}\,dt' \\ &= -\ln\bar{G}_{b1}(t) \end{aligned}$$

The final solutions are obtained by multiplying both expressions by -1 and then exponentiating. □

Although the serial solutions of Proposition 4.3 *always* exist, such is not the case here. It may not be immediately obvious, but there do exist serial densities $f_{a1}(t)$ and $f_{b1}(t)$, for which $\bar{G}_{a1}(t)$ and $\bar{G}_{b1}(t)$, as defined in Proposition 4.4, are not well-defined survivor functions. A necessary condition of any true survivor function is that it approach zero as t approaches infinity.[4]

[4] Townsend (1976b) also presents two sufficient conditions for $\bar{G}_{a1}(t)$ and $\bar{G}_{b1}(t)$ to be survivor functions. The first is that

$$\lim_{t\to+\infty}\frac{F_{b1}(t)}{F_{a1}(t)}=\alpha$$

where $0<\alpha<\infty$. The second condition is that as $t\to+\infty$, both

$$\frac{pf_{a1}(t)}{p\bar{F}_{a1}(t)+(1-p)\bar{F}_{b1}(t)} \quad\text{and}\quad \frac{(1-p)f_{b1}(t)}{p\bar{F}_{a1}(t)+(1-p)\bar{F}_{b1}(t)}$$

approach zero no faster than $a/(b+ct)$. Further, neither of these conditions implies or is implied by the other. Finally it is as yet unknown whether these conditions *taken together* are also *necessary* for $\bar{G}_{a1}(t)$ and $\bar{G}_{b1}(t)$ to be survivor functions.

In other words, if $\bar{G}_{a1}(t)$ is a well-defined survivor function, then $\bar{G}_{a1}(\infty)=P(\mathrm{T}_{a1}>\infty)=0$. This condition is met in the Proposition 4.4 equations if and only if the integrals in the $\bar{G}_{a1}$ and $\bar{G}_{b1}$ solutions diverge as t approaches infinity [then, e.g., $\bar{G}_{a1}(\infty)=e^{-\infty}=0$]. That this divergence is not guaranteed by the restrictions that $f_{a1}(t)$ and $f_{b1}(t)$ be well-defined densities implies the existence of serial models that *no* within-stage independent parallel model can mimic. Thus, as with the exponential models, these parallel models are not as general as serial models, in the sense that the class of parallel models defined by Eqs. 4.3 and 4.4 is contained *within* the class of serial models defined by Eqs. 4.1 and 4.2. This result was recently employed in an experimental test of serial vs. within-stage independent parallel memory retrieval by Ross and Anderson (1981).

It is easily seen that no within-stage independent parallel models exist that are equivalent to the exceedingly simple exponential serial model outlined above with

$$pf_{a1}(t_{a1})=pu_{a1}\exp(-u_{a1}t_{a1})$$

and

$$(1-p)f_{b1}(t_{b1})=(1-p)u_{b1}\exp(-u_{b1}t_{b1})$$

as long as $u_{a1}\neq u_{b1}$. The second-stage densities are, of course, irrelevant. The reader may wish to demonstrate that both $\bar{G}_{a1}(t)$ and $\bar{G}_{b1}(t)$ cannot exist as valid survivor functions in this circumstance.

The solutions to the above equations for $n>2$ are very similar to the $n=2$ case and will not be repeated here. (The interested reader is referred to Townsend 1976b). Chapters 14 and 15 will present further results on the problems of equivalence in general parallel and serial models.

We next briefly take up models that are neither parallel nor serial. This material may be skipped on a first reading.

Hybrid models

Definitions 4.1 and 4.2 do not, of course, exhaust the universe of all possible models of processing systems, even if we were to drop the assumption of within-stage independence in Definition 4.2. We shall refer to all processing models that are not strictly serial or parallel in operation as *hybrid models.*

Hybrid models have not yet generated as much theoretical interest as have serial and parallel models. This is probably due in part to the difficulty of testing them experimentally but also to the challenge of writing an analytic definition that includes more than a small subset of this large class of models. This difficulty is indicated by our definition above, which is a definition of *exclusion* ("A hybrid model is *not*...") rather than the more analytically desirable *inclusion* ("A hybrid model *is*..."). In all likelihood, theorists will be forced to independently analyze each new hybrid model that they encounter in their research. It is our belief however, that such analyses will be aided by a knowledge of the theory of parallel and serial processes. Indeed, in some cases there will likely exist a parallel or serial model (or some combination) that is equivalent to the hybrid model in question.

Among such hybrid models are those where, within trials, processing is parallel part of the time and serial part of the time. For instance, assume the system always processes the first two elements serially and then processes the remaining elements in parallel. Such a model must be considered hybrid; but note that we can bifurcate such a system into two separate subsystems, one that is a strict serial processor and one that is a strict parallel processor. Such a bifurcation need have no intuitive psychological rationale; rather, it can be looked upon merely as a computational tool. This technique is analogous to one popular in mathematical learning theory, that of creating artificial states solely for the purpose of allowing a Markov description of a non-Markovian process (e.g., Bower 1959; Cox & Miller 1965). Here the results are similar. The behavior of each subsystem can be studied using the theory of parallel and serial processes. Other hybrid models amenable to this strategy are those where on some proportion of trials processing is strictly serial while on the remainder of the trials it is strictly parallel.

Of more recent theoretical interest are time-sharing models, that is, hybrid models that work on one element for a while, then, before the first element is completed, switch to another element and work on it. Usually, in perceptual or cognitive applications of these models, it is assumed that the elements are composed of "features" and that "attention" is not switched away from an element until after one or more features complete processing. Such models can often be given both parallel and serial interpretations, although often the parallel description is more intuitive.

Rumelhart (1970) proposed an independent parallel model of letter identification that has a very natural time-sharing interpretation. He assumed that each letter is made up of more elementary features and that on each parallel channel the intercompletion times on the individual features are exponentially distributed (although in some cases the rate may vary over time). The series of intercompletion times that make up the completion of each letter thus form a Poisson process and the individual letter completion times are gamma distributed (see Chapter 3). (This is a very straightforward parallel model, although it is assumed that in certain experimental circumstances - such as the delayed partial report paradigm of Sperling (1960) - attention can be totally reallocated to a subset of the displayed letters. In the

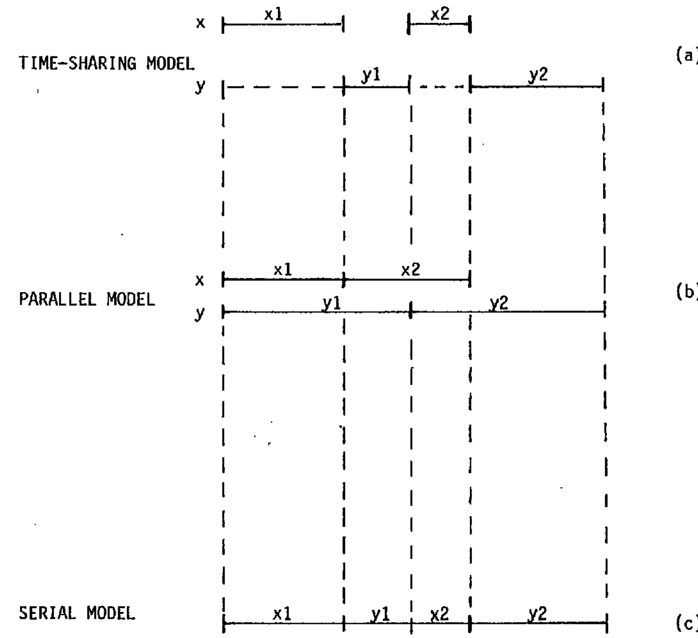


Fig. 4.1. Schematic representing the hypothetical processing of two elements (x) and (y), each consisting of two features (x_1, x_2) and (y_1, y_2). In the example illustrated, the features are completed in the order $\langle x_1, y_1, x_2, y_2 \rangle$. Interpretations are given in terms of three models: a time-sharing model, a parallel model, and a serial model.

latter circumstance the model is, of course, no longer independent on the total set of letters. We are here concerned only with the independent representation.)

The hybrid interpretation of this model (see Townsend 1972: 186–190 for details; see also Chapter 15) is based on the intuition that, at best, the observable events would be the intercompletion times on the various features (even these are usually obscured, as are those on individual letters). Therefore, the equivalent time-sharing model assumes attention may be switched from letter to letter before the one being worked on is completed.

This situation is depicted in Fig. 4.1a for the hypothetical case in which two letters, (x) and (y), each consist of two features, $(x1, x2)$ and $(y1, y2)$, and in which the completion order (at the feature level) is $\langle x1, y1, x2, y2 \rangle$. The parallel interpretation is shown in 4.1b. (Fig. 4.1c will be described below.) This parallel model is indistinguishable from the time-sharing model.

Another type of processing model, constructed by Shevell and Atkinson

(1974), is most readily depicted in a time-sharing framework.[5] This class of models differs from the time-sharing interpretation of Rumelhart's (1970) model in the restrictions placed on processing order. That is, the models of Shevell and Atkinson operate exactly as in Fig. 4.1a, but with the added restriction that exactly one feature must be processed in every element before a second feature can be processed in any element. In other words, the completion order $\langle x1, y1, x2, y2\rangle$ in Fig. 4.1a is a legitimate processing order for the model of Shevell and Atkinson (1974); however, the order $\langle x1, x2, y1, y2\rangle$ is not allowed. Similarly, with three elements the order $\langle x1, y1, x2, z1, y2, z2\rangle$ is impossible because $(x2)$ is completed before $(z1)$. Consequently, an equivalent "parallel" model would have to assume that for every letter the $(n+1)$st step of processing could not be completed until the nth step in all other letters is completed. The only parallel system in which this appears to be possible is *deterministic* in the sense that it knows, a priori, exact processing times of features. For instance, consider again our two-element example. Suppose $(x1)$ is the first feature completed and therefore $(y1)$ must be completed next. Now as $(x2)$ and $(y1)$ are being simultaneously processed the deterministic parallel system can monitor their "time until completion." If this time is ever less for $(x2)$ than for $(y1)$, then the system must divert processing capacity to $(y1)$ and away from $(x2)$. This will result in an increase in $(y1)$'s processing rate as opposed to $(x2)$'s, thereby decreasing the "time until completion" for feature $(y1)$ and increasing it for $(x2)$. In this way the system can guarantee that $(y1)$ is always completed *before* $(x2)$.

In a true stochastic system, however, the "time until completion" is never known exactly. At best we may know the hazard function (the probability the feature is completed in the next instant of time given it is not yet finished; see Chapter 3) for all features at every point in time. If $(x2)$'s hazard function ever appears too large compared with $(y1)$'s, the stochastic system can simultaneously increase the processing rate of $(y1)$ and decrease the rate of $(x2)$ [thereby raising $(y1)$'s hazard function and lowering $(x2)$'s]. But as long as the rate of $(x2)$ is greater than zero, the hazard function of $(x2)$ will also be greater than zero and thus there will occur instances, however rare, when $(x2)$ will be completed before $(y1)$. If the processing rate of $(x2)$ is set to zero, then the completion order $\langle y1, x2\rangle$ is guaranteed, but the system is no longer parallel since during that time there is no simultaneous processing.

[5] This continues the convention established earlier (Townsend 1974b). The term *hybrid* is not altogether felicitous, tending to connote an offspring possessing some of the characteristics of both parents. Some hybrid models may seem to issue from a marriage of parallel and serial models, but others may seem distinct from either. This connotation has unfortunately led some authors to interpret our use of *hybrid* as denoting models that are both serial and parallel (e.g., Harris, Shaw, & Bates 1979). Nevertheless it seems preferable to retain the word *hybrid* at this time rather than coin yet another term.

The present models, although clearly time-sharing by our definitions, are called *parallel* by Shevell and Atkinson (1974).

There is no model that is serial at both the feature and letter level that mimics either the Rumelhart or the Shevell and Atkinson type of model. However, if we define the meaningful stimulus elements as features rather than letters, then a serial interpretation, as depicted in Fig. 4.1c, is possible.

The problem of what to define as an element arises frequently in experimental and theoretical circumstances. For example, in visual recognition studies, it is typical that the element of "unit comparison" is assumed to be at least as large as a whole letter or character. We just saw how decisions such as this can affect the parallel–serial issue and therefore are matters that should be given much thought any time experimental applications of these models are made (Taylor 1976a).

The other alternative with regard to serial mimicry is to gloss over the feature-level representation and describe only the total completion times on the letters or, equivalently, the intercompletion times on the letters. With the detailed structure related to the completion times of the individual features obscured, parallel–serial equivalence follows from slight extensions of the formation of Propositions 4.1 and 4.2 or from the equivalence mappings of Chapter 14. Chapter 15 discusses parallel–serial differences when the fine grain of the element composition is taken into account in the stochastic models describing the underlying systems.

We now turn to consider the possibility that the system is capable of self-terminating its processing as soon as the critical stimulus subset is discovered. This next section will deal primarily with lower-order moments of the processing time distributions (i.e., mean and variance). These will be simpler and more intuitive to work with than the densities and have the advantage of relevance to statistics typically obtained in experimental literature. A more general discussion of this question is deferred until Chapter 7.

Self-terminating vs. exhaustive processing

Again we will concentrate our development on the $n=2$ case when the intercompletion times are distributed exponentially. Let us begin by deriving mean total processing times for the various models. In Chapter 2 we defined the total completion time of an element (or completion time for short) as the time between the beginning of the first element and the completion of the element of interest. In other words, we are no longer interested in the lengths of the intercompletion times per se, but rather in the (mean) length of their sum.

It is well known that the mean of a sum of random times is the sum of the individual means (see Proposition 3.3). Thus, in the case of both parallel and serial processing, we need only calculate the means of the component intercompletion times and add these up. This operation is especially easy to perform for the serial models because the intercompletion times are also the actual processing times of the individual elements.

Sometimes we will be concerned with the total completion time of a single element, usually in a specific serial position. Total completion time without reference to an element implies that we are interested in the total duration that the system is in operation. In many cases, we will assume that a single "target" element carries sufficient information for termination to occur.

Recall that when processing is serial, exhaustive, and exponential and the completion order is $\langle a, b\rangle$, then $f_{a1}(t_{a1}) = u_{a1}\exp(-u_{a1}t_{a1})$, and the mean processing time of the element in position a is therefore $E^s(\mathbf{T}_{a1}) = 1/u_{a1}$. Also, $f_{b2}(t_{b2}) = u_{b2}\exp(-u_{b2}t_{b2})$ and thus $E^s(\mathbf{T}_{b2}) = 1/u_{b2}$. Since the expected total completion time is the sum of these two,

$$E^s_{\mathrm{EX}}(\mathbf{T} \mid \langle a, b\rangle) = \frac{1}{u_{a1}} + \frac{1}{u_{b2}}$$

where EX denotes exhaustive processing. Similarly, when the completion order is $\langle b, a\rangle$,

$$E^s_{\mathrm{EX}}(\mathbf{T} \mid \langle b, a\rangle) = \frac{1}{u_{b1}} + \frac{1}{u_{a2}}$$

Finally, since completion order $\langle a, b\rangle$ occurs with probability p, we have that

$$E^s_{\mathrm{EX}}(\mathbf{T}) = pE^s_{\mathrm{EX}}(\mathbf{T} \mid \langle a, b\rangle) + (1-p)E^s_{\mathrm{EX}}(\mathbf{T} \mid \langle b, a\rangle)$$
$$= p\left(\frac{1}{u_{a1}} + \frac{1}{u_{b2}}\right) + (1-p)\left(\frac{1}{u_{b1}} + \frac{1}{u_{a2}}\right) \tag{4.14}$$

Note that this is the mean of the sum of Eqs. 4.1A and 4.2A.

When processing is self-terminating (ST), mean total completion time will depend on whether position a or b carries the critical information. When the only pertinent element is in position a, then

$$E^s_{\mathrm{ST}}(\mathbf{T} \mid \text{target is } a) = p\frac{1}{u_{a1}} + (1-p)\left(\frac{1}{u_{b1}} + \frac{1}{u_{a2}}\right) \tag{4.15}$$

This expression conveys the fact that with probability p, position a is processed first, and because the system does not need to process b, the mean total completion time on those trials is just the expected first-stage processing time (i.e., $1/u_{a1}$). Similarly, with probability $(1-p)$, position b is completed first, and hence both elements must be processed before the critical position a is completed.

When position b carries all the pertinent information, then

$$E^s_{\mathrm{ST}}(\mathbf{T} \mid \text{target is } b) = p\left(\frac{1}{u_{a1}} + \frac{1}{u_{b2}}\right) + (1-p)\frac{1}{u_{b1}} \tag{4.16}$$

Finally, if both a and b each contain all necessary information, then

$$E^s_{\mathrm{ST}}(\mathbf{T} \mid \text{target is } a \text{ or } b) = p\frac{1}{u_{a1}} + (1-p)\frac{1}{u_{b1}} \tag{4.17}$$

The parallel exponential derivations are more interesting, in that they give us an opportunity to use our parallel–serial equivalence mappings. For instance, in the exhaustive case, we could derive the mean processing time directly from the densities as we did with the serial models. However, in this case the first intercompletion times cannot be so simply written as with the serial models. A direct calculation of the expected first intercompletion time for parallel models involves multiple integrations, which often are messy. Instead, because we now know that for every parallel model of this class there exists an equivalent serial model, we can employ the serial expressions of Eqs. 4.14–4.17 and insert into them the parameter equivalences from Proposition 4.1 to convert them into the appropriate parallel expressions. Thus, beginning with the serial exhaustive mean from Eq. 4.14, if we set $p = v_{a1}/(v_{a1}+v_{b1})$, $u_{a1} = u_{b1} = v_{a1}+v_{b1}$, $u_{b2} = v_{b2}$, and $u_{a2} = v_{a2}$, the serial expression will be transformed into

$$E^p_{\mathrm{EX}}(\mathbf{T}) = \frac{v_{a1}}{v_{a1}+v_{b1}}\left(\frac{1}{v_{a1}+v_{b1}} + \frac{1}{v_{b2}}\right) + \frac{v_{b1}}{v_{a1}+v_{b1}}\left(\frac{1}{v_{a1}+v_{b1}} + \frac{1}{v_{a2}}\right)$$

which is the correct exhaustive mean for the parallel model. This expression can be simplified to

$$E^p_{\mathrm{EX}}(\mathbf{T}) = \frac{1}{v_{a1}+v_{b1}}\left(1 + \frac{v_{a1}}{v_{b2}} + \frac{v_{b1}}{v_{a2}}\right) \tag{4.18}$$

Note that Eq. 4.18 corresponds to the special serial case where $u_{a1} = u_{b1}$.

Employing this same technique with the self-terminating cases, we find (after some simplification) that when a is critical, the parallel mean is

$$E^p_{\mathrm{ST}}(\mathbf{T} \mid \text{target is } a) = \frac{1}{v_{a1}+v_{b1}}\left(1 + \frac{v_{b1}}{v_{a2}}\right) \tag{4.19}$$

and when b is the pertinent position,

$$E^p_{\mathrm{ST}}(\mathbf{T} \mid \text{target is } b) = \frac{1}{v_{a1}+v_{b1}}\left(1 + \frac{v_{a1}}{v_{b2}}\right) \tag{4.20}$$

Finally, when a and b each contain all necessary information,

$$E^p_{\mathrm{ST}}(\mathbf{T} \mid \text{target is } a \text{ or } b) = \frac{1}{v_{a1}+v_{b1}} \tag{4.21}$$

An example

As a very simple example of how self-terminating and exhaustive processes can mimic each other, assume that we can observe the mean total completion time of some system (e.g., a human observer). The system's hypothetical task is to process two elements (x) and (y), and it is told that all pertinent information is contained in element (x) regardless of its *position* in the display. Thus if processing is self-terminating and on some trial the system happens to process element (x) first, then it need not complete processing on the second element in the display. Suppose we present the system with three types of trials: (1) trials on which element (x) appears once and (y)

appears once (called (x,y) trials); (2) trials on which (x) appears in both positions $[(x,x)$ trials]; and (3) trials on which (y) appears in both positions $[(y,y)$ trials]. We then run our experiment and observe that on (x,x) trials the mean total processing time is 100 msec, on (x,y) trials it is 150 msec, and on (y,y) trials it is 200 msec.

Assuming for the moment that processing is serial, what can we conclude about the self-terminating vs. exhaustive issue? First, we know that processing must be exhaustive on the (y,y) trials, since the element (y) contains no pertinent information. But what about (x,x) and (x,y) trials? At first glance, one might argue for self-termination. For instance, assume $p=\frac{1}{2}$ and that whenever either (x) or (y) are processed (either first or second) and whatever position they are in, their mean processing time is always 100 msec. In other words, suppose $p=\frac{1}{2}$ and $1/u_x=1/u_y=100$. Now on (y,y) trials we see from Eq. 4.14 that

$$E^s(\mathbf{T} \mid (y,y))=\tfrac{1}{2}(100+100)+\tfrac{1}{2}(100+100)=200$$

as desired. Further, from Eq. 4.15 we see that this self-terminating model predicts

$$E^s(\mathbf{T} \mid (x,y))=\tfrac{1}{2}\cdot 100+\tfrac{1}{2}(100+100)=150$$

and on (x,x) trials (from Eq. 4.17),

$$E^s(\mathbf{T} \mid (x,x))=\tfrac{1}{2}\cdot 100+\tfrac{1}{2}\cdot 100=100$$

The self-terminating model fits the data perfectly.

Can an exhaustive model make the same predictions? Indeed, it can. If we set $p=\frac{1}{2}$, $1/u_x=50$, and $1/u_y=100$, then from the serial-exhaustive equation

$$E^s(\mathbf{T} \mid (y,y))=\tfrac{1}{2}(100+100)+\tfrac{1}{2}(100+100)=200$$

$$E^s(\mathbf{T} \mid (x,y))=\tfrac{1}{2}(50+100)+\tfrac{1}{2}(100+50)=150$$

and

$$E^s(\mathbf{T} \mid (x,x))=\tfrac{1}{2}(50+50)+\tfrac{1}{2}(50+50)=100$$

Unless we have some good solid evidence about the processing rate of element (x) (i.e., are critical elements processed at the same rate or faster than noncritical elements?), then we are forced to conclude that these data do not permit testability of self-terminating vs. exhaustive processing. This issue will be explored in more detail in Chapter 7.

The independence vs. dependence issue

Independence is a term used sometimes in a strict statistical or probabilistic sense and at other times simply to mean that two variables or dimensions are functionally unrelated to one another. We shall investigate the independence question within the former context, specifically asking whether certain processing events are stochastically independent as opposed to being correlated

in some fashion. Within this context there are a number of types of independence that might be of interest. We will investigate each of several of these via the class of model (e.g., serial or parallel) most naturally associated with it, and then use our equivalence mappings to inspect the opposite class. Although parts of the following discussions can be rather directly generalized to nonexponential distributions, we shall confine our attention to models with exponentially distributed intercompletion times.

The first kind of independence is that of successive intercompletion times. If this *across-stage* independence holds, then, for example, knowledge that the element in position a finishes first in t_{a1} time units tells nothing about how much longer it will take the element in b to finish. More specifically, if processing is serial and across-stage independence holds, then

$$\begin{aligned} f_{a1,b2}(t_{a1},t_{b2} \mid \langle a,b\rangle)&=f_{a1}(t_{a1} \mid \langle a,b\rangle)f_{b2}(t_{b2} \mid \langle a,b\rangle)\\ f_{b1,a2}(t_{b1},t_{a2} \mid \langle b,a\rangle)&=f_{b1}(t_{b1} \mid \langle b,a\rangle)f_{a2}(t_{a2} \mid \langle b,a\rangle) \end{aligned} \tag{4.22}$$

The exponential intercompletion time models that we focus on here obey this constraint. To see this, note that

$$\begin{aligned} f_{a1,b2}(t_{a1},t_{b2} \mid \langle a,b\rangle)&=\frac{f_{a1,b2}(t_{a1},t_{b2};\langle a,b\rangle)}{P^s(\langle a,b\rangle)}\\ &=\frac{pu_{a1}\exp(-u_{a1}t_{a1})u_{b2}\exp(-u_{b2}t_{b2})}{p}\\ &=u_{a1}\exp(-u_{a1}t_{a1})u_{b2}\exp(-u_{b2}t_{b2}) \end{aligned}$$

The marginal densities, conditioned on the completion order $\langle a,b\rangle$, are *defined* to be exactly the component parts of this expression; that is,

$$f_{a1}(t_{a1} \mid \langle a,b\rangle)=u_{a1}\exp(-u_{a1}t_{a1})$$

and

$$f_{b2}(t_{b2} \mid \langle a,b\rangle)=u_{b2}\exp(-u_{b2}t_{b2})$$

and thus independence is satisfied. Furthermore, we may conclude that all the present parallel exponential models also possess this property since each is equivalent to some serial exponential model possessing the property.

A second, closely related kind of independence is that of successive intercompletion times without element identity. This nonconditional across-stage independence essentially ignores all information pertaining to the identity of any processed element. Rather than differentiating as above, between, say, the times $\mathbf{T}_{a1}$ and $\mathbf{T}_{b1}$, we instead label the first intercompletion time as being of length $\mathbf{T}_1$ regardless of which element is completed first. If this independence property holds, knowledge that the first intercompletion time was of length t_1 tells us nothing about the length of the second intercompletion time. Analytically this property can be expressed as

$$f_{1,2}(t_1,t_2)=f_1(t_1)f_2(t_2) \tag{4.23}$$

As we have said, $f_1(t_1)$ includes times both when a is completed first and when b is completed first; that is,

$$f_1(t_1) = f_1(t_1; \langle a,b\rangle) + f_1(t_1; \langle b,a\rangle)$$
$$= pu_{a1}\exp(-u_{a1}t_1) + (1-p)u_{b1}\exp(-u_{b1}t_1) \tag{4.24}$$

The following result states the conditions under which nonconditional across-stage independence holds.

Proposition 4.5: Within the class of serial exponential models, nonconditional across-stage independence holds if and only if $u_{a1} = u_{b1}$ or $u_{a2} = u_{b2}$ or both. On the other hand, all parallel exponential models exhibit this type of independence.

Proof: Following Eq. 4.24 we see that

$$f_2(t_2) = f_2(t_2; \langle a,b\rangle) + f_2(t_2; \langle b,a\rangle)$$
$$= pu_{b2}\exp(-u_{b2}t_2) + (1-p)u_{a2}\exp(-u_{a2}t_2) \tag{4.25}$$

and

$$f_{1,2}(t_1,t_2) = f_{1,2}(t_1,t_2; \langle a,b\rangle) + f_{1,2}(t_1,t_2; \langle b,a\rangle)$$
$$= pu_{a1}\exp(-u_{a1}t_1)u_{b2}\exp(-u_{b2}t_2)$$
$$+ (1-p)u_{b1}\exp(-u_{b1}t_1)u_{a2}\exp(-u_{a2}t_2) \tag{4.26}$$

Multiplying Eqs. 4.24 and 4.25 and then simplifying leads to

$$\begin{aligned} f_1(t_1)f_2(t_2) = p^2[&u_{a1}\exp(-u_{a1}t_1)u_{b2}\exp(-u_{b2}t_2) \\ &+ u_{b1}\exp(-u_{b1}t_1)u_{a2}\exp(-u_{a2}t_2) \\ &- u_{a1}\exp(-u_{a1}t_1)u_{a2}\exp(-u_{a2}t_2) \\ &- u_{b1}\exp(-u_{b1}t_1)u_{b2}\exp(-u_{b2}t_2)] \\ &+ pu_{a1}\exp(-u_{a1}t_1)u_{a2}\exp(-u_{a2}t_2) \\ &+ pu_{b1}\exp(-u_{b1}t_1)u_{b2}\exp(-u_{b2}t_2) \\ &+ (1-2p)u_{b1}\exp(-u_{b1}t_1)u_{a2}\exp(-u_{a2}t_2) \end{aligned} \tag{4.27}$$

Since Eq. 4.26 contains no p^2 term, it is obvious that the coefficient of p^2 in Eq. 4.27 must equal zero if this type of independence (i.e., Eq. 4.23) is to hold. This occurs if either $u_{a1} = u_{b1}$ or $u_{a2} = u_{b2}$ or both. It so happens that when either of these two conditions hold, not only does the p^2 term of Eq. 4.27 drop out, but in addition the rest of Eq. 4.27 reduces to Eq. 4.26, as is sufficient for independence. Thus within the class of serial exponential models, necessary and sufficient conditions for nonconditional across-stage independence are that $u_{a1} = u_{b1}$ or $u_{a2} = u_{b2}$ or both.

The exponential parallel models, on the other hand, are again automatically independent in this sense, since it is always the case that $u_{a1} = v_{a1} + v_{b1} = u_{b1}$. □

These independence conditions appear to be reasonably weak. For instance, assume that $u_{a1} = u_{b1}$. Then it is quite possible, even though independence holds, that $u_{a1} \neq u_{a2}$ (or $v_{a1} \neq v_{a2}$), so that processing speeds up or slows down in the second stage and that $u_{a2} \neq u_{b2}$, so that the processing rate of the second element depends on which element was completed first.

Independence of total completion times

A third type of independence, more molar and in some respects more interesting than the preceding two, is independence of the overall times to process a and b. Here we are not discussing intercompletion times, but rather the *total completion* times for each element in positions a and b and are asking whether or not those times are independent. It is to be expected, however, that this type of independence will depend on relationships among the intercompletion times; indeed, this will become quite apparent as we proceed.

In Chapter 2, we defined the total completion time of an element as the total amount of time the system is in operation before the element of interest is completed. In parallel systems the total completion time is the same as the actual processing time since the system begins operating on all elements as soon as processing begins. In a serial system, however, the total completion time must also include the actual processing times (or equivalently the intercompletion times) of all completed elements. Thus one might expect this type of independence to be more natural to investigate in the parallel models.

Before we begin, we need to slightly modify our notation. Let $\mathbf{T}_a$ be the total (random variable) time to complete the element in position a, $\mathbf{T}_b$ the total time to complete b, and τ_a and τ_b specific values of $\mathbf{T}_a$ and $\mathbf{T}_b$. Independence of the parallel total processing times for a and b holds if and only if

$$g_{a,b}(\tau_a,\tau_b) = g_a(\tau_a)g_b(\tau_b) \tag{4.28}$$

We will begin our investigation of the sorts of models likely to satisfy this property by examining the marginals g_a and g_b. First, the marginal density of $\mathbf{T}_a$ is

$$g_a(\tau_a) = v_{a1}\exp[-(v_{a1}+v_{b1})\tau_a]$$
$$+\int_0^{\tau_a}\{v_{b1}\exp[-(v_{b1}+v_{a1})\tau_b]\}\{v_{a2}\exp[-v_{a2}(\tau_a-\tau_b)]\}\,d\tau_b$$

The first term on the right-hand side is the density when a finishes first and the second is when b finishes first; hence the integration over all possible values of $\tau_b \leqslant \tau_a$. Note that the second-stage component when b is completed first is $\exp[-v_{a2}(\tau_a-\tau_b)]$. In our standard intercompletion time notation this term is $\exp(-v_{a2}t_{a2})$ because when b is completed first, $\tau_b=t_{b1}$ and $\tau_a=t_{b1}+t_{a2}$ and so $t_{a2}=\tau_a-\tau_b$, as in the above expression. After some simplification, it can be shown that

$$g_a(\tau_a)=v_{a1}\exp[-(v_{a1}+v_{b1})\tau_a] + \frac{v_{b1}v_{a2}}{v_{a1}+v_{b1}-v_{a2}}\exp(-v_{a2}\tau_a)\{1-\exp[-(v_{a1}+v_{b1}-v_{a2})\tau_a]\}$$

The corresponding expression for the total completion time, $\mathbf{T}_b$, for the element in position b is

$$g_b(\tau_b)=v_{b1}\exp[-(v_{a1}+v_{b1})\tau_b] + \frac{v_{a1}v_{b2}}{v_{a1}+v_{b1}-v_{b2}}\exp(-v_{b2}\tau_b)\{1-\exp[-(v_{a1}+v_{b1}-v_{b2})\tau_b]\}$$

As it turns out, there is only a very restricted class of parallel exponential models that predict the product of these two marginals equals the joint density $g_{a,b}(\tau_a,\tau_b)$.

Proposition 4.6: Parallel exponential models in the form of Eqs. 4.3A and 4.4A predict independence of total completion times (i.e., Eq. 4.28) if and only if $v_{a1}=v_{a2}=v_a$ and $v_{b1}=v_{b2}=v_b$.

Proof (sufficiency): Letting $v_{a1}=v_{a2}=v_a$ and $v_{b1}=v_{b2}=v_b$, the marginals given above reduce to

$$g_a(\tau_a)=v_a\exp[-(v_a+v_b)\tau_a]+v_a\exp(-v_a\tau_a)[1-\exp(-v_b\tau_a)]$$
$$=v_a\exp(-v_a\tau_a)$$

and $g_b(\tau_b)=v_b\exp(-v_b\tau_b)$. Meanwhile, from Definition 4.2A,

$$g_{a,b}(\tau_a,\tau_b;\langle a,b\rangle)=v_a\exp[-(v_a+v_b)\tau_a]v_b\exp[-v_b(\tau_b-\tau_a)]$$
$$=v_a\exp(-v_a\tau_a)v_b\exp(-v_b\tau_b)$$
$$=g_a(\tau_a)g_b(\tau_b)$$

and

$$g_{a,b}(\tau_a,\tau_b;\langle b,a\rangle)=v_b\exp[-(v_a+v_b)\tau_b]v_a\exp[-v_a(\tau_a-\tau_b)]$$
$$=v_a\exp(-v_a\tau_a)v_b\exp(-v_b\tau_b)$$
$$=g_a(\tau_a)g_b(\tau_b)$$

which proves sufficiency.

(Necessity) We could prove necessity by showing that the product of the marginal densities equals the joint density only under the conditions of the proposition. However, a shorter and easier proof is as follows.

In order that $g_{a,b}(\tau_a,\tau_b;\langle a,b\rangle)=g_a(\tau_a)g_b(\tau_b)$, it is obvious that the left-hand side of this equation must decompose into two separate functions, of τ_a and τ_b, respectively, which each integrate to 1. Now

$$g_{a,b}(\tau_a,\tau_b;\langle a,b\rangle)=\{v_{a1}\exp[-(v_{a1}+v_{b1}-v_{b2})\tau_a]\}[v_{b2}\exp(-v_{b2}\tau_b)]$$

so constraints must be imposed such that

$$\int_0^\infty v_{a1}\exp[-(v_{a1}+v_{b1}-v_{b2})\tau_a]\,d\tau_a=\frac{v_{a1}}{v_{a1}+v_{b1}-v_{b2}}=1$$

Certainly this occurs only when $v_{b1}=v_{b2}$. The concomitant argument for $g_{a,b}(\tau_a,\tau_b;\langle b,a\rangle)$ demonstrates the necessity of $v_{a1}=v_{a2}$. □

Since this result allows the possibility that $v_a\neq v_b$, parallel exponential models possessing independence of total completion times can manifest element and position effects but not changes in the rates across stages of processing. It should thus come as no surprise that the serial model that is mathematically equivalent to the foregoing parallel model also predicts overall independence for a and b. Such a model is, of course, specified by the Proposition 4.1 mappings.

On the other hand, the very simple serial model with parameters $u_{a1}=u_{b1}=u_{a2}=u_{b2}=u$, and p, which is frequently seen in the experimental literature, yields a positive dependence between the total processing times of the elements in a and b. A positive dependency means that the *conditional* probability that a is completed before some time τ given b has already been completed by this time is greater than the *unconditional* probability that a is completed by time τ. In other words, knowledge that b has been completed increases the likelihood that a has also been completed. More precisely, a positive dependency exists if

$$P(\mathbf{T}_a<\tau\mid\mathbf{T}_b<\tau)>P(\mathbf{T}_a<\tau),\quad\text{for all } \tau>0$$

Proposition 4.7: The serial exponential model with parameters p and u (i.e., with $u_{a1}=u_{b1}=u_{a2}=u_{b2}=u$) predicts a positive dependence between the total processing times of the elements in positions a and b.

Proof: First note that

$$P(\mathbf{T}_a<\tau\mid\mathbf{T}_b<\tau)=\frac{P(\mathbf{T}_a<\tau\cap\mathbf{T}_b<\tau)}{P(\mathbf{T}_b<\tau)}$$

The numerator is just the probability that the total completion time of the second element completed is less than τ. In this serial model, the total completion time of this element is the sum of two random intercompletion times that are exponentially distributed with the same rate u. Therefore, the numerator is the cumulative distribution function of a two-stage gamma distribution with rate u (i.e., $1-e^{-u\tau}-u\tau e^{-u\tau}$). The denominator is the probability that b gets completed by time τ, and is composed of the probability that a is completed first (i.e., p) times the probability that both have been completed by that time (i.e., $1-e^{-u\tau}-u\tau e^{-u\tau}$) plus the probability that b is processed first (i.e., $1-p$) times the probability that it is completed by time τ (i.e., $1-e^{-u\tau}$). Putting all this together, we arrive at

$$P(\mathbf{T}_a<\tau \mid \mathbf{T}_b<\tau)=\frac{1-e^{-u\tau}-u\tau e^{-u\tau}}{p(1-e^{-u\tau}-u\tau e^{-u\tau})+(1-p)(1-e^{-u\tau})} \tag{4.29}$$

On the other hand, the unconditional probability that a gets finished by time τ, which we must compare with this conditional probability, is similar to $P(\mathbf{T}_b<\tau)$, an expression we have already found; that is,

$$P(\mathbf{T}_a<\tau)=p(1-e^{-u\tau})+(1-p)(1-e^{-u\tau}-u\tau e^{-u\tau}) \tag{4.30}$$

It suffices now to show that Eq. 4.29 is greater than Eq. 4.30 for all values of τ, u, and p. Thus, we wish to show that

$$\frac{1-e^{-u\tau}-u\tau e^{-u\tau}}{p(1-e^{-u\tau}-u\tau e^{-u\tau})+(1-p)(1-e^{-u\tau})}>p(1-e^{-u\tau}) +(1-p)(1-e^{-u\tau}-u\tau e^{-u\tau})$$

or equivalently that for all $\tau>0$,

$$1-e^{-u\tau}-u\tau e^{-u\tau}>[p(1-e^{-u\tau}-u\tau e^{-u\tau})+(1-p)(1-e^{-u\tau})] \times[p(1-e^{-u\tau})+(1-p)(1-e^{-u\tau}-u\tau e^{-u\tau})]$$

Multiplying the terms on the right-hand side and then simplifying reduces this inequality to

$$1-e^{-u\tau}-u\tau e^{-u\tau}>1-2e^{-u\tau}-u\tau e^{-u\tau}+e^{-2u\tau} +u\tau e^{-2u\tau}+p(1-p)u^2\tau^2e^{-2u\tau}$$

or

$$1-e^{-u\tau}-u\tau e^{-u\tau}-p(1-p)u^2\tau^2e^{-u\tau}>0 \tag{4.31}$$

Now $p(1-p)<\frac{1}{2}$ for all values of p, and so if

$$1-e^{-u\tau}-u\tau e^{-u\tau}-\frac{u^2\tau^2}{2}e^{-u\tau}>0, \quad \text{for all } \tau>0$$

then Eq. 4.31 is also true for all $\tau>0$. The left-hand side of this inequality is the distribution function of a three-stage gamma with rate u and is thus positive-valued for all $\tau>0$. □

The positive correlation of this serial model can perhaps be better understood by viewing its equivalent parallel counterpart; from Proposition 4.2 we see that the mimicking parallel model is found by setting

$$v_{a1}=pu, \quad v_{b1}=(1-p)u, \quad \text{and} \quad v_{a2}=v_{b2}=u$$

Since p is always strictly less than one and greater than zero, we see that the second-stage rates (v_{a2}, v_{b2}) are always greater than the first-stage rates (v_{a1}, v_{b1}), whereas in the independence case they were constrained to be equal across stages.

Intuition suggests that it might be reasonable to suppose that all parallel exponential models in which the second-stage rates are greater than the first-stage rates manifest a positive dependency with respect to total completion times. For instance, if we know that the second-stage rates are greater than the first-stage rates, then knowledge that the first element completed processing at some time before τ *increases* the likelihood that part of the second element's processing was governed by the faster rate, and thus the likelihood that this second element will also be computed by time τ is increased, and a positive correlation results.

To see this in a very simple manner, let us assume that the element in position b completes processing first at time t_{b1} and then investigate the likelihood that a gets finished during the interval (t_{b1}, τ). If independence holds, we know that $v_{a2}=v_{a1}=v_a$ and therefore that

$$P(t_{b1}<\mathbf{T}_{a2}<\tau \mid \mathbf{T}_{b1}=t_{b1})=1-\exp[-v_a(\tau-t_{b1})]$$

If this probability is increased, then a positive correlation must exist, and likewise if the probability is decreased, then a negative correlation exists. If we increase the second-stage rate such that $v_{a2}>v_{a1}$, then

$$P(t_{b1}<\mathbf{T}_{a2}<\tau \mid \mathbf{T}_{b1}=t_{b1})=1-\exp[-v_{a2}(\tau-t_{b1})]$$

A positive dependency now exists since

$$1-\exp[-v_{a2}(\tau-t_{b1})]>1-\exp[-v_a(\tau-t_{b1})]$$

Similarly, if we decrease the second-stage rates, the direction of this inequality will reverse and a negative dependency results. The following proposition summarizes our findings.

Proposition 4.8: If processing is parallel and exponential as in Eqs. 4.3A and 4.4A, *independence* of total completion times occurs whenever $v_{a1}=v_{a2}$ and $v_{b1}=v_{b2}$; a *positive dependence* occurs when $v_{a2}>v_{a1}$ and $v_{b2}>v_{b1}$; and a *negative dependence* results when $v_{a2}<v_{a1}$ and $v_{b2}<v_{b1}$. □

The straightforward serial concomitant to the negatively correlated parallel model is found from Proposition 4.1 by setting

$$u_1=u_{a1}=u_{b1}=v_{a1}+v_{b1}, \quad u_{a2}=v_{a2}<pu_1,$$

$$u_{b2}=v_{b2}<(1-p)u_1, \quad \text{and} \quad p=\frac{v_{a1}}{v_{a1}+v_{b1}}$$

In this model, there is a significant slowing down in the processing of the individual elements during a single trial. In the negatively correlated parallel model, the individual processing need not slow down nearly as drastically. In the serial case, u_{a1} must be substantially larger than u_{a2} or u_{b2}, but if processing is parallel, v_{a1} and v_{b1} need be only slightly greater than v_{a2} and v_{b2}, respectively.

The capacity issue

By *capacity* we shall mean the ability to get work done. In this sense, the capacity of information theory (e.g., Shannon & Weaver 1949) is one particular type of capacity. Recently, capacity in basic processing tasks has begun to command a considerable amount of attention from theorists and experimenters alike (Kahneman 1973; Norman & Bobrow 1975; Kantowitz, unpublished manuscript; Navon & Gopher 1979; Townsend & Ashby 1978).

To some degree, the capacity of a system seems less easily obscured in the processing structure than certain other issues since it can always be determined whether the overall system that is engaged in performing a task is of limited or unlimited capacity by varying the load that must be handled by the system. Capacity can then potentially be determined by examining the way in which errors and processing time change in conjunction with alterations in load. Thus, intuitively, a typical serial system is limited capacity because as n increases, the time necessary to complete all the elements also increases. On the other hand, pinning down the ultimate source of the limitation can be as tricky as any of the other issues. For example, in experiments where an observer must report back as many unrelated letters as possible from a briefly exposed visual display (the whole report paradigm), it has been difficult to designate the exact stage where the first major processing limits occur, the major hypothesis being a visual identification stage vs. a later short-term memory stage (see Chapter 11 for some new evidence on this particular problem).

No real physical system can ever be of absolute "unlimited capacity," and so the term is usually used to mean that the processing efficiency is not deleteriously affected at some particular level of processing, when the load is moderately increased. For instance, in a system with N parallel channels, each with its own independent source of capacity, the processing rate on each channel will not be altered for any load of n elements, so long as $n \leqslant N$. This system is therefore of unlimited capacity for $n \leqslant N$, at the level of the individual element or channel. The concept here of *level* is important, for processing time might well increase at some other level, for instance, at the level of exhaustive processing of all n elements; and in fact, it will in this parallel example, when the distributions are probabilistically independent and have nonzero variance.

Supercapacity is employed to denote systems that actually speed up as the load increases. A serial system could be built to speed up as n increases in such a fashion that the average time to process all n elements is constant. This type of system is unlimited capacity at the exhaustive level and supercapacity at the level of an individual element. Finally, *limited capacity* will denote a decrement in processing ability at some level as the load increases. A particularly interesting class of limited capacity parallel systems can be produced by assuming that the processing times tend to increase on the individual elements as the load n increases. We shall examine systems of these varieties below.

A conceptualization of capacity was proposed by Townsend and Ashby (1978) in which a system is pictured as being able to expend different amounts of energy from trial to trial. The idea here is that energy is expended to get work done and so a probability distribution on expended energy always leads to a probability distribution on work done, and thus the capacity-producing ability of a system can be characterized by either of these probability distributions. Two approaches can now be taken. First, the amount of energy expended, or work accomplished, can be fixed and a probability distribution determined on the time required to carry that amount of work out; or second, the processing duration can be fixed and a probability distribution can be produced that represents the likelihood that any given amount of energy or work is completed by that particular time.

In the present investigations, we are focusing on tasks that demand the processing of discrete objects, which we are typically referring to as *elements* to keep the discussion fairly general. Hence, the amount of work done in a particular duration can be given in terms of the number of elements completed, so that the energy or work dimension consists of the set of positive integers. On the other hand, we wish to apply our results to real continuous time, so the time dimension should continue to be the positive real line. In the Townsend and Ashby (1978) approach, the basic unit of capacity expenditure by a system is usually written in terms of its instantaneous energy, or its power, rather than its energy per se, where energy equals the integral or, in our case, the sum of power across an interval of time. For us, the amount of power expended at any point in time is either 0 or 1, depending on whether an element has completed processing. The stochastic process that results when time is fixed and the number of completed elements is studied is called a *counting process,* and when the number of elements finished is fixed and time is allowed to vary, the ensuing distribution is on the *waiting times* (e.g., Parzen 1962).

A rather general model for capacity in such systems is provided by the so-called nonhomogeneous Poisson process. It maintains the Poisson assumptions of independence of previous events and increments of one item (impulse etc.) at a time, but allows the rate of processing $w(t)$ to vary with time rather than being a constant w, as in the ordinary Poisson process.

Suppose t_0 designates the instant of the last completion and we wish to compute the probability density on the ensuing intercompletion time. In a nonhomogeneous Poisson process, this can be written as

$$f(t-t_0)=w(t)\exp\left[-\int_{t_0}^{t} w(t')\,dt'\right]$$

whereas the survivor function is just

$$\bar{F}(t-t_0)=\exp\left[-\int_{t_0}^{t} w(t')\,dt'\right]$$

The hazard function therefore turns out to equal the rate of processing at time t, that is,

$$h(t)=\frac{f(t-t_0)}{\bar{F}(t-t_0)}=w(t)$$

In a conventional Poisson process $h(t)$ is, of course, equal to a constant w. The expected intercompletion time in that case is $1/w$, and thus in an interval of length t the expected number of completions is

$$W(t)=\int_0^t w\,dt'=wt$$

Generalizing this definition to a nonhomogeneous Poisson process leads to

$$W(t)=\int_0^t w(t')\,dt'$$

It can now be shown (e.g., Parzen 1962; Papoulis 1965) that in a nonhomogeneous Poisson process, the probability distribution on the number of completions by time t is given by

$$P[\mathbf{K}(t)=k]=\frac{[W(t)]^k e^{-W(t)}}{k!}, \quad k=0,1,2,\ldots,t>0$$

The expected value of this random variable, that is, the expected number of completions in an interval of length t is, as with the conventional Poisson process, $E[\mathbf{K}(t)]=W(t)$. In our conceptualization, however, elements completed signify energy expended and work done. Thus, $P[\mathbf{K}(t)=k]$ is also the probability distribution on energy or work done in time t and $E[\mathbf{K}(t)]$ can be viewed as the expected energy expended or the average work done during that interval.

On the other hand, the power function is a random function equaling 0 or 1 at each point in time, and the probability that the power output equals 1 in any given small interval of time of length Δt is equal to $w(t)\Delta t$, that is, the conditional probability that an element will be completed in the next instant (given that one is being processed), $w(t)$, times the length of the interval Δt. Therefore, the expected power output at a point t in time is just

$$\begin{aligned}1\cdot P(\text{power}=1;t)+0\cdot P(\text{power}=0;t)&=P(\text{power}=1;t)\\&=w(t)\Delta t\end{aligned}$$

Thus, roughly speaking, the average power output at time t is just the hazard function $w(t)$. This result can also be obtained by differentiating the average energy function as follows:

$$\frac{dE[\mathbf{K}(t)]}{dt}=\frac{dW(t)}{dt}=\frac{d[\int_0^t w(t')\,dt']}{dt}=w(t)$$

It can be shown (e.g., Parzen 1962) that this derivative of average energy is equal to the expectation of the derivative of the random function $\mathbf{K}(t)$, representing energy. That is,

$$\frac{dE[\mathbf{K}(t)]}{dt}=E\left[\frac{d\mathbf{K}(t)}{dt}\right]=w(t)$$

Therefore the expected power $w(t)$ may be interpreted as the average power disbursed at time t.

Of course, the derivative of a random variable must be defined in the proper measure theoretic sense. Thus in the case of $d\mathbf{K}(t)/dt$, there must exist a function $K'(t)$ defined for $t\geqslant 0$, such that

$$\lim_{\Delta t\to 0} E\left\{\left[\frac{\mathbf{K}(t+\Delta t)-\mathbf{K}(t)}{\Delta t}-K'(t)\right]^2\right\}=0$$

All is well in this particular case.

As an aside, one should be aware that the statement in Chapter 3 that *any* density $f(t)$ on waiting times can be defined in terms of its hazard function $h(t)$ as

$$f(t)=h(t)\exp\left[-\int_0^t h(t')\,dt'\right]$$

is not contradicted by the nonhomogeneous process considered here. The point is that the above representation is for any density on the *first* completion time and therefore makes no assumption about what happens between later completions.

In much of this book we will operate under the assumption that $w(t)=w$ between completions, but we will often permit the rate w to vary across stages and serial positions and/or the elements themselves. Within our present characterization of capacity, the rate w is interpretable as the average power or the capacity at the individual element level. In addition, the sum of the w values can be interpreted as the capacity (or power) of the system at that stage. In the special case when the values of w are constant over the different elements and serial positions, a standard Poisson counting process results and the waiting time distribution is gamma.

We will now proceed to consider a few specific models of processing that are characterized by different capacity dynamics. Our concentration will be solely on mean processing times.

Serial models

Serial 1: the standard serial model

It will whet our intuition to begin with the well-known standard serial model. Let u be the capacity expressed as a rate, and for purposes of exposition suppose that we can draw u as if it half-filled a box, so that $u=$ ▄. We express u in this way as $\frac{1}{2}$ the "capacity" of the box or tank, so that there will be some room to add more if necessary.

We will work with $n=2$ for simplicity, show what happens to the capacity across stages 1 and 2, and give the formulas for the mean self-terminating and exhaustive processing times, as well as the average processing time for an individual element and the minimum processing time (the time to the completion of the first element) both for $n=2$ and for general n. A graph will exhibit the general form of the functions.

Let $\mathbf{T}$ be the random total completion time of processing, and to ease the notational problems of the section let $\mathbf{X}_i$ be the actual time it takes the system to process the element completed ith. Thus in both parallel and serial models the minimum processing time is $\mathbf{X}_1$, and when the total load is n elements, the average individual element (actual) processing time is $\bar{\mathbf{X}}=(1/n)\sum_{i=1}^{n}\mathbf{X}_i$. In serial exhaustive models the total completion time is $\mathbf{T}=\sum_{i=1}^{n}\mathbf{X}_i$, whereas in parallel exhaustive models $\mathbf{T}=\mathbf{X}_n$.

The capacity allocations for the two stages in the standard serial models are shown in Fig. 4.2, where it can be seen that at stage 2, the capacity originally allocated to the element in position 1 is now devoted entirely to the element in position 2. The X shows that the first element is completed at the end of stage 1.

Figure 4.3 reveals the four types of curves mentioned above for the case when each individual element is processed with rate u. The well-known formulas corresponding to the graphs are as follows:

	General	*Case of $n=2$*
$E_{\mathrm{EX}}(\mathbf{T})=$	$\dfrac{n}{u}$	$\dfrac{2}{u}$
$E_{\mathrm{ST}}(\mathbf{T})=$	$\dfrac{n+1}{2u}$	$\dfrac{3}{2u}$
$E(\bar{\mathbf{X}})=E(\mathbf{X}_1)=$	$\dfrac{1}{u}$	$\dfrac{1}{u}$

It is important to keep in mind that the "individual element" function and

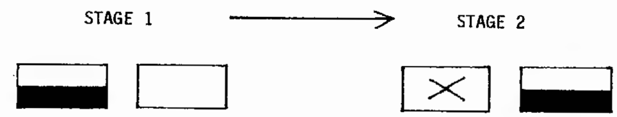


Fig. 4.2. Capacity allocations in the standard serial model. Each box represents one element, in the sense that the processing rate on that element is proportional to the degree to which the box is filled, with a half-filled box representing rate u. An X in a box signifies that the element has been completed.

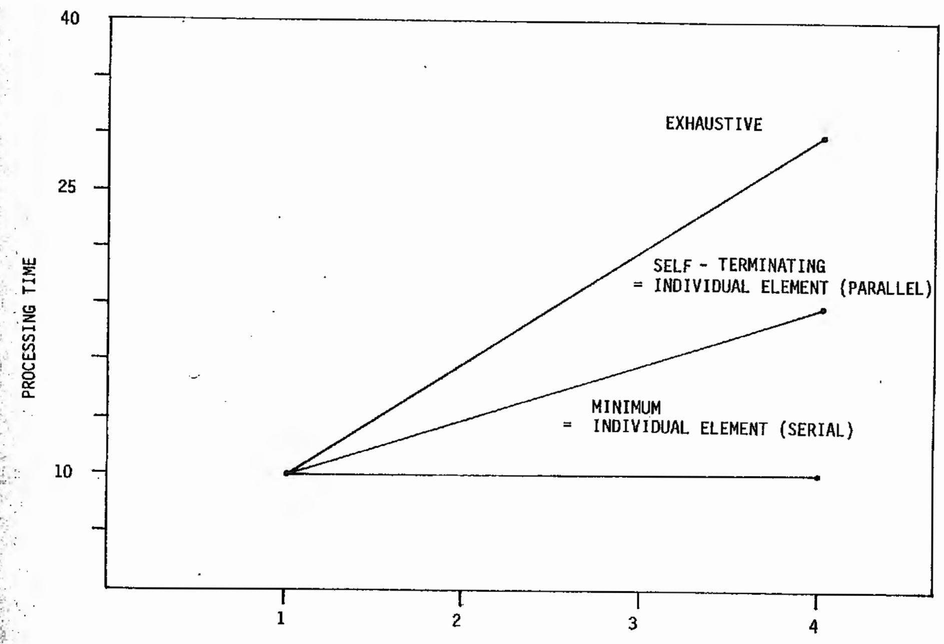


Fig. 4.3. Expected processing time vs. n curves for the standard serial model and for the fixed-capacity parallel model with reallocation in the case of exhaustive, self-terminating, minimum (i.e., first stage), and average individual element processing times.

curve will always refer to the actual processing time rather than the total completion time of the element. Thus, the self-terminating curve in Fig. 4.3 would be the average total completion time of an element but is not the average actual processing time in a serial system. In the succeeding section we will see that the self-terminating time equals the actual processing time of an individual element in the parallel model that is equivalent to the standard serial model (thus the notation in Fig. 4.3 on the bottom function).

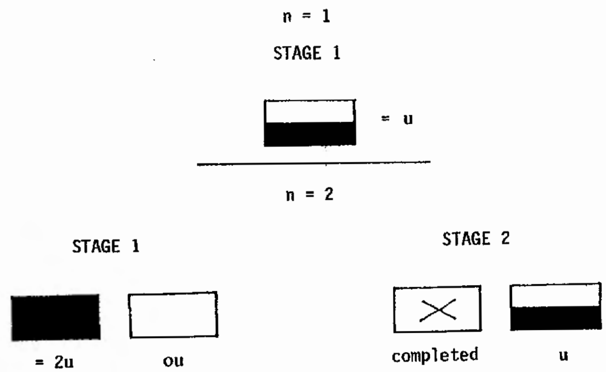


Fig. 4.4. Capacity allocations in the serial model with unlimited capacity at the self-terminating level.

Serial 2: unlimited capacity at the self-terminating level

The next serial model is less hackneyed. It is the serial model that is identical to an *unlimited capacity* independent parallel model, both being based on exponential intercompletion times (that assumption will be made, as stated above, throughout the remainder of the section, and will not be repeated again – though remember that we are scrutinizing only the means of the distribution). Basically, it describes a system that speeds up as n increases, but slows down or "gets tired" within trials. The capacity allocations for each stage for the cases when $n=1$ and $n=2$ are shown in Fig. 4.4.

Notice that when $n=2$, the stage 1 rate is twice that for $n=1$, but that under the increased load there is a slowdown in stage 2 back to the old rate. These properties mimic the fact that in an unlimited capacity independent parallel model the overall rate during stage 1 increases as n grows, but as elements are progressively completed during a trial the overall rate gradually diminishes. The corresponding parallel model will be examined later. The pertinent formulas are

General	*Case of* $n=2$
$E_{\mathrm{EX}}(\mathbf{T})=\sum_{i=1}^{n}\frac{1}{iu}$	$\frac{3}{2u}$
$E_{\mathrm{ST}}(\mathbf{T})=\frac{1}{n}\sum_{i=0}^{n-1}\sum_{j=0}^{i}\frac{1}{(n-j)u}=\frac{1}{u}$	$\frac{1}{u}$
$E(\bar{\mathbf{X}})=\frac{1}{n}\sum_{j=0}^{n-1}\frac{1}{(n-j)u}$	$\frac{3}{4u}$
$E(\mathbf{X}_1)=\frac{1}{nu}$	$\frac{1}{2u}$

The corresponding curves are given in Fig. 4.5. In this model, the mean

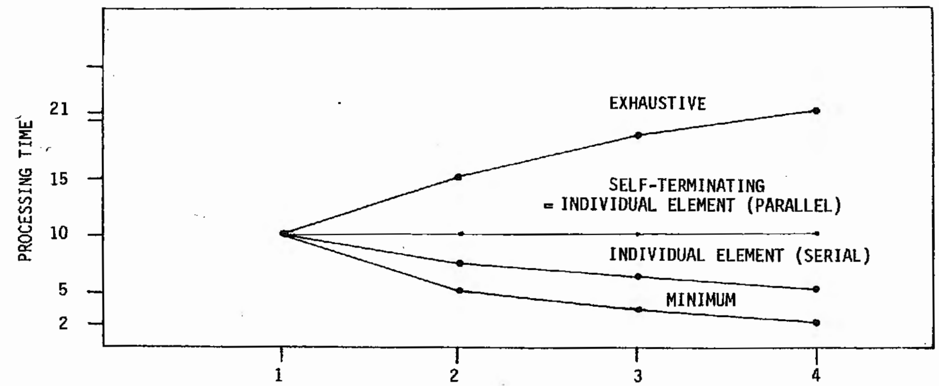


Fig. 4.5. Expected processing time vs. n curves for the serial model with unlimited capacity at the self-terminating level and for the independent parallel model with unlimited capacity.

exhaustive time increases as a function of n and so, like the standard serial model, it is limited capacity at the exhaustive level. However, unlike the standard serial model, which was limited capacity both on the self-terminating as well as exhaustive level, the present model is unlimited capacity at the self-terminating level and supercapacity at the minimum processing time level, since the average minimum processing time actually decreases as n increases. The reader is invited to try his or her hand at working out the predictions for the serial model that yields a flat exhaustive processing function, that is, that is unlimited capacity at the exhaustive level. What are the self-terminating, individual element, and minimum processing time levels in such a model?

Let us now turn to some parallel models. More of these will be investigated since they tend to be less well understood and since interesting differences in prediction are produced depending on whether it is assumed that capacity can be reallocated from completed elements.

Parallel models

Parallel 1: unlimited capacity and independent

This model gives identical predictions to its alter ego above (serial model 2), but has the advantage of being more natural in many situations. One possible realization of this model would be the system with N independent channels, mentioned earlier, in which each channel has its own separate source of capacity and each channel begins on one of the n $(n \leqslant N)$ elements as soon as processing starts. With sufficient separation of visual angle, elementary recognition may act in this way, with the distinct areas on the retina acting as the intakes to separate independent parallel channels (see

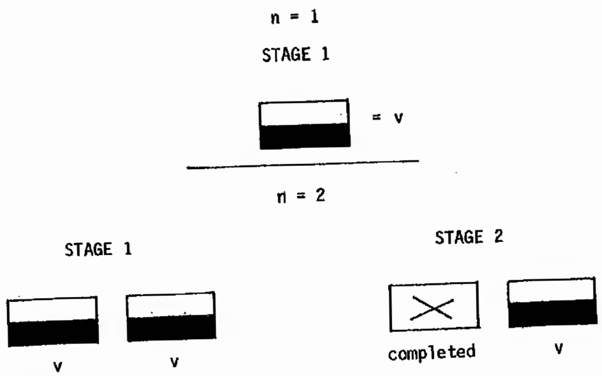


Fig. 4.6. Capacity allocations in the independent parallel model with unlimited capacity. A box half-full represents processing rate v.

Eriksen & Lappin 1965). We will present evidence later in this book that the processing involved in identifying and reporting as many letters as possible in a visual display may also be parallel and independent (but limited capacity; see Chapter 11).

The distribution of capacity in this model is illustrated in Fig. 4.6. Note that the capacity on the unfinished element is the same during stage 2 as in stage 1. For convenience, the element in position 1 is shown as being completed first, although either might be. The pertinent formulas are:

General	Case of $n=2$
$E_{\mathrm{EX}}(\mathbf{T})=\sum_{i=1}^{n}\frac{1}{iv}$	$\frac{3}{2v}$
$E_{\mathrm{ST}}(\mathbf{T})=E(\bar{\mathbf{X}})=\frac{1}{v}$	$\frac{1}{v}$
$E(\mathbf{X}_1)=\frac{1}{nv}$	$\frac{1}{2v}$

The supercapacity prediction at the level of the minimum processing time occurs here, because the rates of all n elements are summed during stage 1, the stage that leads to the first completion. The idea is similar to asking what the average trial of the first toss of a head is when one tosses 1 coin, 5 coins, or 10 coins simultaneously on each trial. Obviously, this average will be lower for 5 coins than for 1, and lower for 10 coins than for 5. Figure 4.5 graphs the predictions of the model.

Parallel 2: unlimited capacity with reallocation

The capacity resources start out the same in this model as in the previous one, since the stage 1 rate on every element is always v; however,

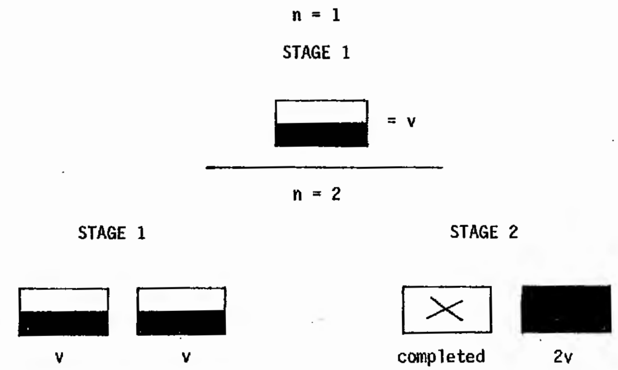


Fig. 4.7. Capacity allocations for the unlimited capacity parallel model with reallocatable capacity.

the other elements. Thus, the total capacity in any stage, as given by the sum of the rates, remains constant at nv. This situation is illustrated in the cases of $n=1$ and 2 in Fig. 4.7.

The corresponding mean processing times are as follows:

General	Case of $n=2$
$E_{\mathrm{EX}}(\mathbf{T})=\sum_{i=1}^{n}\frac{1}{nv}=\frac{1}{v}$	$\frac{1}{v}$
$E_{\mathrm{ST}}(\mathbf{T})=E(\bar{\mathbf{X}})=\left(\frac{n+1}{2}\right)\frac{1}{nv}=\frac{n+1}{2nv}$	$\frac{3}{4v}$
$E(\mathbf{X}_1)=\frac{1}{nv}$	$\frac{1}{2v}$

It is almost startling how much difference the reallocation property can make, actually rendering a system unlimited capacity at the exhaustive level and supercapacity at the others. Note, however, that the minimum time prediction is unchanged from the nonreallocation model since everything is the same in the two models during stage 1. The corresponding curves are given in Fig. 4.8.

Parallel 3: fixed capacity and independent

In this model, we find a description of a special kind of limited capacity where there is assumed to be a constant quantity of capacity, v, that is spread across the various elements before processing starts, but then stays the same, without reallocation, across the remainder of the stages. This allocation strategy is illustrated in Fig. 4.9, where it can be seen that the capacity tanks are only one-quarter full when $n=2$, indicating the spread of v (which half fills any one tank) across the two elements.

The mathematical expectations are:

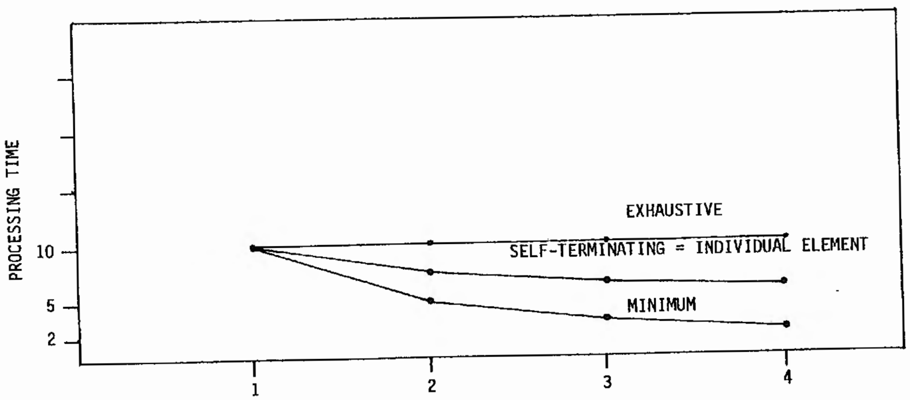


Fig. 4.8. Expected processing time vs. n curves for the unlimited capacity parallel model with reallocatable capacity.

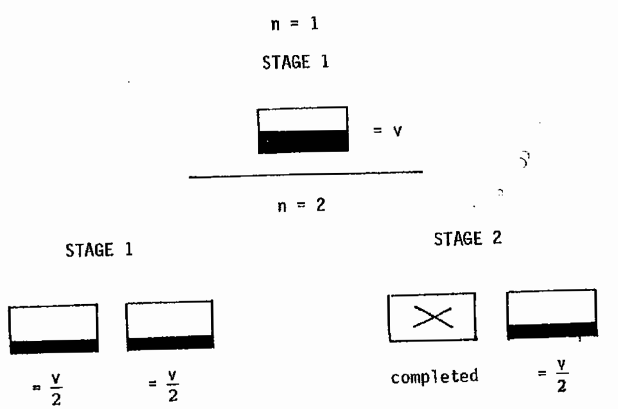


Fig. 4.9. Capacity allocations in the independent parallel model with fixed capacity.

General	Case of $n=2$
$E_{\mathrm{EX}}(\mathbf{T})=\dfrac{1}{n(v/n)}+\dfrac{1}{(n-1)(v/n)}+\cdots+\dfrac{1}{(v/n)}$	$\dfrac{3}{v}$
$=\dfrac{n}{v}\sum_{i=1}^{n}\dfrac{1}{i}$	
$E_{\mathrm{ST}}(\mathbf{T})=E(\bar{\mathbf{X}})=\dfrac{1}{(v/n)}=\dfrac{n}{v}$	$\dfrac{2}{v}$
$E(\mathbf{X}_1)=\dfrac{1}{n(v/n)}=\dfrac{1}{v}$	$\dfrac{1}{v}$

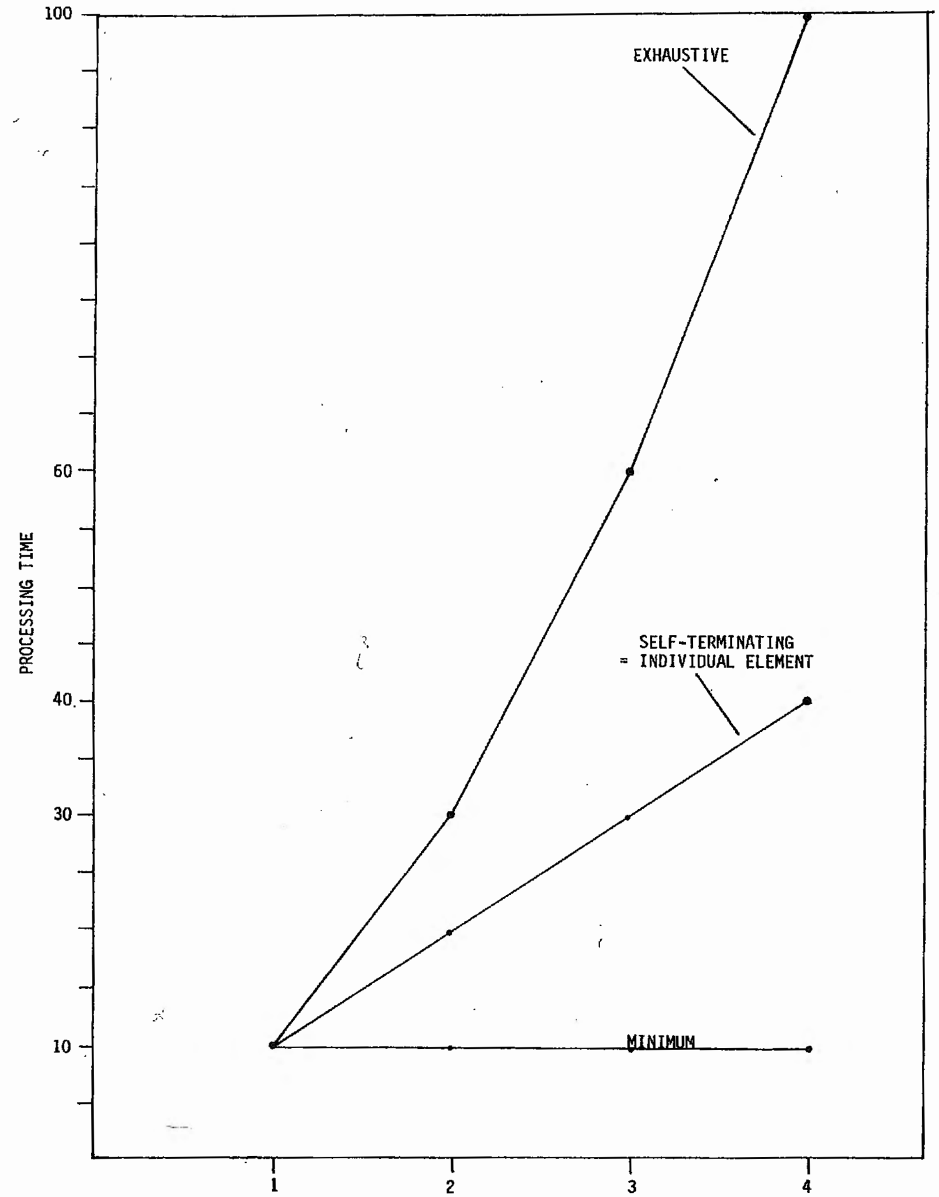


Fig. 4.10. Expected processing time vs. n curves for the independent parallel model with fixed capacity.

The reader can observe that overall, production rate decreases across stages as one would expect, to the final rate v/n; the rate on an individual element. The curves, which appear in Fig. 4.10, reveal that the exhaustive function is

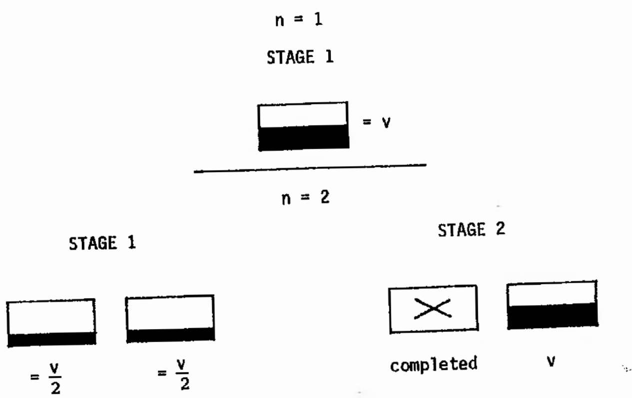


Fig. 4.11. Capacity allocations in the fixed capacity parallel model with reallocatable capacity.

positively accelerated and the self-terminating function is a straight line, with the same slope that the standard serial model predicts for its *exhaustive* curve. This model therefore makes rather dramatic predictions.

Parallel 4: fixed capacity with reallocation

The same type of limited capacity as in the previous model is found here, but now it can be reallocated within trials. Hence, we have the situation illustrated in Fig. 4.11, where the unfinished element acquires all of the available capacity during stage 2. The appropriate expressions are then as follows:

General	*Case of* $n=2$
$E_{\mathrm{EX}}(\mathbf{T}) = \dfrac{1}{n(v/n)} + \dfrac{1}{(n-1)[v/(n-1)]} + \cdots + \dfrac{1}{v} = \dfrac{n}{v}$	$\dfrac{2}{v}$
$E_{\mathrm{ST}}(\mathbf{T}) = E(\bar{\mathbf{X}}) = \dfrac{n+1}{2v}$	$\dfrac{3}{2v}$
$E(\mathbf{X}_1) = \dfrac{1}{n(v/n)} = \dfrac{1}{v}$	$\dfrac{1}{v}$

This model, as the reader undoubtedly anticipates, is equivalent to the standard serial model, and so the predictions for the mean processing times are shown in that illustration, Fig. 4.3. The only apparent disparity is that the mean individual element time is $1/u$ in the standard serial model but is $(n+1)/2v$ in the parallel model. This is only an illusory contradiction, however. The models are equivalent on the distributions of *completion* times (and therefore the means, which we are presently considering), which is all that is typically observable. The equivalence is manifestly not on the actual processing times, for then the models would be the same model. We are showing the actual processing means for an individual element, so these obviously must differ in the two mutually mimicking models.

In the literature, this model is often referred to as the *capacity reallocation model,* although, as we have seen, other parallel models possessing the reallocation property can be formulated. However, since this model mimics the popular standard serial model, it is the best known of the reallocation models. We shall examine it again in much more detail in Chapter 6.

Parallel 5: moderately limited capacity and independent

Finally, we examine a model that mimics the exhaustive prediction of the standard serial model, but is not identical to it. As we saw earlier, the standard serial model predicts positive dependencies between the total completion times of the elements, as does the preceding model, which *is* equivalent to it. The present model, on the other hand, predicts independence of total completion times.

In the present case, it will behoove us to consider the explicit mathematical formulas first, before the stage representations. In fact, for illustrative purposes let us see if we can derive the capacity requirements that lead to the linear exhaustive function $E_{\mathrm{EX}}(\mathbf{T}) = n/v$ that is characteristic of the standard serial model. The technique we will employ can be used in many different circumstances.

We begin by assuming that the capacity allocation to a particular element depends only on the processing load on the system. If there are n elements to be processed, let us call the rate on each element $v(n)$. In this case, the mean exhaustive processing time is

$$E_{\mathrm{EX}}(\mathbf{T}) = \sum_{i=1}^{n} \frac{1}{iv(n)} = \frac{1}{v(n)} \sum_{i=1}^{n} \frac{1}{i}$$

Since we want this mean to be a linear function of n, we know the following identity must be satisfied:

$$E_{\mathrm{EX}}(\mathbf{T}) = \frac{1}{v(n)} \sum_{i=1}^{n} \frac{1}{i} = \frac{n}{v}$$

This equation can be easily solved for the allocation function $v(n)$, which yields linear exhaustive curves

$$v(n) = \frac{v}{n} \sum_{i=1}^{n} \frac{1}{i}$$

We may observe that $v(n)$ decreases gradually as a function of n [in fact, approximately as $\log(n)/n$] and is therefore limited capacity but not so limited as in the fixed capacity models where $v(n) = v/n$.

Now let us look at the other functions:

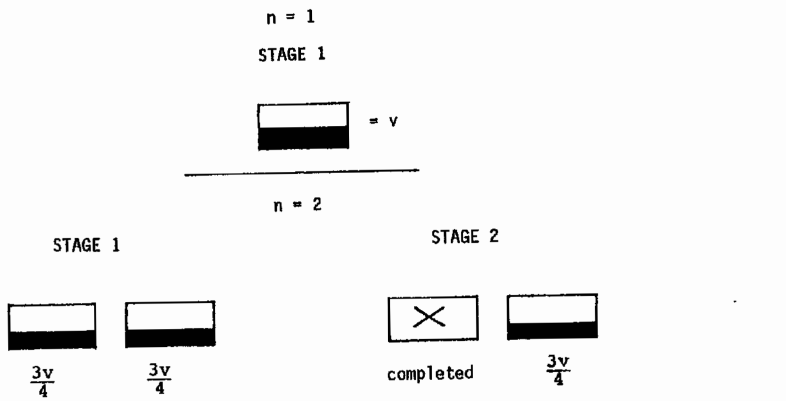


Fig. 4.12. Capacity allocations in the independent parallel model with moderately limited capacity.

General	Case of $n=2$
$E_{\mathrm{EX}}(\mathbf{T})=\dfrac{n}{v}$	$\dfrac{2}{v}$
$E_{\mathrm{ST}}(\mathbf{T})=E(\bar{\mathbf{X}})=\dfrac{1}{v(n)}=\dfrac{n}{v\sum_{i=1}^{n}(1/i)}$	$\dfrac{4}{3v}$
$E(\mathbf{X}_1)=\dfrac{1}{nv(n)}=\dfrac{1}{v\sum_{i=1}^{n}(1/i)}$	$\dfrac{2}{3v}$

We are now in a position to examine the stage representations and the graphs of these functions. In Fig. 4.12, it can be seen that the capacity allotted to an element in the $n=2$ case falls to $\frac{3}{8}$ (remember that $v=\frac{1}{2}$) as opposed to $\frac{1}{4}$ in the fixed capacity parallel model (Parallel 3), illustrating the fact that the present model is less limited in capacity.

The curves corresponding to the mean processing times are shown in Fig. 4.13. It is intriguing that the self-terminating curve, while in principle not strictly linear, could probably not be told from linear in a real experiment. The ratio of the exhaustive to self-terminating slopes might give somewhat more hope of differentiating the two models, since that ratio is close to 3:1 in the present case, rather than the classic 2:1 ratio found in the standard serial model. Still more hopeful would be an experiment where the minimum time could be assessed, since it decreases here but remains constant in the standard serial model and its parallel equivalent.

Before closing this section, it might be mentioned that, in all models we consider, the magnitudes of the rates, which we have implicitly assumed to be indicative of capacity, do not, by themselves, specify to what degree they are affected by factors such as stimulus intensity and to what degree they represent the inherent capacity of the processor. For example, even in what we might consider to be an unlimited capacity system, we might still expect pro-

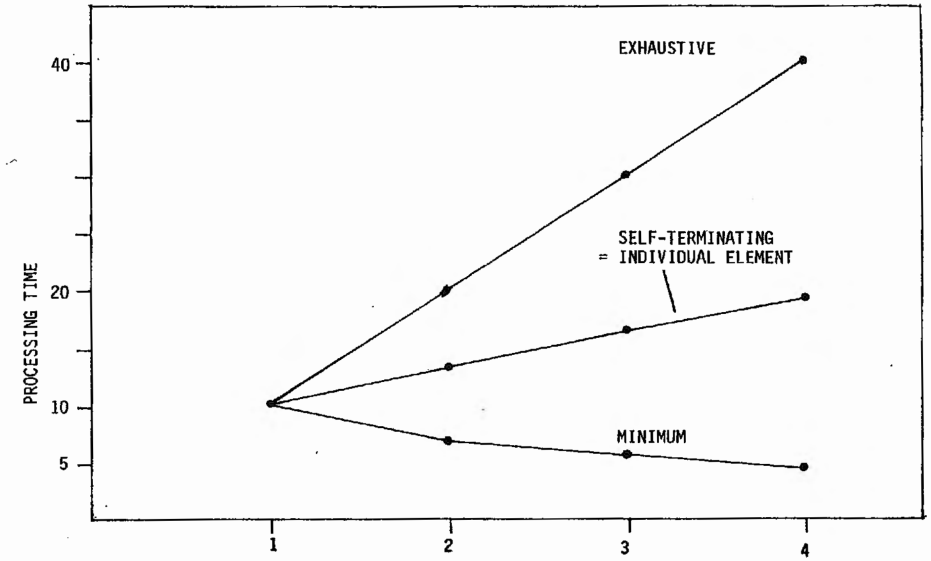


Fig. 4.13. Expected processing time vs. n curves for the independent parallel model with moderately limited capacity.

cessing rates to decrease if stimulus intensity is turned way down. Thus, it is probably more accurate to say that the processing rates represent the inherent capacity of the processor, *relative to the input characteristics*. In particular circumstances, it may be possible to break down the overall rate into components associated with separate aspects of the experimental setup (cf. Rumelhart 1970), but the organism's contribution has usually been implicit.

Generalization to distributions not based on exponential intercompletion times

A reasonable and pertinent question concerns what happens with regard to capacity when more general distributions are taken up. This question will be given some attention here, because although capacity is considered in a number of contexts throughout the book, there is no other entire chapter devoted exclusively to the issue (Chapter 8 comes the closest), unlike the parallel vs. serial or self-terminating vs. exhaustive processing issues. We will demonstrate below that the qualitative form of the exponential case of unlimited capacity in independent parallel processing is mimicked in the completely general case, when a natural generalization of "unlimited capacity" is utilized. However, although we can come up with reasonable definitions of limited capacity in the general case, it is not trivial to discover the qualitative form of the mean completion times in these other (nonunlimited capacity) instances; and, in fact, it may be that no such completely general statement can be made. After indicating something of an approach that may be useful in special cases but that fails to solve this problem in general, we briefly address an alternative method based on medians rather than means. We will occupy ourselves with independent parallel models in this section.

Unlimited capacity

It seems that an exceedingly natural generalization of unlimited capacity to broader distributions, when no serial position effects are present (i.e., the distributions are the same for all positions of the elements), is to define a model as unlimited capacity if and only if the distribution on each of the individual elements is constant (stays the same) no matter how many other elements are present. Thus, if $G_n(t)$ represents the individual element processing time distribution function when a total of n elements are to be processed, then under this definition of unlimited capacity $G_n(t)=G(t)$ for all values of n. This corresponds to the constancy of the rate v in the exponential models. Townsend and Ashby (1978) and Chapter 8 consider extensions of this notion where serial position effects are present.

We might mention that the general problem of the distribution of extrema (i.e., maxima and minima) has typically forced experts in the area to turn to statistics other than the expectation to describe the central tendencies, due to the mathematical intractability of expectations in this context (see Gumbel 1958 or Galambos 1978 as a standard source). Thus, we should not be too surprised if we encounter some obstacles in our attempt to extend our explorations.

The next result describes what we know about our four mean processing times in the general case of unlimited capacity.

Proposition 4.9: In the general case of independent parallel processing with unlimited capacity (as defined above), the following results are true:

(i) $E_{\mathrm{EX}}(\mathbf{T})=\int_0^\infty [1-G^n(t)]\,dt$, which is an increasing, negatively accelerated function of n, no matter what the distribution $G(t)$;

(ii) $E_{\mathrm{ST}}(\mathbf{T})=E(\bar{\mathbf{X}})=\int_0^\infty \bar{G}(t)\,dt$, which is a flat function, independent of n, no matter what the form of $G(t)$;

(iii) $E(\mathbf{X}_1)=\int_0^\infty \bar{G}^n(t)\,dt$, which is a decreasing positively accelerated function of n, for all $G(t)$.

Proof: (i) The distribution function of the maximum of n identically distributed random variables is $G^n(t)$ and the mean is just the integral of the survivor function. The way this function changes with n can be deduced from the first- and second-order differences. The first will show that the expectation increases with n. Define $E_{\mathrm{EX}_n}(\mathbf{T})$ as $E_{\mathrm{EX}}(\mathbf{T})$ when the total load on the system is n elements. Now

$$\Delta^1 E_{\mathrm{EX}_n}(\mathbf{T})=E_{\mathrm{EX}_n}(\mathbf{T})-E_{\mathrm{EX}_{n-1}}(\mathbf{T})$$
$$=\int_0^\infty G^{n-1}(t)[1-G(t)]\,dt>0 \quad \text{for all } n\geqslant 2$$

As indicated, there is always a positive increment in the magnitude so that $E_{\mathrm{EX}_n}(\mathbf{T})$ is an increasing function of n. Now, the second-order difference is

$$\Delta^2 E_{\mathrm{EX}_n}(\mathbf{T})=[E_{\mathrm{EX}_{n+1}}(\mathbf{T})-E_{\mathrm{EX}_n}(\mathbf{T})]-[E_{\mathrm{EX}_n}(\mathbf{T})-E_{\mathrm{EX}_{n-1}}(\mathbf{T})]$$
$$=\int_0^\infty G^{n-1}(t)[2G(t)-G^2(t)-1]\,dt$$

Because

$$2G(t)-G^2(t)-1=-[G^2(t)-2G(t)+1]=-[G(t)-1]^2$$

is always $\leqslant 0$ for all positive t, it follows that so is the integral and therefore

$$\Delta^2 E_{\mathrm{EX}_n}(\mathbf{T})\leqslant 0 \quad \text{for } n\geqslant 2$$

The exhaustive mean is thus shown to be an increasing, negatively accelerated function of n irrespective of the distribution function of individual elements, $G(t)$.

(ii) Obvious.

(iii) Left to the reader. □

Limited capacity

We will now attempt to extend the investigation to limited capacity independent parallel models. Unfortunately, our conclusions will not be as clear-cut as in the unlimited capacity case and therefore some readers may wish to skip over this section on a first reading. In order to take into account capacity changes we will write the distribution function for an individual element as $G_n(t)$ to indicate the dependence on n. It seems reasonable to postulate that, when capacity is limited, $G_n(t)$ is ordered in terms of n so that $G_n(t)\geqslant G_{n+1}(t)$; that is, for any positive t, the distribution functions are decreasing functions of n. This captures the limitation in capacity at the distributional level and is in line with subsequent developments on capacity in Chapter 8; in particular, this assumption implies that the individual element mean processing time increases as n increases. In addition, we know mean exhaustive processing time must increase since it did even with unlimited capacity.

We can determine the rate of increase of the mean exhaustive processing time function by examining the second-order difference,

$$\Delta^2 E_{\mathrm{EX}_n}(\mathbf{T})=[E_{\mathrm{EX}_{n+1}}(\mathbf{T})-E_{\mathrm{EX}_n}(\mathbf{T})]-[E_{\mathrm{EX}_n}(\mathbf{T})-E_{\mathrm{EX}_{n-1}}(\mathbf{T})]$$

as we did in the proof of Proposition 4.9. The key quantity in this difference is the function

$$2[G_n(t)]^n-[G_{n+1}(t)]^{n+1}-[G_{n-1}(t)]^{n-1} \tag{4.32}$$

If this expression is always of the same sign, the expected exhaustive processing time will always have the same acceleration as that sign; thus, it will be negatively accelerated if the above expression is less than zero and positive if it is greater than zero. If it equals zero, then the mean increases linearly. We can therefore write a test inequality,

$$2[G_n(t)]^n - [G_{n+1}(t)]^{n+1} - [G_{n-1}(t)]^{n-1} \overset{?}{<} 0$$

or more revealingly,

$$A(n) \overset{?}{<} \frac{A(n+1)+A(n-1)}{2}$$

where $A(n)=[G_n(t)]^n$ and so on. If the left-hand side of the test is always less than or equal to the right, $A(n)$ is said to be a convex (down) function of n, whereas it is said to be concave if it is always greater than the right-hand term. Our interest is in convexity holding for all $t>0$, and we will henceforth mean this when we refer to *convexity*. It would be more natural perhaps to have the condition of convexity on the original distribution functions rather than on powers of them. Fortunately, if convexity is in force through the original distribution functions, then it holds as above; that is,

$$G_n(t) < \frac{G_{n+1}(t)+G_{n-1}(t)}{2} \quad \text{for all } t>0$$

implies that

$$[G_n(t)]^n < \frac{[G_{n+1}(t)]^{n+1}+[G_{n-1}(t)]^{n-1}}{2} \quad \text{for all } t>0$$

The proof of this fact will be omitted here, but it turns out to go through because the power function p^n is also a convex function of $n\,(0<p<1)$. Intuitively, what we have in the present circumstance of limited capacity with convexity is that the distribution function gets shifted over to the right as n increases, but less so for larger n. Put another way, the capacity available for individual elements decreases in smaller and smaller amounts as n gets larger. This property is sufficient to cause the mean exhaustive curve to be negatively accelerated, as in the unlimited capacity case.

What happens when the individual *distribution* functions are concave in n instead of convex? Unfortunately, it is not necessarily the case that the function $A(n)$ is also concave. So, for now, we will have to be satisfied with the statement that if

$$[G_n(t)]^n > \frac{[G_{n+1}(t)]^{n+1}+[G_{n-1}(t)]^{n-1}}{2} \quad \text{for all } t>0$$

then the mean exhaustive curve will be positively accelerated, as in the fixed capacity, independent, and exponential model we dealt with earlier. Even worse than this is that in a number of ordinary cases, the second-order difference of the distribution functions is not always of the same sign for all $t>0$, or it may not be trivial to see whether it is or not. When this happens the only way we can determine whether or not the mean exhaustive function is positively or negatively accelerated (if either) is to integrate the function (Eq. 4.32) over t, since

$$\Delta^2 E_{\mathrm{EX}_n}(\mathbf{T}) = \int_0^\infty \{2[G_n(t)]^n - [G_{n+1}(t)]^{n+1} - [G_{n-1}(t)]^{n-1}\}\,dt$$

In fact, even in the exponentially based fixed capacity independent model we encountered earlier, $G_n(t)$ is not convex in n for all t. This can be economically demonstrated by treating n as if it were continuous and differentiating $G_n(t)$ twice with respect to n (the second derivative is positive at points $n=1,2,\ldots$ if and only if the second-order difference is positive at these points):

$$\frac{d^2G_n(t)}{dn^2} = \frac{d^2\{1-\exp[-(v/n)t]\}}{dn^2} = \frac{vt}{n^3}\exp\left(-\frac{v}{n}t\right)\left(2-\frac{v}{n}t\right)$$

Clearly, for any given value of n, this expression will be positive for some values of t and negative for others, and thus shows a lack of convexity for some t. Integrating this term does result in zero as it should, because we know from our earlier work that the individual element expectation of this model is a straight line. Taking the second derivative of the exhaustive distribution function,

$$[G_n(t)]^n = \left[1-\exp\left(-\frac{v}{n}t\right)\right]^n$$

ends in a complicated function of n and t that is difficult to evaluate. However, we earlier found that the expectation, garnered through our proficiency with exponential intercompletion times, is a positively accelerated function of n. Thus, although the second-order difference in the individual distribution functions (or in the case of completion time minima, the survivor functions) can in principle be of aid in drawing conclusions about the way expectations change with n, they may be nondiagnostic in pragmatic cases.

Quantiles and capacity

The moments of a distribution have pretty much occupied center stage in mathematical psychology, primarily because of mathematical tractability and widespread use in probability and statistics, although they are definitely not without their problems (see Ratcliff 1979). Occasionally quantiles such as the median have been employed, especially in reaction time, and some effective proselytization carried out in their behalf (e.g., Thomas 1971). It is virtually certain that moments will continue to be of considerable import in theorizing as well as empirical work, and much of the focus in the present book is on the moments, particularly the mean. Nevertheless, quantiles do possess some desirable characteristics, such as being less dependent on so-called outliers (extreme reaction times, often suspected of being unrelated to the basic psychological processes undergoing study).[6] When they have the

[6] One very attractive property of medians, which is not true of means, is that the median of any strictly monotonic function of a random variable is equal to the function of the median. In other words, suppose $\ell(\mathbf{T})$ is any strictly monotonic function of

additional advantage of mathematical tractability they can become attractive alternatives to moments.

This appears to be the case in the study of the effects of capacity on central tendency. As an illustrative example we will analyze the manner in which median processing times behave as a function of n in the case of the exponential, fixed capacity, independent parallel models. Then we will show how to ascertain the behavior of the median under more general capacity assumptions; finally, we demonstrate how one can construct a distribution function that will produce any desired median processing time vs. n curve.

First, in the case of exponential processing and fixed capacity we have:

Proposition 4.10: In the exponential case of independent parallel processing with fixed capacity, let $\text{med}(\mathbf{T})$ be the processing time median; then

(i) $\text{med}_{\text{EX}}(\mathbf{T}) = (-n/v)\ln[1-(\tfrac{1}{2})^{1/n}]$, which is an increasing positively accelerated function of n;

(ii) $\text{med}_{\text{ST}}(\mathbf{T}) = \text{med}(\bar{\mathbf{X}}) = (n/v)\ln 2$, which is a linear function of n;

(iii) $\text{med}(\mathbf{X}_1) = (1/v)\ln 2$, which is a constant and hence independent of n.

Proof: (i) Let $t_m = \text{med}_{\text{EX}}(\mathbf{T})$; then by definition

$$[G_n(t_m)]^n = \left[1-\exp\left(-\frac{v}{n}t_m\right)\right]^n = \frac{1}{2}$$

Solving for t_m yields

$$t_m = -\frac{n}{v}\ln\left[1-\left(\frac{1}{2}\right)^{1/n}\right]$$

Treating n as a continuous variable and taking the first and second derivatives of t_m with respect to n results in

$$\frac{dt_m}{dn} = \frac{1}{v}\left\{\frac{1}{1-(\frac{1}{2})^{1/n}}\left(\frac{1}{2}\right)^{1/n}\frac{1}{n}\ln 2 - \ln\left[1-\left(\frac{1}{2}\right)^{1/n}\right]\right\} \geqslant 0$$

and

$$\frac{d^2t_m}{dn^2} = \frac{1}{v}\ln^2(2)\frac{1}{n^3}\left(\frac{1}{2}\right)^{1/n}\frac{1}{1-(\frac{1}{2})^{1/n}}\left[\frac{(\frac{1}{2})^{1/n}}{1-(\frac{1}{2})^{1/n}}+1\right]$$
$$\geqslant 0 \quad \text{for } n=1,2,\ldots$$

Thus, t_m is an increasing positively accelerated function of n.

(ii) Letting $t_m = \text{med}_{\text{ST}}(\mathbf{T}) = \text{med}(\bar{\mathbf{X}})$ results in

$$G_n(t_m) = 1-\exp\left(-\frac{v}{n}t_m\right) = \frac{1}{2}$$

Solving for t_m yields the result.

---

the random variable $\mathbf{T}$; then $\text{med}[\ell(\mathbf{T})] = \ell[\text{med}(\mathbf{T})]$, where $\text{med}(\mathbf{T})$ is the median of $\mathbf{T}$. As an example of this property, suppose $\ell(\mathbf{T}) = \mathbf{T}^2$ (where we assume $\mathbf{T}$ is restricted to the nonnegative real line); then $\text{med}(\mathbf{T}^2) = [\text{med}(\mathbf{T})]^2$. On the other hand, it is well known that $E(\mathbf{T}^2) \neq [E(\mathbf{T})]^2$.

(iii) Finally, letting $t_m = \text{med}(\mathbf{X}_1)$, we have that

$$[\bar{G}_n(t_m)]^n = \left[\exp\left(-\frac{v}{n}t_m\right)\right]^n = \frac{1}{2}$$

Solving for t_m again produces the result. □

Note that in all three cases, the median follows the exponential expectation in terms of general curvature.

In the most general case, capacity effects are expressed only as a dependency of processing load on the individual element distribution function, via $G_n(t)$. How does the median behave as n is varied in this case? As an example, assume processing is exhaustive and for convenience let $t_m = \text{med}_{\text{EX}}(\mathbf{T})$. Then, by definition,

$$[G_n(t_m)]^n = \tfrac{1}{2}$$

Solving for t_m results in

$$\text{med}_{\text{EX}}(\mathbf{T}) = t_m = G_n^{-1}[(\tfrac{1}{2})^{1/n}]$$

so that the median is written as the inverse (distribution) function of the nth root of $\frac{1}{2}$, an inverse that always exists when the density g_n exists (as we usually assume) due to the monotonicity of $G_n(t)$. Note that the inverse is through the inverse function of t, not of n.

The converse problem arises when we are confronted with some particular median vs. n capacity curve, say $t_m = R(n)$, and we are asked to determine what sort of individual element distribution function could yield such a result. For instance, in the exhaustive case, if we assume unlimited capacity (i.e., $G_n = G$), we must solve the following equation for G:

$$t_m = G^{-1}[(\tfrac{1}{2})^{1/n}] = R(n)$$

Doing so yields

$$G(t_m) = G[R(n)] = (\tfrac{1}{2})^{1/n}$$

Thus G, as a function of $R(n)$, must equal $(\frac{1}{2})^{1/n}$. The general form of G can be discovered graphically by marking down points on the time axis corresponding to $R(n)$ for each value of n, and then drawing a point above each of these with the value of $(\frac{1}{2})^{1/n}$ on the ordinate of $G(t)$. Any distribution function satisfying the requisite relationship must go through these points. Of course, this does not determine the distribution uniquely, but it does give the general form that any such distribution has to assume, if it exists.

An example illustrating this procedure is given in Fig. 4.14. In Fig. 4.14a, a negatively accelerated median processing time vs. n function is shown. From this graph, five points on the individual element processing time distribution function can be determined. This is enough to tell us that, in this instance, $G(t)$ is also negatively accelerated, at least for t between 300 and 750 msec. This process is shown in Fig. 4.14b.

In the next chapter our progress is diverted somewhat as we consider the

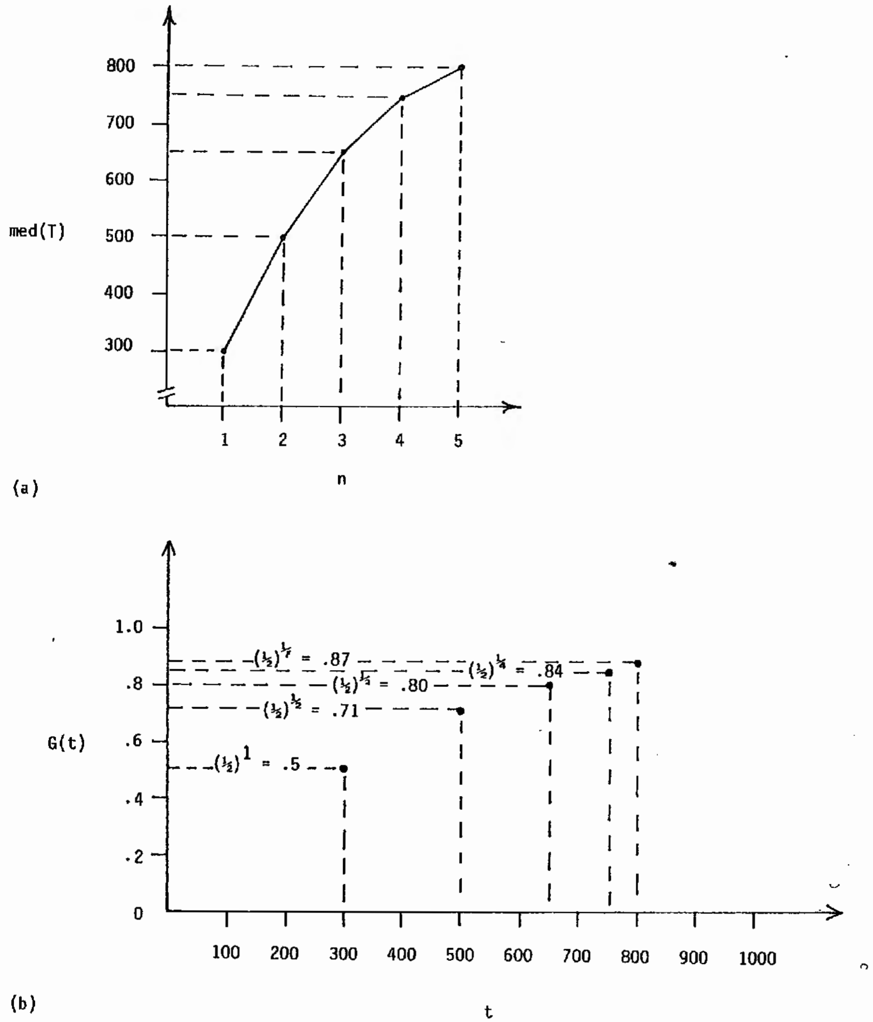
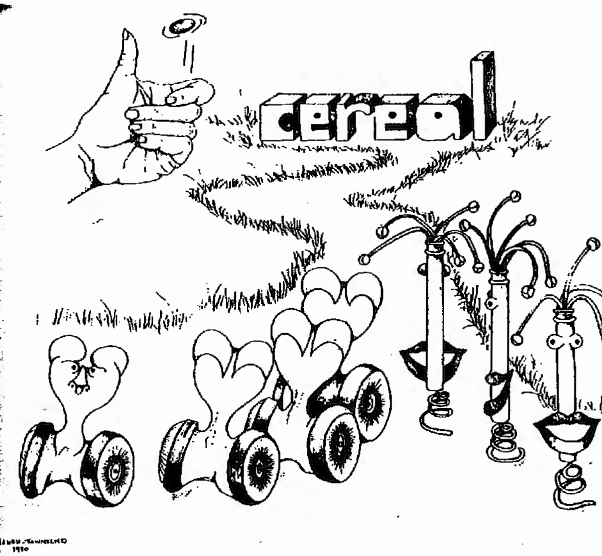
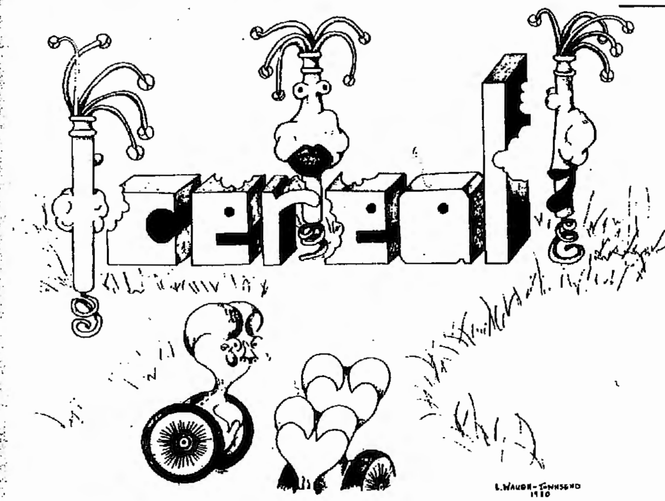


Fig. 4.14. Example of the way in which a median processing time vs. n curve can be used to reconstruct the individual element processing time distribution function when an unlimited capacity parallel model is assumed.

more specialized topics of compound processing models. Chapter 6 then picks up where we leave off now, but from a much more empirical perspective. Basically, it is an examination of the experimental paradigms to which the models we have developed in this chapter are most commonly applied. Since Chapter 5 does deal with a more specialized subject, the reader may wish to skip to Chapter 6 on first reading.



In a *compound processing system,* two regular systems stand available before processing begins. A probabilistic mechanism selects one or the other system to operate on a given trial. Here the two regular systems are parallel, one being represented by the Heart People and the other by the Spinach Heads (*top*). The coin toss (i.e., the probabilistic selection mechanism) determines that on the present trial, the Spinach Heads win and are permitted to process (e.g., eat) the object of processing (cereal) (*bottom*).



5 *Compound processing models*

We developed the notion of (regular) serial and (regular) parallel systems and models in Chapter 4. In general, regular serial models may be fixed-order or variable-order. In *variable-order serial models* there is a probability distribution on the different possible processing orders of the elements in the various

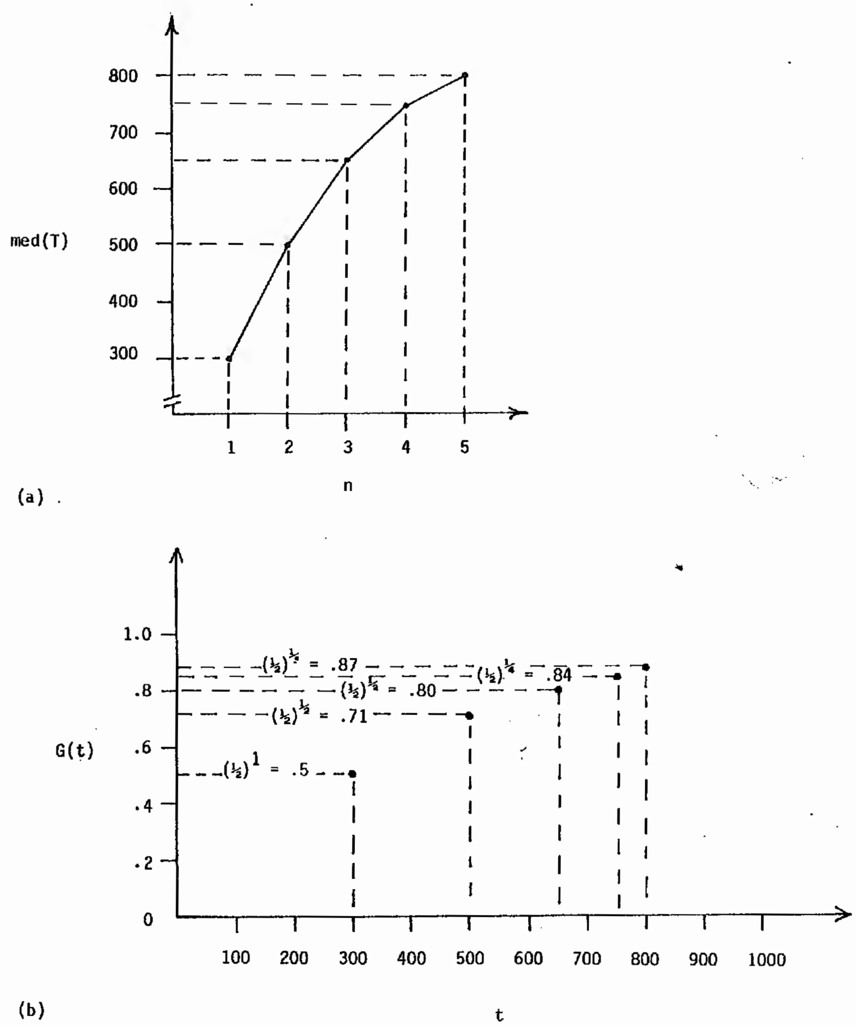
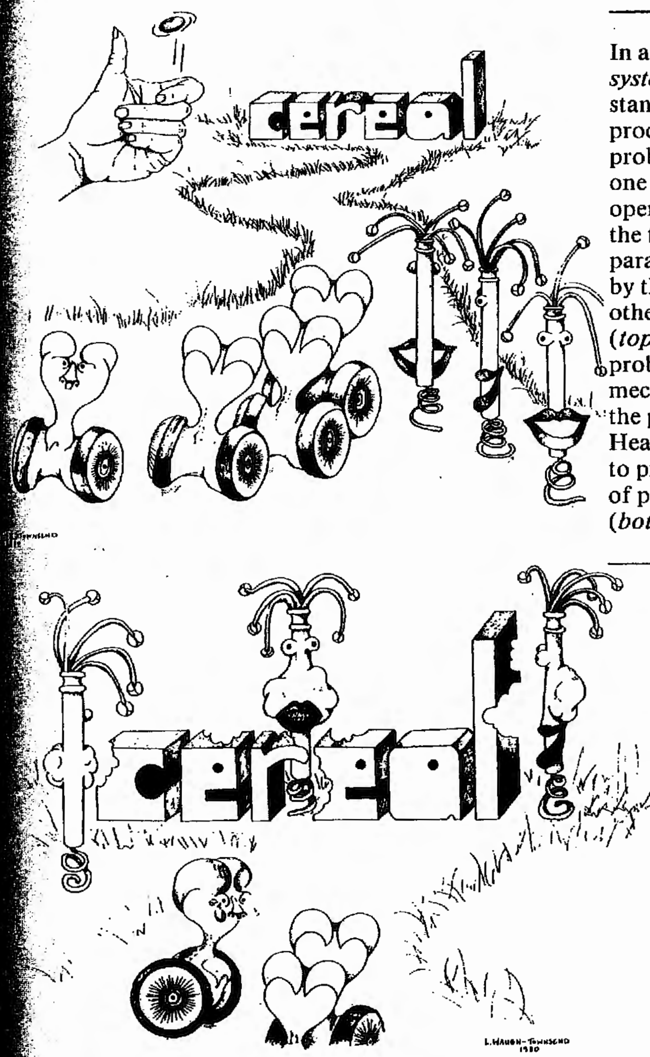


Fig. 4.14. Example of the way in which a median processing time vs. n curve can be used to reconstruct the individual element processing time distribution function when an unlimited capacity parallel model is assumed.

more specialized topics of compound processing models. Chapter 6 then picks up where we leave off now, but from a much more empirical perspective. Basically, it is an examination of the experimental paradigms to which the models we have developed in this chapter are most commonly applied. Since Chapter 5 does deal with a more specialized subject, the reader may wish to skip to Chapter 6 on first reading.



In a *compound processing system,* two regular systems stand available before processing begins. A probabilistic mechanism selects one or the other system to operate on a given trial. Here the two regular systems are parallel, one being represented by the Heart People and the other by the Spinach Heads (*top*). The coin toss (i.e., the probabilistic selection mechanism) determines that on the present trial, the Spinach Heads win and are permitted to process (e.g., eat) the object of processing (cereal) (*bottom*).

5 *Compound processing models*

We developed the notion of (regular) serial and (regular) parallel systems and models in Chapter 4. In general, regular serial models may be fixed-order or variable-order. In *variable-order serial models* there is a probability distribution on the different possible processing orders of the elements in the various positions – the order is according to position. On each trial, a particular order of processing the positions is chosen with its associated probability. A *fixed-order serial model* describes a special kind of serial system where the same order of processing is taken on each and every trial. This is, of course, produced by setting the probabilities on every order but one equal to 0. The order with nonzero probability then has a probability equal to 1 of being taken.

Regular parallel models treated in this work always are variable-order, as long as they are probabilistic, that is, stochastic. Intuitively this is because the particular order occurring on a given trial happens by chance, although some orders may be more likely than others. We refer to simultaneous processing models where certain orders never occur as a type of hybrid model. Naturally, the probability distribution on the different possible orders is related to the rates (or more generally to the joint probability distribution on processing times) on the various positions. Note again that if any specific order of processing has a probability of 0 of occurring, then the system or model is not truly parallel. Thus, for instance, a regular parallel model cannot actually be equivalent to a fixed-order serial model, because all the orders but one would have to have probability 0 of happening, that is, be impossible. However, a probabilistic parallel model can still approximate this type of behavior by setting the rates on all the positions but one close to, but not equal to, 0 during any one stage.

In this chapter, we look briefly at more complex serial and parallel systems, which accordingly permit more flexibility than the previous systems considered. These we term *compound systems and models*. In almost all other parts of the book we will be dealing with regular models and will drop the "regular" in such cases. Instances where compound or hybrid models are employed will be so designated. A compound serial system is defined as a system in which on each trial a unique serial system is selected from a set of different regular serial systems according to a discrete probability distribution. In general, a given regular serial system chosen from the set of such systems can have both its probability distribution on the possible processing orders as well as the set of rates on the various positions distinct from those of the other regular serial systems that are available in the set. There are many degenerate cases where the compound system reduces to a regular serial system. Suppose, for instance, that a set of regular serial systems contains only two serial systems, one being a fixed-order serial system in one direction (from position 1 to n) and the other a fixed-order system whose order of processing is the reverse (from position n to 1). The compound model of this system is equivalent to the model of a regular serial system with a probability distribution that places nonzero probabilities on only two orders – in this example, two opposite orders. Here, because the compound system is equivalent to a regular system, we would ordinarily use the regular serial model as a description. Typically, however, compound serial systems produce behavior of a more general nature than regular serial systems, and the compound *models* tend likewise to be more general.

Compound parallel systems are of a similar nature to compound serial systems. On each trial, a regular parallel system is selected from a set of such systems according to a probability distribution. Then, the particular system selected is employed on that trial. As with the serial variety, compound parallel systems may be degenerate in the sense that their models may be equivalent to regular parallel models, but typically they produce behavior of a more general nature.[1]

It is important to understand that even the rates (or generally conditional probability distributions) in a regular parallel system or model may change *within* a trial depending on which elements are completed first and in what order. As we saw earlier, for instance, the rate on c when it happens to finish last out of three elements can depend on whether a or b finished first.

However, in a compound parallel system, the rates on all the elements that apply even during the very first stage, before any of the elements are completed, can change from trial to trial. In a sense, it is as if a different parallel system were selected on each trial according to a set of probabilities. Thus, a particular system (or equivalently, a set of rates) might be selected (again with probability q), but with probability $1-q$ some other system (set of rates) would be selected.

Where might a compound system arise, in psychological terms? Within experimental situations, whenever circumstances change in some manner that makes it advantageous for the observer to alter either the way he or she allocates or distributes processing capacity in parallel processing or to vary the preferred order of processing (and possibly the rates as well) in serial operations, the possibility of compound systems comes up. Of course, some systems may not possess the ability to substitute one system for another in the manner that compound systems allow.

Suppose a target element previously presented must be compared with a

[1] We regretfully have had to modify the terminology used in earlier papers. There, *regular serial models* were referred to as *mixed serial models,* whereas an *unmixed serial model* was just a *fixed-order serial model* in the present usage. Further, *mixed parallel models* (Townsend 1972: 178) were what we are now calling *compound parallel models,* and *unmixed parallel models* were our present *regular parallel models.* Compound serial models have not been heretofore discussed. Thus, both *regular serial* as well as *compound parallel* models were referred to as *mixed*, and *regular parallel* as well as *fixed-order serial* models were called *unmixed.* The change in terminology amounts to a reconsideration of what is the appropriate system to call *regular.* Although this problem is to some extent a matter of convention, it now seems to us unnecessarily restrictive to demand that a serial system or model have a fixed order, in order to be called *regular serial.* Once this step has been taken, that is, to broaden the definition of *regular serial* to include variable-order serial models and systems, it then seems natural to refer to probabilistic combinations (i.e., probability mixtures) of the regular serial or parallel systems or models with a new adjective. For fear that retaining the term *mixed* would lead to additional confusion, we thought it preferable to employ the term *compound* for the more complex systems and their models.

subsequent visual display of two elements with one element on the right and the other on the left. The observer pushes a "yes" button if the target is present and a "no" button otherwise. Suppose further that 50% of the trials have the target element present in the second set and that the experimenter decides to initiate each trial with a prestimulus cue light. If the cue light is red, then the proportion of target-present trials on which the target is on the left is .75. If, on the other hand, the cue light is blue, then the proportion of target-present trials on which the target will be on the left is .25. Here is a circumstance where it makes eminent sense to shift one's attention from trial to trial, depending on the cue light. This could result in compound parallel or serial processing. One might therefore expect that the proportion of trials in which the observer allocates more attention to the left will depend on the frequency that the experimenter presents the red rather than the blue cue light (this proportion would be the value of q). Many other cognitive situations can be imagined in which compound processing may be important. For instance, in certain problem-solving tasks, it may make sense for the subject to occasionally adopt different priorities on a set of heuristic operations rather than to always follow the same routine. Finally, in real-life situations, relative to any given perceptual or cognitive milieu, compound parallel or serial processing may be more realistic than supposing that one always confronts a pattern of stimuli with the same parallel distribution of attention or the same priority of processing directions in serial operations. The structure of compound models is made in the spirit of the probabilistic views put forth by Brunswik (1955) and the learning models of Atkinson and Estes (1963).

In the remainder of the present section, we will deal only with models based on exponential intercompletion times, which overall produce probability mixtures of gamma distributions when one looks at the distribution on the sum of the exponentially distributed intercompletion times. This is a fundamental restriction since statements about whether compound serial and compound parallel (indeed about whether regular serial and regular parallel) models are equivalent depend on whether one is confined to one particular class of probability distributions defined on the intercompletion times.

We now indicate how regular and compound models differ from one another and how compound parallel models differ from serial compound models of the exponential intercompletion time type. We will not provide rigorous proofs since the algebra is simple but exceedingly tedious and quite uninstructive intuitively. One line that a rigorous demonstration can take is to assume equivalence and then by a series of steps using the method of undetermined coefficients (Courant 1936) show the two distributions cannot, in general, be equivalent. Thus, the region of parameter values where equivalence does not hold reveals the models that are outside the prediction scope of the other models. Our more intuitive approach will, however, also suggest the parameter restrictions that are necessary and/or sufficient for equivalence to occur. Later, we will give an illustration of how compound models might be used by reference to a study from the perceptual literature.

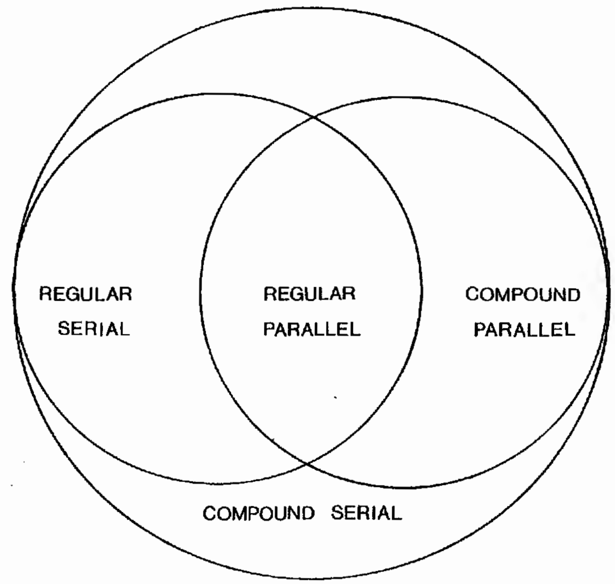


Fig. 5.1. Venn diagram showing containment relations among regular and compound models based on exponential intercompletion times.

Let us first attain an overview of the equivalence relationships among compound models based on exponential intercompletion times. The situation is shown in Fig. 5.1. It can be seen that the regular parallel models are contained in the class of regular serial models. This fact was demonstrated in earlier papers and in earlier chapters in the present work. However, it can be seen that compound parallel models are not contained in the class of regular serial models, nor are all compound serial models contained in the class of compound parallel models. It was indicated in Townsend (1972: 178) that the class of compound (there called "mixed") parallel models is not contained in the class of regular (there called "mixed") serial models, but compound serial models were not investigated. It is interesting to note that the class of compound serial models *does* contain the class of compound parallel models. Thus, the serial classes of models are more general, under our present taxonomy, than the parallel classes of models at the regular as well as the compound level of complexity.

At this point, we return to the mathematical part of the development. For ease of comparison, we repeat the formula for the regular parallel and serial models. It will suffice for our demonstration to treat the case of two elements, $n=2$, and to use only the first-stage rates. Because of the latter simplification, we can drop the designation of the stage usually given in the second subscript of the rates. Finally, in the compound systems, we will restrict our work to the instance where there are only two systems from which to choose.

The regular serial exponential densities, which by now can probably be anticipated by the reader, are

$$P(t<\mathbf{T}_a \leqslant t+dt \cap a \text{ is first}) = pf_a(t) = pu_a \exp(-u_a t) \qquad (5.1)$$

for instance, when the element in position a is completed before b and

$$P(t<\mathbf{T}_b \leqslant t+dt \cap b \text{ is first}) = (1-p)f_b(t) = (1-p)u_b \exp(-u_b t) \qquad (5.2)$$

for the case when the element in position b is finished first. Similarly, the regular parallel exponential densities for $n=2$ are the familiar

$$P(t<\mathbf{T}_a \leqslant t+dt \cap a \text{ is first}) = g_a(t)\bar{G}_b(t) = v_a \exp(-v_a t)\exp(-v_b t) \qquad (5.3)$$

for the order a first then b, $\langle a,b\rangle$, and

$$P(t<\mathbf{T}_b \leqslant t+dt \cap b \text{ is first}) = g_b(t)\bar{G}_a(t) = v_b \exp(-v_b t)\exp(-v_a t) \qquad (5.4)$$

for the reverse order $\langle b,a\rangle$.

The equations for the compound serial and parallel models are simply probabilistically weighted combinations, that is, probability mixtures of the more atomistic models. In the case of compound serial models then, we find

$$\begin{aligned} P(t<\mathbf{T}_a \leqslant t+dt \cap a \text{ is first}) &= rpf_a(t) + (1-r)p'f_a'(t) \\ &= rpu_a \exp(-u_a t) + (1-r)p'u_a' \exp(-u_a' t) \end{aligned} \qquad (5.5)$$

for the order $\langle a,b\rangle$, and

$$\begin{aligned} &P(t<\mathbf{T}_b \leqslant t+dt \cap b \text{ is first}) \\ &\quad = r(1-p)f_b(t) + (1-r)(1-p')f_b'(t) \\ &\quad = r(1-p)u_b \exp(-u_b t) + (1-r)(1-p')u_b' \exp(-u_b' t), \quad 0 \leqslant r,p,p' \leqslant 1 \end{aligned} \qquad (5.6)$$

for the processing order $\langle b,a\rangle$. The parameter r gives the probability with which system I (with rate u_a and probability p of processing a first) is selected. System II possesses the rate u_a' and probability parameter p'.

Let q be the probability that the parallel system described by (within-stage-independent) densities g_a, g_b is assigned and $1-q$ be the probability that the parallel system described by densities g_a', g_b' is assigned to the task.

Then the formulas for the compound parallel models are

$$\begin{aligned} P(t<\mathbf{T}_a \leqslant t+dt \cap a \text{ is first}) &= qg_a(t)\bar{G}_b(t) + (1-q)g_a'(t)\bar{G}_b'(t) \\ &= qv_a \exp[-(v_a+v_b)t] \\ &\quad + (1-q)v_a' \exp[-(v_a'+v_b')t] \end{aligned} \qquad (5.7)$$

for the instances when a is finished first, and

$$\begin{aligned} &P(t<\mathbf{T}_b \leqslant t+dt \cap b \text{ is first}) \\ &\quad = qg_b(t)\bar{G}_a(t) + (1-q)g_b'(t)\bar{G}_a'(t) \\ &\quad = qv_b \exp[-(v_a+v_b)t] + (1-q)v_b' \exp[-(v_a'+v_b')t], \quad 0 \leqslant q \leqslant 1 \end{aligned} \qquad (5.8)$$

for that set of trials or occasions when b is completed first.

By inspection we are able to observe that the serial expressions yield models that cannot be completely mimicked by the parallel models. This follows because the rates of Eq. 5.5, u_a and u_a', are totally distinct, in general, from those of Eq. 5.6. In the compound parallel models, the rates of Eq. 5.7 are, on the contrary, the same rates that appear in Eq. 5.8. Hence, whereas Eq. 5.7 can be made to equal Eq. 5.5 by judicious selection of the parallel parameters, it would then be impossible, with arbitrary u_b and u_b', to make the parallel expression Eq. 5.8 be equivalent to Eq. 5.6. How can the serial model be made to be mathematically the same as the parallel model? The answer is by restricting $u_a = u_b$ and $u_a' = u_b'$. Then we may set $r=q$ and use the parallel-to-serial mappings developed in earlier chapters. Looking in the opposite direction, we see that we have also shown that the parallel models are contained in the class of serial models so that any compound parallel model can be mimicked completely (that is, in distribution) but that not any compound serial model can be mimicked in distribution by a compound parallel model.

Now let us look back to the regular models. It is easy to perceive that the regular parallel class of models is contained in all the other classes. On the other hand, interestingly enough, the regular serial class of models is *not* contained in the class of compound parallel models. The reason is basically identical to that for the comparison between the compound parallel and serial models. That is, the rates in the regular serial model may be selected in a completely arbitrary manner for the two alternative processing orders $\langle a,b\rangle$ and $\langle b,a\rangle$. In contrast, the parallel rates for the separate orders cannot be so selected. Perusing Eqs. 5.7 and 5.8 again and comparing them with Eqs. 5.1 and 5.2, it becomes clear that neither is the class of compound parallel models contained in the class of regular serial models.

The parallel expressions of Eq. 5.7 and Eq. 5.8 give probability densities on the completion time of the first element to be processed that are probability mixtures of exponential densities, and these are not, in general, themselves exponential densities. Thus, the compound parallel family encompasses, but has members lying outside, the exponential family. On the other hand, the regular serial expressions reveal that serial processing can produce, by way of the generality of their rates, exponential densities for the two orders of completion that do not fall within the boundary of the class of compound parallel models of mixed exponential densities.

What constraints on the compound parallel and regular serial classes of models are required so that they produce equivalent densities on the minimum completion time? Since the regular serial models are based on exponential distributions, it follows that the compound parallel densities must somehow reduce to these also. This is accomplished by setting $v_a + v_b = v_a' + v_b'$, since then Eq. 5.7 and Eq. 5.8 become exponential densities with the same rate $v_a + v_b$ and coefficients given by $qv_a + (1-q)v_a'$ and $qv_b + (1-q)v_b'$, respectively. It is also easily shown that there is then a regular parallel model equivalent to this special reduced compound parallel model, so that we have been forced to restrict the parallel models to the regular class. Going the other way, we must set the serial rates equal, that is, $u_a = u_b$, just as we had to do with the regular parallel and serial models to induce equivalence.

It may be recalled from Chapter 4 that the regular (exponential) serial models can yield behavior that the regular (exponential) parallel models cannot predict, although these differences may be difficult to employ in usual experiments due to the use of only the means, or to problems associated with uncovering the actual processing densities (i.e., the intercompletion time densities). Unfortunately, matters are even more difficult in the case of compound models.

Although the compound parallel and regular serial models are not ordinarily equivalent, even with parameters chosen so that the classes are *not* reduced (and therefore the specific models are within mathematically disparate families), the appearance of the processing densities can look very much alike. In fact, they can appear sufficiently similar that it may be too much to expect to readily differentiate them experimentally, even neglecting the problems of extracting the intercompletion time densities out of the overall composite distribution that includes encoding time, motor latency, and the like. Figure 5.2 shows a particular regular serial model and a compound par-

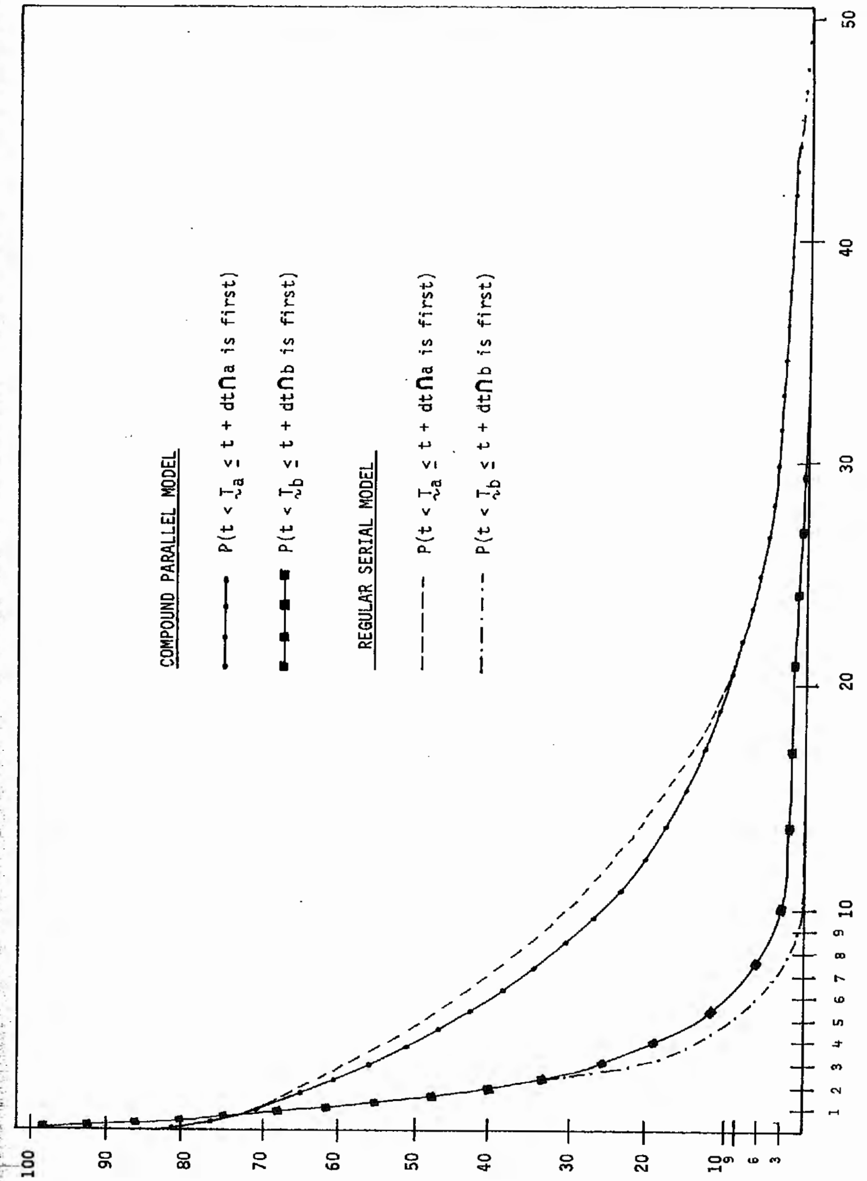


Fig. 5.2. Example comparison of compound parallel and regular serial density functions. Here

$P(t < \mathbf{T}_a \leqslant t + dt \cap a \text{ is first})$

$\quad = qv_a \exp[-(v_a + v_b)t_a] + (1-q)v_a' \exp[-(v_a' + v_b')t_a]$

$P(t < \mathbf{T}_b \leqslant t + dt \cap b \text{ is first})$

$\quad = qv_b \exp[-(v_a + v_b)t_b] + (1-q)v_b' \exp[-(v_a' + v_b')t_b]$

for a compound parallel model, and

$P(t < \mathbf{T}_a \leqslant t + dt \cap a \text{ is first}) = pu_a \exp(-u_a t_a)$

$P(t < \mathbf{T}_b \leqslant t + dt \cap b \text{ is first}) = (1-p)u_b \exp(-u_b t_b)$

for a regular serial model. The relevant parameters are as follows:

$q = .8 \quad v_a = .09 \quad v_b = .01$

$1 - q = .2 \quad v_a' = .05 \quad v_b' = .45$

for compound parallel parameters, and, for regular serial parameters,

$u_a = .1 \quad 1 - p = .2$

$u_b = .5 \qquad p = .8$

allel model. The parameters of each have been chosen so that they are far outside the reduced classes of models that permit equivalence. Indeed, the graphs are not precisely the same, but it can be seen that they are rather close; and the parameter values were selected quite roughly with no systematic attempt to produce the closest possible fits of one to the other. To paraphrase, the mixed exponential distribution of the compound parallel model is able to *approximate* the unique exponential behavior of the regular serial model, which regular parallel models were unable to do.

An experimental example

Let us return for a moment to develop a little further the hypothetical experiment given earlier where a red or blue cue light could precede a stimulus display. It may be reasonable to suppose that a "parallel" observer always employs a parallel system with rates $v_a < v_b$ on blue cue light trials. That is, the observer is using cues for a perfect guide on when to switch systems. In this case let $q = .75$, although the probability that the first system is employed given a red light occurred is 1. It is important to observe that in such a situation, where an external variable controls the operating system, it may be feasible to observe the separate systems. In other words, although if analyzed without knowledge of the cue light structure the responses would be composed of compound systems, when the data is conditioned on which cue light occurred it represents the processing of a *regular* serial or parallel model.

Recall that if the red cue light appeared, the previously presented target was displayed on the left with probability .75 and on the right with probability .25. If a blue cue light was shown, the probabilities of target placement were reversed.

Exactly what the rates and order probabilities will be in a serial system is usually impossible to specify before the experiment. If

$$\frac{v_a}{v_a + v_b} = .75 \text{ (red light)} \quad \text{and} \quad \frac{v_a'}{v_a' + v_b'} = .25 \text{ (blue light)}$$

in an independent parallel system, then the frequency of times that the left element is completed first matches the frequency of times that the target appears on that side. On the other hand, if able and so disposed, the observer might set $v_b = 0$, $v_a' = 0$ so that

$$\frac{v_a}{v_a + v_b} = 1 \quad \text{and} \quad \frac{v_a'}{v_a' + v_b'} = 0$$

retaining all attentional capacity for the position most likely to contain the target. In this latter possibility, the observer literally becomes a serial processor, with processing order determined by the cue light. Although the serial order is deterministic on any one trial, over trials we observe a variable-order serial system in operation.

Of course, typically the experimenter does not directly control the mixing parameter q, although there may be many reasons for the observer to change it. In particular, before being quite practiced in a task, no relatively fixed set of parallel rates may have been settled on so that the observer is essentially acting as a compound parallel system. Similarly, nature may or may not give cues to organisms in particular situations that they may employ to alter their resource allocations via rates and/or processing orders.

In such experiments as the above cued task, one may examine RT as the relevant dependent variable pursuant to testing compound models. However, it is sometimes feasible also to investigate accuracy patterns.

One experiment that has a structure close to these ideas is that of Shaw and LaBerge (1971). In that study, the observers read displays of letters for several milliseconds and indicated which of three potential target letters were present among the others. The instructed reading path (order) through the display was manipulated by assigning differential point values to the two separated display locations. The accuracy varied with which path was assigned and the authors argued from these results that processing was serial. While they certainly suggest differential attention, a possible change of parallel rates across their conditions (i.e., compound parallel processing) currently appears to us at least as reasonable as a serial interpretation.

Without getting into the details of the design, we note that there were two "orders" of processing the two stimulus positions that were reinforced on a between-session basis. This leads to an opportunity, as in the hypothetical cue light experiment above, to separate the two processing orders or parallel attention distributions. We must therefore emphasize that we are actually fitting distinct *regular* models to the two opposed reinforcement conditions. The main dependent variable was the proportion correct at the two distinct locations. In order to illustrate the possibilities for compound parallel and serial models here, we attempt to fit some reasonable candidates to the mean proportion correct taken across observers. Of course, for the strongest conclusions to be drawn, fits should be obtained for individual observers. As we will see, a compound independent parallel model is guaranteed to make perfect predictions. Nevertheless, the present work will serve to indicate which of the specific models can adhere to the average pattern of accuracy with any degree of success. The data are from Shaw and LaBerge (1971: Experiment 2).

In order to obtain predictions for some models, we first note that the display duration employed by Shaw and LaBerge serves to limit the time interval that the observer has available to process (i.e., identify) the stimuli. We shall make the simplifying assumption that the internal available processing duration equals the maximum exposure duration given to any observer, 23 msec.

The exact value of the internal duration is not of critical concern; rather, the values of that duration relative to the rates are what matter. This will be discussed further below.

We first fit three compound serial models and then look at an independent parallel model. We will confine ourselves to models that have no more parameters than there are data points, that is, to models with at most four parameters.

Equations 5.9 and 5.10 give the first compound serial model's (model I) predictions for the two reinforcement conditions, which we call A and B, respectively. Condition A refers to the preferred order $\langle a, b\rangle$, that is, the observer is paid more for being correct on the first position (which we call a) than on the second (b), and condition B, of course, refers to the opposite: $\langle b, a\rangle$ is the preferred order. Let $P_i(\text{err} \mid J)$ be the probability of an error in condition $J = A, B$ when the target was in position $i = a, b$. The symbols p_A and p_B are the probabilities of processing position a first in conditions A and B, respectively. Let us briefly analyze condition A; condition B is handled similarly. The data are shown in the right-hand column.

Compound serial model I

	Theoretical	Data	
Condition A			
	$P_a(\text{err} \mid A) = p_A e^{-ut}\cdot\frac{2}{3} + (1-p_A)[e^{-ut} + ute^{-ut}]\cdot\frac{2}{3}$	.155	(5.9)
	$P_b(\text{err} \mid A) = p_A[e^{-ut} + ute^{-ut}]\cdot\frac{2}{3} + (1-p_A)e^{-ut}\cdot\frac{2}{3}$	.413	(5.10)
Condition B			
	$P_a(\text{err} \mid B) = p_B e^{-ut}\cdot\frac{2}{3} + (1-p_B)[e^{-ut} + ute^{-ut}]\cdot\frac{2}{3}$	.265	(5.11)
	$P_b(\text{err} \mid B) = p_B[e^{-ut} + ute^{-ut}]\cdot\frac{2}{3} + (1-p_B)e^{-ut}\cdot\frac{2}{3}$	.150	(5.12)

With probability p_A the observer is assumed to process a first and the probability that he or she has not completed a by time t is

$$P(a \text{ uncompleted in } t \mid a \text{ is first}) = \int_t^\infty ue^{-ut'}dt' = e^{-ut}$$

If it is not completed, then the observer must guess, and in the Shaw and LaBerge experiment there were three signal possibilities, so we can assume that the probability of guessing *incorrectly* is $\frac{2}{3}$. On the other hand, with probability $1-p_A$ position b is processed first and hence the observer can fail to complete position a if neither are finished (with probability e^{-ut}) or if exactly one of the elements is completed (the one in position b). The latter probability may be computed directly, as in the case of e^{-ut}, or it may be realized that the probability of completing exactly one element in the time interval t is just that found from the Poisson distribution, and hence is $ut\cdot e^{-ut}$, and again the guessing factor is then appended. The position b expression is just the reverse of the a expression, because with probability p_A the observer can only be correct by *perception* on position b by completing both positions in time t; but if position b is processed first (which occurs with probability $1-p_A$), then only one element need be completed – so an incor-

rect response can occur in the latter instance only if nothing is finished by time t. The terms for condition B are analyzed in an analogous manner. The numbers appearing at the far right in all four cases are the averages computed from the experimental values (Table 2 in Shaw and LaBerge 1971).

Rather than perform a numerical fit via a chi-square or least squares minimization, we shall proceed by estimating the parameters from some of the conditions and then trying them out in others. If the model fits, the same parameters should work in all the appropriate places. In this first compound serial model there are 3 parameters, the two p terms and the u, one less parameter than there are degrees of freedom in the data.

Note first that adding Eq. 5.9 and Eq. 5.10 effectively eliminates p_A and allows us to attempt to locate a value of u which, when used with $t = 23$ msec, will yield $.155 + .413 = .568$. An estimate of u of $\hat{u} = .06$ produces this value. We now go back to Eq. 5.9 to estimate p_A algebraically, employing the parameter $\hat{u}$ we just estimated and, of course, $t = 23$. Our estimate of p_A turns out to be $\hat{p}_A = 1.05$, which is impossible, and we therefore already suspect that this model is not correct. Plugging back the largest acceptable value of $\hat{p}_A = 1$ yields $P_a(\text{err} \mid A) = .168$ and $P_b(\text{err} \mid A) = .400$, which are pretty far off the obtained values. If we now move to the predictions for condition B and continue to use $\hat{u} = .06$, as we must with this model, we estimate $\hat{p}_B = -.581$, which is ridiculous, and using the least acceptable value of $\hat{p}_B = 0$ gives predictions very far from what we want.

Perusing the structure of Eqs. 5.9–5.12 again, we discover that there was a fairly strong and easily tested nonparametric prediction from this model, namely that Eq. 5.9 + Eq. 5.10 = Eq. 5.11 + Eq. 5.12. That this relation does not hold empirically offers further evidence against model I.

Compound serial model II has 4 parameters and assumes that the rate of processing u, when the target is in position a, can be different from when the target is in position b. There does not seem to be a really good rationale for why this should be true, but we will see if the model works. Equations 5.13–5.16 show the predictions.

Compound serial model II

	Theoretical	Data	
Condition A			
	$P_a(\text{err} \mid A) = p_A e^{-u_1 t}\cdot\frac{2}{3} + (1-p_A)[e^{-u_1 t} + u_1 te^{-u_1 t}]\cdot\frac{2}{3}$	.155	(5.13)
	$P_b(\text{err} \mid A) = p_A[e^{-u_2 t} + u_2 te^{-u_2 t}]\cdot\frac{2}{3} + (1-p_A)e^{-u_2 t}\cdot\frac{2}{3}$	.413	(5.14)
Condition B			
	$P_a(\text{err} \mid B) = p_B e^{-u_1 t}\cdot\frac{2}{3} + (1-p_B)[e^{-u_1 t} + u_1 te^{-u_1 t}]\cdot\frac{2}{3}$	.265	(5.15)
	$P_b(\text{err} \mid B) = p_B[e^{-u_2 t} + u_2 te^{-u_2 t}]\cdot\frac{2}{3} + (1-p_B)e^{-u_2 t}\cdot\frac{2}{3}$	.150	(5.16)

It is more difficult to pull apart the parameters in model II than it was for the first model, so we employed a χ^2 fit and test routine. The estimated parameters were $\hat{p}_A=.818$, $\hat{p}_B=0$, $\hat{u}_1=.082$, $\hat{u}_2=.057$, and the predicted values were $P_a(\text{err}\mid A)=.136$, $P_b(\text{err}\mid A)=.371$, $P_a(\text{err}\mid B)=.293$, $P_b(\text{err}\mid B)=.179$. While following the pattern of results, the fit leaves much to be desired, considering that all degrees of freedom in the data were used up.

Model III assumes that the rates are distinct for the two separate conditions, and we will call these u_A and u_B. The predictive equations are as follows.

Compound serial model III

	Theoretical	Data
Condition A		
	$P_a(\text{err}\mid A)=p_A e^{-u_A t}\cdot\frac{2}{3}+(1-p_A)[e^{-u_A t}+u_A te^{-u_A t}]\cdot\frac{2}{3}$	.155
	$P_b(\text{err}\mid A)=p_A[e^{-u_A t}+u_A te^{-u_A t}]\cdot\frac{2}{3}+(1-p_A)e^{-u_A t}\cdot\frac{2}{3}$	.413
Condition B		
	$P_a(\text{err}\mid B)=p_B e^{-u_B t}\cdot\frac{2}{3}+(1-p_B)[e^{-u_B t}+u_B te^{-u_B t}]\cdot\frac{2}{3}$	.265
	$P_b(\text{err}\mid B)=p_B[e^{-u_B t}+u_B te^{-u_B t}]\cdot\frac{2}{3}+(1-p_B)e^{-u_B t}\cdot\frac{2}{3}$	.150

Here we can employ the same values for condition A as for model I since the present equations are identical to those. However, we call that rate u_A, since the rates now are supposed to be different in the two conditions. We then clearly can estimate completely separate values of p_B and u_B. Here matters are worse than in the case of condition A because, although $\hat{u}_B$ turns out to be plausible (.079), our estimate of p_B is $\hat{p}_B=-.209$, and trying out $\hat{p}_B=0$ gives a quite bad prediction (.109 as opposed to .150).

The last compound serial model we will look at (model IV) is in some ways the most interesting, but also more intractable than the others. It says that the rates differ according to the order taken through the two positions, u for order $\langle a,b\rangle$ and u' for order $\langle b,a\rangle$. The expressions follow immediately.

Compound serial model IV

	Theoretical	Data	
Condition A			
	$P_a(\text{err}\mid A)=p_A e^{-ut}\cdot\frac{2}{3}+(1-p_A)[e^{-u't}+u'te^{-u't}]\cdot\frac{2}{3}$	.155	(5.17)
	$P_b(\text{err}\mid A)=p_A[e^{-ut}+ute^{-ut}]\cdot\frac{2}{3}+(1-p_A)e^{-u't}\cdot\frac{2}{3}$	.413	(5.18)
Condition B			
	$P_a(\text{err}\mid B)=p_B e^{-ut}\cdot\frac{2}{3}+(1-p_B)[e^{-u't}+u'te^{-u't}]\cdot\frac{2}{3}$	.265	(5.19)
	$P_b(\text{err}\mid B)=p_B[e^{-ut}+ute^{-ut}]\cdot\frac{2}{3}+(1-p_B)e^{-u't}\cdot\frac{2}{3}$	.150	(5.20)

Note that both of the rate parameters appear in each of the four expressions. This at least suggests the possibility that this model might be better able to handle the present data through this more complex patterning of the model parameters. However, close scrutiny of Eqs. 5.17–5.20 provides some doubt. The doubt springs from the large empirical value corresponding to Eq. 5.18 relative to the others and the near-equality of Eq. 5.17 and Eq. 5.20. In any case, a chi-square fit yielded predictions of .166, .396, .265, and .150 corresponding to Eqs. 5.17–5.20, respectively. Note that the last two predictions were right on the nose, but the first two deviate somewhat from the observed values. The parameters were $\hat{p}_A=1.0$, $\hat{p}_B=.174$, $\hat{u}=.061$, $\hat{u}'=.083$, $\hat{p}_A$ and the rates being close to our earlier, more crudely obtained estimates. The final chi-square cannot be rationally evaluated since the number of parameters is equal to the number of degrees of freedom in the data. This model could be describing some of the processing behavior of the observers.

Interestingly enough, when the time parameter t was also allowed to vary in the data fit routine on model IV, it turned out to bear almost the same relation to the parameters as when it was set at $t=23$ msec. That is, the estimated value of t was 2.04 time units and the estimates of u and u' were larger than the previous ones (in the fit with $t=23$) by a factor of 10. The predictions were exactly the same as before. Thus, although one is here unable to completely isolate t independently of the u terms, within the context of this model we may have identified the *relationship* between the processing time and the processing rates.

Now that we have finished a number of compound serial models, how about trying a compound parallel model? The most natural one to work with is probably the independent parallel model, where the rates may be distinct across positions but do not change on a given position across stages, that is, $v_{a1}=v_{a2}$, for example. On the other hand, we must assume that the two parallel systems making up the compound parallel system have different rates; otherwise the system would degenerate to a regular parallel system.

In the serial cases, it was supposed that the two separate reinforcement conditions A and B were each associated with one of the two serial models (one having probability p_A, the other p_B). Of course, the most general possible models could not be used because they would have 10 parameters, 6 more than the number of degrees of freedom in the data. In the parallel case we also assume that condition A is associated with one independent parallel system and condition B with the other. Thus, we again have observability of the two systems that we would not have, for instance, if the reinforcement conditions were not stated for a given trial until after the trial was over.

The compound parallel model used here assumes that the observer distributes attention differentially in the two conditions and that the attention given to position a will also typically differ from that given to position b. The predictive equations are as follows.

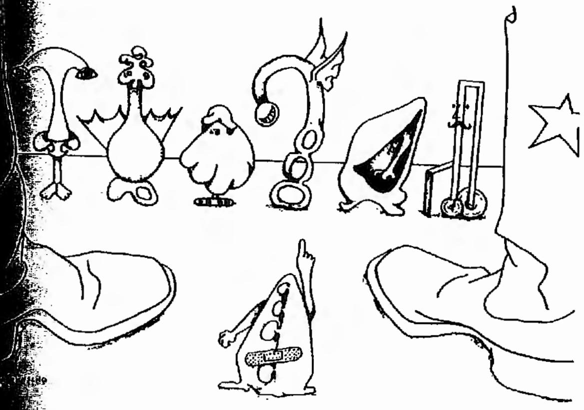
Compound parallel model I

	Theoretical	Data	
Condition A			
	$P_a(\text{err}\,\vert\,A)=\exp(-v_{aA}t)\cdot\frac{2}{3}$	.155	(5.21)
	$P_b(\text{err}\,\vert\,A)=\exp(-v_{bA}t)\cdot\frac{2}{3}$	.413	(5.22)
Condition B			
	$P_a(\text{err}\,\vert\,B)=\exp(-v_{aB}t)\cdot\frac{2}{3}$	.265	(5.23)
	$P_b(\text{err}\,\vert\,B)=\exp(-v_{bB}t)\cdot\frac{2}{3}$	.150	(5.24)

The parameters are just the rates on the two positions for the two conditions; thus, v_{aA} is the processing speed on position a in condition A. It is immediately clear that this model will fit perfectly because the parameters are distinct in the four above equations. The parameter estimates are $\hat{v}_{aA}=.063$, $\hat{v}_{bA}=.021$, $\hat{v}_{aB}=.040$, and $\hat{v}_{bB}=.065$. This model has, of course, a compound serial equivalent, and the reader is invited to derive it, using the mappings stated earlier. In the present instance, this is quite easy, because we need only give the regular serial model equivalent to the four parallel models corresponding to Eqs. 5.21–5.24. Although this independent compound parallel model and its serial equivalent are not testable with these data, it is instructive that a very natural parallel model finds it so easy to handle a situation in which several fairly intuitive and reasonably sophisticated serial models evidenced considerable difficulty.

Shaw and LaBerge's serial notions were expressed in terms of "preservation operations," but it could well be that such functions act in parallel. The major objections brought forth by those authors against parallel processing related to parallel rate differences (in our language) where one would not expect them on the basis of retinal locus considerations. However, as the reader may readily demonstrate if he or she has not already, it is easy to be looking at one object (i.e., fixating) while placing one's processing attention on one or more other objects, with just about whatever asymmetry is desired. Thus one can look at the doorknob but "focus" one's visual attention on the window to the right and become almost effectively blind to the doorknob. The window may not be so clearly seen as if its image were in the fovea, but rough characteristics can still be discerned. Similarly, two objects may be equally far from the fovea yet one can be attended to while the other is virtually ignored. It is a separate question as to whether the two may be simultaneously attended to as well as one (the capacity issue). It may also be difficult to completely shut out the information from one of the stimuli, but it seems certain that one can control attention without gross retinal differences in images.

We mention in closing that equivalence and diversity theorems concerning compound parallel vs. serial models have not yet been worked out for other than those based on exponential intercompletion time distributions.



In a *display search* experiment, a visual array of items is shown to an observer who must determine whether a previously designated target is present in the display. Here a Turgle who suffered the loss of one of its buds (which give off seeds for procreative purposes) correctly affirms the presence of the thief (one of the Ringed Angel tribe that preys on Turgle buds) in the lineup. The officer-in-charge stands protectively at hand.

6 Memory and visual search theory

Probably the most pervasive application of the stochastic search models introduced in Chapter 4 has been to the study of human memory scanning and visual search. Broadly defined, these areas are interested in the processes involved in the retrieval of information from man's transient memories. Memory-scanning experiments have traditionally been concerned with search through what has come to be known as *short-term memory*, whereas visual search tasks are thought to involve a more modality-specific (i.e., visual) iconic store; however, as we shall see, even these seemingly straightforward interpretations are not without their controversy.

In a typical search experiment, a list of stimulus elements or items is loaded into the transient memory of interest and the observer is asked to search through the list as quickly as possible for the presence or absence of some critical item. Usually the observer indicates his or her decision by pressing a button indicating either "yes, the critical item is contained in the list" or "no, the item is not in the list."

The difference between memory-scanning and visual search tasks is operationally a difference of the temporal ordering of stimuli. In memory-scanning tasks the stimulus list is shown to the observer and then some short time later the critical item – or *target* item, as it is often called – is displayed. As soon as the target item is presented to the observer, he or she can begin scrutinizing the now-internal search list for the target's presence and can respond "yes" or "no" as soon as the target's membership or nonmembership in the search set is verified. Memory-scanning tasks, introduced by Sternberg in 1966, are sometimes called LT or *late target* tasks (Townsend & Roos 1973) since the

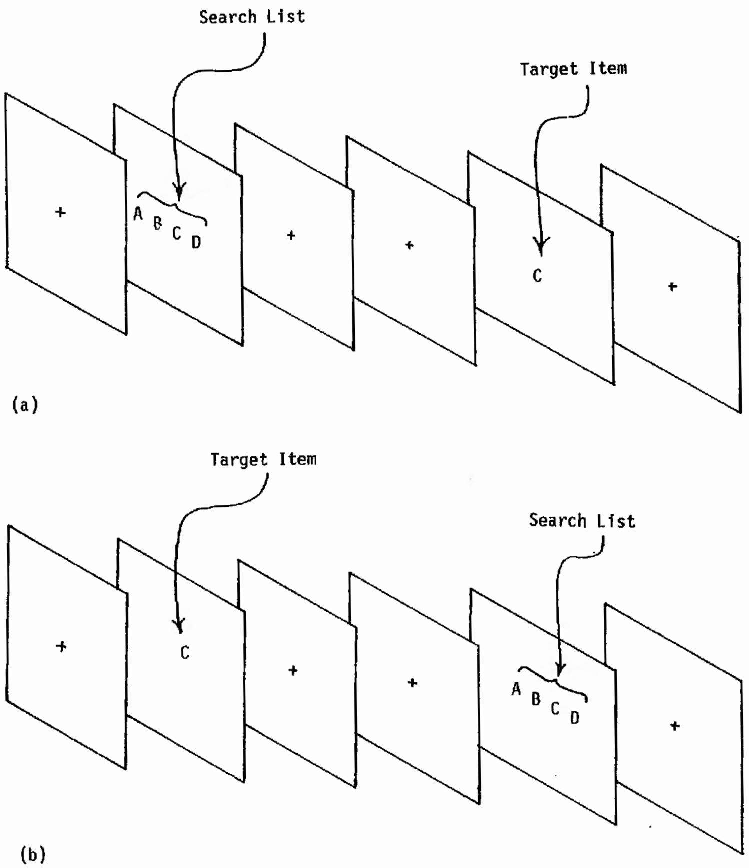


Fig. 6.1. Schematic representing a series of hypothetical events occurring on one particular trial of a memory scanning or LT task (a) and on a trial of a visual search or ET task (b).

target is presented *after* the search list (often called the memory list in LT paradigms). A schematic representing the events occurring on a hypothetical trial in a memory-scanning experiment is shown in Fig. 6.1a.

The events in a visual search or ET (for *early target*) task happen in the reverse chronological order. First the target item is shown to the observer and then this is followed somewhat later by the search list. The measurement of response time begins with the onset of the search list and ends when the observer depresses one of the response keys. A schematic of a hypothetical trial of this type is given in Figure 6.1b. The visual search paradigm, which has its roots in the "detection paradigm" formulated by Estes and Taylor (1964), was first introduced in its present form by Atkinson et al. (1969).

In both the ET and LT paradigms, response time is the dependent variable of primary interest. The observer is typically instructed to respond as quickly

as possible without making any errors. For this to be possible the task must be easy enough for the observer to perform perfectly under conditions of free response. For this reason, stimulus legibility is typically quite good and care is usually taken to ensure that the presentations are foveal. This factor tends to discourage the use of large search lists, which have the additional disadvantage of overloading the memory store and in this way reducing accuracy. In addition, all exposure durations are relatively long, generally in the range of 150 to 250 msec. Longer durations are not used only because they allow the possibility of saccadic eye movements. Under these conditions human observers can usually maintain their error rates below 5%.

We mentioned above that it is believed that the perceptual comparisons involved in ET and LT experiments take place in different memory stores. On the basis of capacity effects and learning curve properties for ET and LT designs, Townsend and Roos (1973) argued that the visual search or ET comparisons may take place in a visual "form system," whereas the memory scanning or LT comparisons may occur in an acoustic "form system" (see Fig. 11.1). The visual form system is assumed to be an icon or short-term store in which the stored representations are visual in nature and it is thought to play an important role in ET tasks because in this paradigm the response processes begin with the presentation of the search list. A visual icon is thought to be the earliest store for visually presented stimuli (Sperling 1960; Neisser 1967; Sakitt 1976); thus if comparisons could take place at this level, response time should be minimized.

In LT paradigms the search list is presented before the response processes begin and so there is plenty of time for the internal representations to move to other memory stores. A short-term memory in which the stored representations are of an acoustic nature is a good candidate because it is thought that observers often internally rehearse the search list before the target is presented.

Gilford and Juola (1976), however, reported a study where the orthographic factors of familiarity and meaningfulness had similar effects on the forms of the ET and LT response time functions, and from these results they argued that the comparisons occur in the same memory store. They did not specify the nature (e.g., modality code) of the comparison location or mechanism. As we mentioned briefly, one of the arguments against both types of comparison taking place in the same memory system is that it takes observers quite a bit more practice to perform a memory-scanning task without error than it does a visual search task. Were the same memory stores used, we would expect both designs to be characterized by the same learning rates.

Moreover, we should note that the Gilford and Juola (1976) findings could be explained within the context of separate comparison locations if the effects of varying orthographic features have similar influences on an acoustic form system and a visual form system. Certainly there must exist concomitant structure for the constraints of orthographic factors in both the visual and verbal-acoustic systems. We propose that experiments manipulat-

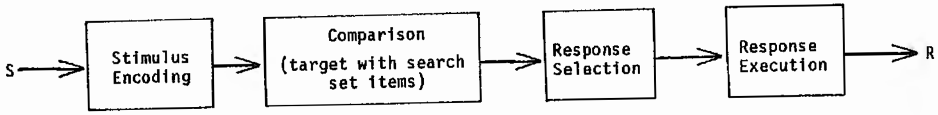


Fig. 6.2. Schematic of a popular discrete stage model of the internal events thought to occur during memory scanning and visual search tasks.

ing the acoustic and visual similarity of targets to the nontargets may be helpful in deciding the locus of comparisons in the two tasks. For example, high visual similarity should greatly increase RT in the ET or visual search design but not in the LT or memory-scanning design if processing occurs in a visual form system in the former but occurs in an acoustic form system in the latter.

Wherever it takes place, the comparison process has been assumed to be the primary locus of the RT differences that are observed when the length of the search list is varied. Other widely accepted processing subsystems whose durations are usually thought *not* to be affected by search set size are stimulus encoding, response selection, and response execution. Together they comprise a popular model of the response processes involved in ET and LT tasks (Smith 1968; Sternberg 1969a, b). A schematic of the model is given in Fig. 6.2.

The stimulus-encoding process is envisioned as the early subsystem (in a visual form system) in which the internal neural code representation of the stimulus is constructed. In memory-scanning experiments this is the target item, whereas in visual search tasks this is the search list itself. Stimulus encoding has been thought to be of unlimited capacity in the sense that increasing the number of items to be encoded presumably does not affect total encoding time (e.g., Shiffrin and Gardner 1972; Gardner 1973). If this assumption is correct, then increasing the search set size in ET tasks should not increase encoding time, at least when all display items can be viewed within a single fixation. The duration of this processing stage is assumed to be affected, however, by experimental factors such as stimulus intensity and perhaps stimulus probability (Sternberg 1969a; Miller & Pachella 1976). Evidence that the visual form system is limited capacity will be considered in Chapter 11.

We will be most interested in the comparison process in this chapter, whether it takes place in, say, the visual form system or the acoustic form system (or elsewhere).

It is usually assumed that comparison and encoding, along with the other RT subprocesses, are discrete and nonoverlapping subsystems (i.e., Donderian or strictly serial subsystems). The effects of dropping this assumption are only now beginning to be studied. We investigate this issue further in Chapter 12.

The final "cognitive" process is response selection. Once comparison is completed and the observer knows whether any of the search set items match

the target, the response selection process determines the appropriate response. A somewhat anthromorphic example might be "... a match occurred, so the right-hand button should be pressed." Reaction time and assumedly response selection time are well known to increase strongly with the number of response alternatives (Hick 1952; Berlyne 1957), although this factor is almost always confounded with stimulus ensemble size. In this respect it might be considered an output analogue to the comparison process since comparison time is thought to increase sharply with the number of stimulus alternatives. In response selection, the observer might be thought to be searching through a list of response alternatives for one that is appropriate to, or that matches, the output of the comparison process.

Since we have (somewhat arbitrarily) decided that we are more interested in details of the comparison process than in response selection, we must take care that response uncertainty does not covary with stimulus uncertainty. This is cleverly achieved in both ET and LT paradigms. Response uncertainty is always one bit, no matter what the search set size, since the only appropriate responses are always "yes" and "no" or perhaps "present" and "absent." The development of such a paradigm thus represents a significant improvement over, say, stimulus identification tasks in which stimulus and response uncertainty are equal.

Little is to be said of the last processing subsystem, response execution, except that if one is interested primarily in the other cognitive processes, then a response should be chosen that minimizes response execution time variability. This goal is achieved by keeping movement distances small if buttons or levers are used, and by favoring fingers as response agents over the larger arms and legs. Also, to reduce execution error, response agents should be separated spatially on the body. For example, fewer errors and faster RT's will probably be made in two choice tasks when fingers from different hands are used to depress response keys rather than adjoining fingers on the same hand (e.g., Shulman & McConkie 1973). Given these criteria, the choice of having observers press a button with a finger of either hand to signify "yes" or "no," as is done in most ET and LT studies, seems a good one.

We now consider typical ET and LT results and the popular model they were first thought to support. We have already outlined a very general model, but as it stands details of the comparison process are missing. It is these details that we will be most concerned with throughout the rest of this chapter.

One of the first published studies utilizing either of the experimental designs we have been discussing was a memory-scanning or LT experiment performed by Sternberg in 1966. Using letters of the English alphabet as stimuli, he varied search set size from one to six items and obtained the now-classical results associated with both experimental paradigms. Mean RT was found to increase linearly with search (memory) set size and at the same rate of about 38 msec per search set item for both "yes" and "no" or target-present and target-absent trials. This result has been replicated many, many times in both ET and LT paradigms (Atkinson et al. 1969; Burrows & Okada 1971, 1973;

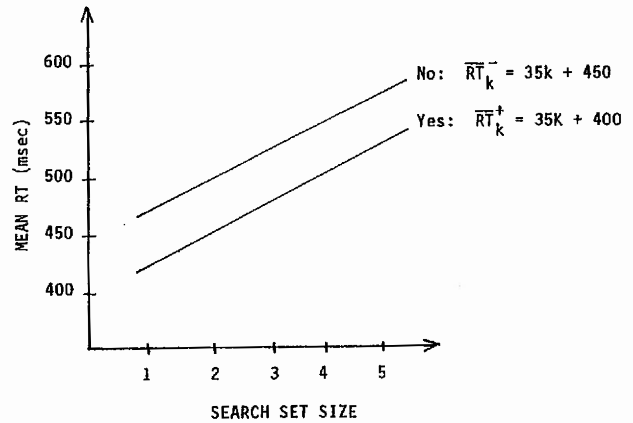


Fig. 6.3. Prototypical target-present and target-absent mean RT results obtained in standard memory scanning and visual search tasks.

Chase & Calfee 1969; Theios, Smith, Haviland, Traupmann, & Moy 1973; Townsend & Roos 1973; see Sternberg 1975 for a much more extensive list of replications and exceptions in the LT task).

An idealized example of results typically found in these paradigms is given in Fig. 6.3. On the basis of results such as these, Sternberg (1966) proposed that the comparison process is mediated by a serial exhaustive search in which the individual item-processing times of all items, targets and nontargets, is the same. This very popular model came to be known as the *standard serial exhaustive search model*. It predicts the results of Fig. 6.3 very nicely and with only three parameters.

First, the general formula will be derived. Let $\mathbf{T}^+_{i,n}$ be the random individual-item processing time of the target item in (serial) position i of the search set when search set size is n, and let $\mathbf{T}^-_{i,n}$ be the corresponding random time if the item is a nontarget. Then the expected RT on target absent trials for a general serial exhaustive model when search set size is n is

$$E(\mathbf{RT}^-_n) = \sum_{i=1}^{n} E(\mathbf{T}^-_{i,n}) + t_- \tag{6.1}$$

where t_- is the average time of all RT processes (the so-called *residual processes duration*), other than comparison, on target-absent trials. This time is often called the *base time* and is usually assumed to include stimulus encoding (if it precedes comparison), response selection, and response execution.

The expected target-present RT is slightly more complicated because we must take into account the different serial positions in which the target can appear,

$$E(\mathbf{RT}^+_n) = \frac{1}{n}\sum_{j=1}^{n}\left\{\left[\sum_{\substack{i=1\\ i\neq j}}^{n} E(\mathbf{T}^-_{i,n})\right] + E(\mathbf{T}^+_{j,n})\right\} + t_+ \tag{6.2}$$

In the standard serial exhaustive model these expressions are greatly simplified since there is only one processing rate to consider.

Proposition 6.1: In the standard serial exhaustive model $E(\mathbf{RT}^-_n) = E(\mathbf{T})n + t_-$ and $E(\mathbf{RT}^+_n) = E(\mathbf{T})n + t_+$.

Proof: In the standard serial exhaustive model

$$\mathbf{T}^-_{i,n} = \mathbf{T}^-_{j,n} = \mathbf{T}^+_{i,n} = \mathbf{T}^+_{j,n} = \mathbf{T} \quad \text{for all } i,j \leqslant n$$

and for all values of n. Thus, Eq. 6.1 becomes

$$E(\mathbf{RT}^-_n) = \sum_{i=1}^{n} E(\mathbf{T}) + t_- = E(\mathbf{T})n + t_-$$

Similarly, Eq. 6.2 becomes

$$E(\mathbf{RT}^+_n) = \frac{1}{n}\sum_{j=1}^{n}\left\{\left[\sum_{\substack{i=1\\ i\neq j}}^{n} E(\mathbf{T})\right] + E(\mathbf{T})\right\} + t_+$$

$$= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} E(\mathbf{T}) + t_+ = \frac{1}{n}\sum_{j=1}^{n} nE(\mathbf{T}) + t_+$$

$$= \sum_{j=1}^{n} E(\mathbf{T}) + t_+ = E(\mathbf{T})n + t_+ \quad \square$$

Thus the standard serial exhaustive model predicts the target-present and target-absent curves to be linear functions of n with the same slope $E(\mathbf{T})$. If we now set

$$E(\mathbf{T}) = 35 \text{ msec}, \quad t_- = 450 \text{ msec}, \quad \text{and} \quad t_+ = 400 \text{ msec}$$

then

$$E(\mathbf{RT}^-_n) = 35n + 450 \quad \text{and} \quad E(\mathbf{RT}^+_n) = 35n + 400$$

which are exactly the RT functions displayed in Fig. 6.3.

Sternberg based his arguments for seriality on the linearity of the observed mean RT functions. Since adding one item to the search set always caused a constant increase in mean RT, it was as if adding an item to the search set added a discrete stage of constant mean duration to the RT process, just as would be expected of a serial system. His argument that search is exhaustive was based on the parallelity of the target-present and target-absent functions. Adding a nontarget item to the search set has the same effect on target-present trials as it does on target-absent trials. This is to be expected of an exhaustive process. If search is self-terminating, though, on the average about half the time the newly added item will be processed before the target and about half the time it will not yet be processed by the time the target is completed and search is terminated. This means that with a self-terminating search, on target-present trials the increase in mean RT caused by adding an item to the search set should be only half of what it is on target-absent trials. The target-present curve should then reflect this fact by having half the slope of the target-absent curve. Since, empirically, target-present and absent curves were found to have the same slope, Sternberg decided in favor of an exhaustive search strategy.

These arguments are intuitive and compelling, and if these were all the data available to us, the story might end here. But since 1966, several factors have arisen that cloud the picture. First it was learned that many other search models can also predict the ubiquitous straight-lined, equal-sloped target-present and absent curves, although only a few possess the economy of parametric structure associated with the standard serial exhaustive model. Second and equally necessary to the continuation of our story, empirical results were discovered that, as it stood, the model could not predict.

Problems with the standard serial exhaustive search model

Although the standard serial exhaustive search model easily and naturally predicts linear and parallel target-present and absent curves, there exist many other findings that the model has great difficulty accommodating. We will examine four of these in this section.

Serial position effects

In typical memory-scanning and visual search experiments, the target item appears randomly among the stimulus positions in the search list on target-present trials. Thus, in addition to computing mean RT conditioned on the search set size, as we have done above, we could also look at mean RT for target-present trials when the display size is *n and* the target appears in a particular position of the display, say the kth.

Mean RT curves that are conditioned on the serial position of the target within the display are, naturally enough, called serial position curves. A serial position curve can be constructed for each display size of the experiment. Often, mean RT is found to vary with the position of the target in the display, so that the resulting serial position curves are not flat functions (Morin, DeRosa, & Stultz 1967; Atkinson et al. 1969; Clifton & Birenbaum 1970; Townsend & Roos 1973). Any deviation from a flat function is referred to as a serial position effect. In memory-scanning experiments serial position effects are especially likely to occur if the interval between presentation of the memory set and the target item is brief (Sternberg 1975).

It is easy to see that the standard serial exhaustive model cannot predict any serial position effects. If all items are processed at the same rate, then the total processing time of all displays with an equal number of items should be the same; that is, serial position effects should not be manifested. One way to circumvent this difficulty is to allow individual item-processing rates to depend on the serial position of the item (Townsend 1974b). A possible rationale for the differential rates on different positions follows, for example, from

serial positions. This possibility ties in nicely with traditional memory experiments and theorizing and is consonant with assumptions made by, say, trace strength models.

Although such an exhaustive, different rates model could be given either a parallel or serial interpretation, to facilitate comparison with the standard serial exhaustive model let us see how a serial interpretation acts in the present context. First, because the model is serial and exhaustive we can begin with Eqs. 6.1 and 6.2. Now let us assume that all nontargets have the same exponential processing rate u, so that $E(\mathbf{T}_{i,n}^{-})=1/u$, for all values of i and n, but that target items, although still exponential, may vary over serial position and search set size. Thus, $E(\mathbf{T}_{i,n}^{+})=1/u_{i,n}^{+}$. Under these processing rate assumptions Eqs. 6.1 and 6.2 reduce to

$$E(\mathbf{RT}_n^-)=\frac{n}{u}+t_- \quad \text{and} \quad E(\mathbf{RT}_n^+)=\frac{n-1}{u}+\frac{1}{n}\sum_{j=1}^{n}\frac{1}{u_{j,n}^{+}}+t_+$$

Clearly, serial position effects are represented by the $1/u_{j,n}^{+}$. These two curves will be linear functions of n with equal slopes if and only if

$$E(\mathbf{RT}_n^-)-E(\mathbf{RT}_n^+)=c=\text{constant}$$

or equivalently,

$$\frac{n}{u}=\frac{n-1}{u}+\frac{1}{n}\sum_{j=1}^{n}\frac{1}{u_{j,n}^{+}}+c+t_+-t_-$$

or equivalently whenever

$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{u_{j,n}^{+}}=\frac{1}{u}+t_--t_+-c$$

A sufficient condition is then that

$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{u_{j,n}^{+}}=\frac{1}{u}$$

That is, the average time to process a positive comparison would be equal to the time to process a negative comparison. Put another way, the harmonic mean of the positive rates is equal, for all n, to the negative comparison rate.

In theory, virtually any type of serial position effects can be predicted by this model. Pragmatically, though, a somewhat different picture emerges. The constraint that the requirement of equal slopes imposes on the target rates is strong enough to make prediction of certain types of serial position effects unparsimonious. For instance, we shall see in the next chapter that to predict either monotonically increasing or decreasing serial position curves requires $n(n+1)/2$ target rate parameters in addition to the nontarget parameter. Unless one has a way to fix these parameters a priori, this completely exhausts the degrees of freedom generated by the serial position curves and thus renders the model of little practical value as a predictive device.

To get a rough idea of the difficulties the model faces, note that the obvious recency effect) is to postulate that target items are processed faster the farther to the right they appear in the display (for one reason or another) but that nontargets are always processed at the same rate. This might be the easiest way for a serial exhaustive model to predict left-to-right decreasing serial position curves, but unfortunately the cost is great. The model no longer predicts parallel target-present and absent curves. This is because with larger set sizes, faster and faster target-processing times are averaged into the mean target-present processing times. This causes the target-present curve to increase more slowly than the target-absent curve, thus violating parallelity. As we shall see in the next chapter, the only way to circumvent this problem is to increase the number of parameters to an unreasonable degree.

The effect of redundant targets

A second robust result that serial exhaustive models have trouble predicting is that target-present mean RT generally decreases as more replicas of the target item are added to the stimulus display (Bjork & Estes 1971; Baddeley & Ecob 1973; van der Heijden & Menckenberg 1974). The natural prediction of a standard serial exhaustive model is, of course, that the target-present mean RT with n items in the display should always be the same no matter how many of the items are duplicates of the target. We can, however, modify the model in a fairly simple fashion so that it can make this prediction. We merely need to assume that target items are processed more quickly than nontarget items so that the more targets in the display the faster the processing time. This scheme only adds one (rate) parameter to the model and does not disturb the prediction of linear and parallel target-present and absent curves. Some of the intercept difference between the two functions is now accounted for within the comparison stage - specifically, the mean processing time difference between targets and nontargets. It should be noted, though, that this modification of the model still does not allow serial position effects to be predicted.

Reaction time variances

A third finding that causes the model difficulty is that RT variances tend to increase with display size more quickly for target-present curves than for target-absent curves (Schneider & Shiffrin 1977). Contrarily, the prediction by the standard serial exhaustive model is that RT variance should increase at the same rate in both cases. On the other hand, self-terminating models do easily predict the observed pattern of results. Since we consider this prediction in detail in the next chapter, we offer only an intuitive rationale for the phenomenon here. Basically, the interaction occurs because two factors contribute to the target-present variance when search is self-terminating, whereas only one factor contributes to the target-absent variance.

First, in a stochastic system, there is always some variance associated with the individual element processing times, and this source of variation will naturally enough tend to increase as the processing load (i.e., n) increases. Second, on target-present trials, different numbers of elements will be processed on different trials even when there are the same number of elements in the search set. This will tend to increase the variance, and in fact the effect will be larger for larger n since then the number of items actually processed on a given trial can vary over a greater range. This causes the target-present RT variance to grow with stimulus set size faster than the target-absent RT variance.

The fact that serial self-terminating models predict large RT variances was known as early as 1903 when Hylan introduced arguments to the effect that serial self-terminating processing should be associated with larger variances than "parallel" (presumably exhaustive) processing.

Display configuration

The final evidence we will consider that is damaging to serial exhaustive models is the effect of display configuration. The results discussed above all utilized fairly standard linear arrays; however, there are several reasons why linear arrays might not always be the best to use. Principal among these is the fact that adding items to the stimulus array tends to increase both lateral interference and the visual angle subtended by the array. Either of these two problems could conceivably cause, say, stimulus encoding time to covary with stimulus array size and thus to violate our assumption that search set size affects only the comparison process. Further, reading habits, which one may wish to neutralize, may become a factor. Circular stimulus arrays tend to minimize all of these problems, and several researchers have opted for this configuration (Egeth, Jonides, & Wall 1972; Gardner 1973; van der Heijden & Menckenberg 1974).

Typically it is found that mean RT barely increases or even remains constant with increases in circular display size (Egeth et al. 1972), evidence usually taken as more indicative of a parallel scan than a serial search. Serial models, however, can, with great difficulty, predict flat RT functions if they postulate a tremendous increase in individual item processing rate with search set size or if they postulate some change in the base time with increases in display size. Even so, it is not clear why either the individual-item processing time or the base time should decrease so dramatically with search set size for circular arrays but not for linear ones. On the other hand, one could make a similar, if not quite as strong, argument against parallel models. The flat RT functions, often found with circular arrays, suggest a supercapacity or a deterministic unlimited capacity model, while results from linear arrays suggest a fairly substantial capacity limitation. Why should there be such a difference? It is possible that overcoming increased lateral interference and the increased average distance to the foveal center associated with items in linear arrays draws off some of the capacity allocated to comparison. This then would result in a limited capacity comparison process when linear displays are used even though comparison might be unlimited capacity with circular displays. Another possibility, inelegant perhaps, is that search is serial with linear arrays but parallel with circular ones.

Objections to other models

Although serial exhaustive models have a very difficult time predicting certain empirical phenomena, other natural candidates that might be called on as replacements have their own difficulties. It is our contention, however, that much of the criticism targeted at self-terminating models and at parallel models has been unduly harsh and the resulting conclusions somewhat hasty. We examine these issues now in more detail.

Self-terminating search

Equal-sloped target-present and target-absent curves

The classical objection to self-terminating search strategies is, of course, the parallel target-present and target-absent curves so routinely found in memory-scanning and visual search studies. We already indicated that, in general, self-terminating models have little difficulty making this prediction, and for this reason comparing target-present and target-absent slopes is not a good way to decide between self-terminating and exhaustive strategies. For instance, consider an independent parallel self-terminating model. On target-present trials the nontarget-processing times will have no effect on the RT, and so if the experimenter places a target in each position with probability $1/n$, then the expected target-present RT when there are n items will be

$$E(\mathbf{RT}_n^+) = \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{T}_{i,n}^+) + t_+ = E(\overline{\mathbf{T}}_n^+) + t_+ \tag{6.3}$$

where t_+ is the expected base time, $E(\mathbf{T}_{i,n}^+)$ is the expected processing time of the target item when it is in position i, and $E(\overline{\mathbf{T}}_n^+)$ is the average of these expected processing times over the n serial positions. It is now easily seen that any linear relationship between the $E(\overline{\mathbf{T}}_n^+)$ for different values of n imparts a linearity to the target-present mean RT curve [i.e., the $E(\mathbf{RT}_n^+)$ vs. n curve]. Further, we have as yet placed no constraints on any of the nontarget-processing times, so these are free to be chosen in a way such that the target-absent curve is also linear with positive slope. Since the two curves are determined by a mutually exclusive set of parameters, it is reasonable that their slopes can be related in any arbitrary manner.

To see what such a model might look like, note that the expected RT on target-absent trials is

$$E(\mathbf{RT}_n^-) = E(\max_{i \leq n} \mathbf{T}_{i,n}^-) + t_- \tag{6.4}$$

To simplify the model let us assume that the nontarget-processing rates are the same for all serial positions and therefore depend only on the search set size n. We also assume, as before, that individual item comparison times are exponentially distributed.

Proposition 6.2: A parallel self-terminating model with exponentially distributed processing times predicts target-present and target-absent curves will be linear functions of n with the same slope D if the nontarget rate when the load is n elements is given by

$$v_n^- = \frac{1}{Dn}\sum_{i=1}^{n}\frac{1}{i} \doteq \frac{\log n}{Dn}$$

and if the harmonic mean of the target rates equals $1/Dn$, that is, if

$$\frac{n}{\sum_{i=1}^{n}(1/v_{i,n}^+)} = \frac{1}{Dn}$$

Proof: Letting $E(\mathbf{T}_{i,n}^+) = 1/v_{i,n}^+$ in Eq. 6.3 and assuming the $\mathbf{T}_{i,n}^-$ in Eq. 6.4 are all independent with identical exponential distributions (with rate v_n^-) reduces the mean RT predictions to

$$E(\mathbf{RT}_n^+) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{v_{i,n}^+} + t_+$$

and by Eq. 3.23

$$E(\mathbf{RT}_n^-) = \left[\frac{1}{nv_n^-} + \frac{1}{(n-1)v_n^-} + \cdots + \frac{1}{v_n^-}\right] + t_- = \frac{1}{v_n^-}\sum_{i=1}^{n}\frac{1}{i} + t_-$$

To obtain linear functions of n with slope D it must be the case that

$$\frac{1}{v_n^-}\sum_{i=1}^{n}\frac{1}{i} = Dn$$

and hence that

$$v_n^- = \frac{1}{Dn}\sum_{i=1}^{n}\frac{1}{i} \doteq \frac{\log n}{Dn}$$

This will achieve the desired effect for the target-absent RTs. Similarly, for the target-present curve it must be true that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{v_{i,n}^+} = Dn \quad \text{or} \quad \frac{n}{\sum_{i=1}^{n}(1/v_{i,n}^+)} = \frac{1}{Dn} \quad \square$$

Note that in this model, capacity is even more limited on positive comparisons than it is on negative since the harmonic mean of the target rates decreases proportionally to n and therefore *faster* than the nontarget rates.

In any case, these parameter mappings yield target-present and absent mean RT curves that are both linear with slope D. That is,

$$E(\mathbf{RT}_n^-) = Dn + t_- \quad \text{and} \quad E(\mathbf{RT}_n^+) = Dn + t_+$$

We will learn that other self-terminating models can predict linear and parallel target-present and absent curves. It is generally true, however, that these models are not the simplest in their class and are certainly not simpler than their exhaustive counterparts. Even so, we are unable to invoke Occam's razor here because of the difficulties with serial exhaustive models that we discussed in the preceding section. For this reason, we are forced to look at more complex models, be they self-terminating or exhaustive.

Increases in minimum reaction time

A second finding that has been used to discredit self-terminating models is that the minimum RT increases with the size of the search set (Sternberg 1975; Lively 1972; Lively & Sanford 1972). The rationale behind this argument is quite simple. If search is self-terminating, then no matter how large the search set, there should always be some nonzero probability that the target is the first item processed and thus the minimum possible RT should be the same for all search set sizes. There are several weaknesses with this argument; however. First and most important, it assumes capacity is not limited at the level of the individual element or, in other words, that the individual item-processing time does not become greater as the processing load (i.e., the search set size) increases. In limited capacity self-terminating models, minimum RT will increase with search set size because the mean processing time for the first item to complete processing increases.

This argument, while true for both serial and parallel models, is more plausible in the parallel case because of the awkwardness of interpreting limited capacity at the level of the individual element or item in serial systems. We will confront this situation repeatedly. In a serial model, the only (natural) way the processing time of the first item selected can depend on the total number of items in the search set is if the system knows or reacts to this total number before comparison begins. A not totally absurd proposition is that some kind of preprocessing might affect the subsequent serial comparison rate. Another, especially pertinent to ET situations, is that some type of lateral interference could effectively lower capacity even though the comparison process itself is serial.

Nevertheless, a more natural form of serial limited capacity is system fatigue as processing continues. In such a scheme, the individual item-processing time would increase with the number of items already completed rather than a priori with the total number of items in the search set. Thus the processing time for the kth item would be greater than for the $(k-1)$th. The problem with this conception, however, is that if a self-terminating strategy is assumed, no increase in minimum RT with processing load is predicted.

The second problem with the minimum RT argument is statistical in nature. Even if the population minimum does not increase with search set size, the sample RT minimum will, in general. The rationale is as follows: As the search set size is increased, the proportion of times that the target is the first item completed decreases. In most experiments, the same number of trials are run at each search set size, so that in general there will be fewer instances on which search was terminated after only one item was processed for the larger search set sizes. This means that the sample size that determines the minimum RT will be much smaller for the larger search set sizes. Given the same distribution, the sample minimum tends to decrease as sample size is increased, but in a way that depends on the specific distribution (Gumbel 1958). Thus, even if search is self-terminating and capacity is unlimited, the *sample* RT minimum will tend to increase with the search set size.

This effect will depend heavily on sample size. For experiments with many trials the predicted increase in the sample minimum will generally be so small that for all intents and purposes it can be ignored. It is only for smaller sample sizes that this effect becomes significant. As a rather idealized example of this, assume a process with a constant base time of 200 msec and with exponentially distributed individual item-processing times with a mean of 40 msec, and assume an experiment with 20 trials per search set size. With a sample size of N and an exponential distribution with rate v the sample minimum has probability density Nve^{-Nvt} and so the expected sample minimum is $1/Nv$.

Thus, under the above conditions the expected minimum RT when search set size is 1 is

$$E(\min \mathbf{RT}_1) = 200 + \frac{40}{20} = 202$$

When the search set size is 2 (or more) we will make the simplifying assumption that the expected minimum RT is affected only by those trials on which the target is the first item completed. With a processing load of 2 items, this will occur, on the average, on half of the trials, and so

$$E(\min \mathbf{RT}_2) = 200 + \frac{40}{20/2} = 204$$

Similarly,

$$E(\min \mathbf{RT}_n) = 200 + \frac{40}{20/n} = 200 + 2n$$

The increase in the expected minimum RT is thus seen to be 2 msec for each item added to the search set. For instance, with search set size ranging from 1 to 5 items we would expect a 10-msec increase in minimum RT from this unlimited capacity self-terminating model. Such an increase may or may not be detectable in an experimental setting. However, if there is even a slight limitation in capacity or if the sample size is reduced even further, the observable

In any case, these parameter mappings yield target-present and absent mean RT curves that are both linear with slope D. That is,

$$E(\mathbf{RT}_n^-)=Dn+t_- \quad \text{and} \quad E(\mathbf{RT}_n^+)=Dn+t_+$$

We will learn that other self-terminating models can predict linear and parallel target-present and absent curves. It is generally true, however, that these models are not the simplest in their class and are certainly not simpler than their exhaustive counterparts. Even so, we are unable to invoke Occam's razor here because of the difficulties with serial exhaustive models that we discussed in the preceding section. For this reason, we are forced to look at more complex models, be they self-terminating or exhaustive.

Increases in minimum reaction time

A second finding that has been used to discredit self-terminating models is that the minimum RT increases with the size of the search set (Sternberg 1975; Lively 1972; Lively & Sanford 1972). The rationale behind this argument is quite simple. If search is self-terminating, then no matter how large the search set, there should always be some nonzero probability that the target is the first item processed and thus the minimum possible RT should be the same for all search set sizes. There are several weaknesses with this argument, however. First and most important, it assumes capacity is not limited at the level of the individual element or, in other words, that the individual item-processing time does not become greater as the processing load (i.e., the search set size) increases. In limited capacity self-terminating models, minimum RT will increase with search set size because the mean processing time for the first item to complete processing increases.

This argument, while true for both serial and parallel models, is more plausible in the parallel case because of the awkwardness of interpreting limited capacity at the level of the individual element or item in serial systems. We will confront this situation repeatedly. In a serial model, the only (natural) way the processing time of the first item selected can depend on the total number of items in the search set is if the system knows or reacts to this total number before comparison begins. A not totally absurd proposition is that some kind of preprocessing might affect the subsequent serial comparison rate. Another, especially pertinent to ET situations, is that some type of lateral interference could effectively lower capacity even though the comparison process itself is serial.

Nevertheless, a more natural form of serial limited capacity is system fatigue as processing continues. In such a scheme, the individual item-processing time would increase with the number of items already completed rather than a priori with the total number of items in the search set. Thus the processing time for the kth item would be greater than for the $(k-1)$th. The problem with this conception, however, is that if a self-terminating strategy is assumed, no increase in minimum RT with processing load is predicted.

The second problem with the minimum RT argument is statistical in nature. Even if the population minimum does not increase with search set size, the sample RT minimum will, in general. The rationale is as follows: As the search set size is increased, the proportion of times that the target is the first item completed decreases. In most experiments, the same number of trials are run at each search set size, so that in general there will be fewer instances on which search was terminated after only one item was processed for the larger search set sizes. This means that the sample size that determines the minimum RT will be much smaller for the larger search set sizes. Given the same distribution, the sample minimum tends to decrease as sample size is increased, but in a way that depends on the specific distribution (Gumbel 1958). Thus, even if search is self-terminating and capacity is unlimited, the *sample* RT minimum will tend to increase with the search set size.

This effect will depend heavily on sample size. For experiments with many trials the predicted increase in the sample minimum will generally be so small that for all intents and purposes it can be ignored. It is only for smaller sample sizes that this effect becomes significant. As a rather idealized example of this, assume a process with a constant base time of 200 msec and with exponentially distributed individual item-processing times with a mean of 40 msec, and assume an experiment with 20 trials per search set size. With a sample size of N and an exponential distribution with rate v the sample minimum has probability density Nve^{-Nvt} and so the expected sample minimum is $1/Nv$.

Thus, under the above conditions the expected minimum RT when search set size is 1 is

$$E(\min \mathbf{RT}_1)=200+\frac{40}{20}=202$$

When the search set size is 2 (or more) we will make the simplifying assumption that the expected minimum RT is affected only by those trials on which the target is the first item completed. With a processing load of 2 items, this will occur, on the average, on half of the trials, and so

$$E(\min \mathbf{RT}_2)=200+\frac{40}{20/2}=204$$

Similarly,

$$E(\min \mathbf{RT}_n)=200+\frac{40}{20/n}=200+2n$$

The increase in the expected minimum RT is thus seen to be 2 msec for each item added to the search set. For instance, with search set size ranging from 1 to 5 items we would expect a 10-msec increase in minimum RT from this unlimited capacity self-terminating model. Such an increase may or may not be detectable in an experimental setting. However, if there is even a slight limitation in capacity or if the sample size is reduced even further, the observable increase may be substantial. In addition, if the individual item-processing time distribution has a smaller variance-to-mean ratio than the exponential, a larger increase in the minimum RT can also be generally expected.

On the other hand, an increase in sample size will greatly reduce the effect. For instance, in the above example, if 40 trials are run with each search set size instead of 20, the predicted increase in minimum RT is only half as great. Each successive doubling of the sample size halves the increase, and so we see that the effect is virtually eliminated if the sample size is large enough. In any event, minimum RT increases do not seem as problematic for the self-terminating model as naive intuition suggested.

In general, self-terminating models have little difficulty with the empirical results we earlier labeled as troublesome for serial exhaustive models. This is particularly true of serial position effects, multiple target results, and the RT variance literature. The reason is that for these phenomena, the serial exhaustive model runs into difficulty primarily because of the exhaustive nature of the search rather than the seriality involved.

Parallel search

Historically, parallel search strategies have not been well understood. For instance, parallel interpretations are sometimes discounted in the literature when there is any increase in RT with processing load. This is, of course, a misguided and hasty conclusion. Even unlimited capacity models generally predict an increase in exhaustive mean RT with increases in processing load. In fact, in general, only a very small class of deterministic, supercapacity, and/or correlated parallel models predict no RT increase with processing load. For instance, a supercapacity, independent parallel model with exponentially distributed individual item processing times predicts a flat exhaustive mean RT curve (with a mean of c msec) if the rate of each item, when the search set size is n, is

$$v_n = \frac{1}{c} \sum_{1}^{n} \frac{1}{i}$$

Another parallel model that predicts a constant exhaustive processing time assumes unlimited capacity on elements ($v_n = v$) with complete reallocation of capacity within trials; individual total completion times are thus correlated positively in this model (see Chapter 4 on capacity).

Linear target-present or target-absent curves

More common is to find arguments against parallelity when the increase in RT is found to be at least linear (Sternberg 1966). We already know this argument to be valid only for a restricted type of parallel model (independent, unlimited capacity). Just a few pages ago we encountered a parallel (self-terminating) model that predicts linear increasing target-present and absent curves (see Proposition 6.2). In fact, the prediction of a negatively

accelerated mean RT function is characteristic of only a certain class of parallel models, namely those for which the capacity limitations are not too severe. To further underscore this point, later in this chapter we consider a parallel capacity reallocation model postulated by Atkinson et al. (1969) and by Townsend (1969, 1974b) that exactly mimics the standard serial exhaustive model proposed by Sternberg (1966). This model assumes a fixed and hence limited capacity source that is equally divided among all uncompleted items. As soon as an item is completed, the capacity assigned to it is instantaneously reallocated and spread evenly among the remaining uncompleted items.

Unitary vs. multiple capacity sources

A more serious criticism involves implicit capacity assumptions that some (e.g., Sternberg 1975) have argued are made by parallel models (and especially limited capacity parallel models). This argument was specifically directed at the capacity reallocation model. Sternberg (1975) argued that it implicitly assumes a unitary capacity source and that dual task studies in which the inclusion of a secondary task has only a small effect on the mean RT of the primary task are evidence against unitary capacity sources (Darley, Klatzky, & Atkinson 1972; Ellis & Chase 1971).

The question of how many capacity sources there are is a complex and hotly debated issue (Navon & Gopher 1979) and we must proceed with caution here. Although there is still disagreement, the idea that there is more than one capacity source is probably the more popular at this time. The capacity reallocation model does assume a fixed unitary pool of capacity that is drawn on by all subprocesses of the comparison process. This does not mean, however, that the model presumes that all human information processing is supplied by a unitary capacity source. Comparison may have its own capacity source that is not shared with other information-processing subsystems and that might be reallocatable on a within-trial basis. This question is really logically distinct from that concerning distinct sources for different tasks or subsystems. In any case, in spite of the relative popularity of the multiple capacity source viewpoint, the evidence in its favor is not overwhelming. For instance, the studies on which Sternberg has based his arguments appear to us inconclusive. For example, Darley et al. (1972) showed that the size of a small memory load (from one to four items) does not affect a same–different task (in this case, an ET study in which the memory search set size is always 1), but they neglected to include a control condition and thus failed to show that their memory task required any capacity at all or even that their same–different task did (see Kantowitz 1974 for a discussion of the importance of control conditions in double stimulation paradigms). If a task does not require capacity or requires only an insignificant amount, even a unitary capacity source model predicts that its use as a secondary task in a dual task paradigm will have little disrupting effect on the mean RT of the primary task.

On the other hand, Ellis and Chase (1971) had observers do a size discrimination task at the same time they did a memory-scanning task, and they did show that both of these tasks required capacity. However, the observers were asked to integrate the information and to make only one response (respond "no" if no target is present or if there is a size difference), thus making interpretation of their results more difficult.

No matter how this conflict is resolved, it is important that this criticism of the capacity reallocation model not be taken as prejudicial to parallel processing in general. For instance, assuming unlimited capacity is functionally the same as assuming that each "channel" of the parallel process has its own capacity source. Alternatively, we could view unlimited capacity as one large pool from which each channel could fill its capacity dipper without draining the pool dry. Thus, unlimited capacity has both a unitary and a multiple capacity source interpretation. This is generally true at all capacity levels. In fact, the situation is often clouded even further by the existence of still other interpretations. For instance, with limited capacity we might imagine that each channel has its own capacity source but that these sources overlap, or must share among themselves in some way. This type of situation is a sort of combination of unitary and multiple source models.

The point to be made here is that, in general, assuming a parallel search strategy places few constraints on the details of the capacity structure. The two issues are logically independent in the same way the parallel vs. serial issue is logically independent of the self-terminating vs. exhaustive issue. The one assumption crucial to parallel processing is that *some* capacity be allocated to each element not yet completed.

Finally, although the fixed capacity reallocation model is often selected for rebuttals against parallel models, nonreallocation (e.g., independent) parallel models can just as easily mimic standard serial mean RT predictions, as we saw in Proposition 6.2.

The exponential distribution

A final criticism that really has nothing to do with whether processing is parallel but is nevertheless usually associated with parallel models is directed at the use of the exponential distribution in RT theorizing (Sternberg 1975). This is an issue we briefly discussed in Chapter 3. It is somewhat unclear how the criticism came to be associated with parallel models in the first place since it is also commonly incorporated into serial models (Restle 1961; McGill 1963; McGill & Gibbon 1965). The complaint is that exponential distributions have, because of their memoryless property (see Chapter 3), historically been used to model waiting times and that this fact somehow rules them out as models of processing times.

It is true that one reason for the popularity of the exponential distribution is its mathematical tractability, but this is not the only reason. Reaction time density functions notoriously have high tails, and high tails are associated with intensity or hazard functions whose rate of increase is very slow. (See Ashby 1982a for a more detailed discussion of this point.) If the density function tail is high enough, the hazard function will even decrease, meaning that the longer the process has been continuing the less likely it will be completed in the next instant.

There is thus good evidence that the hazard function of one or more RT components will be nonincreasing or will increase only slowly. Exponential distributions have flat hazard functions and they are thus good candidates for models of RT components. Knowing this, it is not surprising that empirical evidence for exponentially distributed RT components is available (Snodgrass, Luce, & Galanter 1967; Ratcliff & Murdock 1976; Ashby & Townsend 1980; Ashby 1982a). Empirical evidence, not historical precedent, should decide issues such as this.

It is also germane in this context to emphasize that when considering the mean RTs the exact underlying distribution is of little importance. Thus, virtually any family of densities may be employed in a parallel model to mimic the *means* of the standard serial model, as long as some type of pliability with respect to capacity structure exists. The latter requirement is weak; for example, simply shifting a distribution to the right as n increases may represent a capacity degradation.

Specific alternatives to the serial exhaustive model

We will now take the time to consider in some detail three models, conceptually quite different from the standard serial exhaustive model, which nevertheless predict linear and parallel target-present and absent curves. By no means is this intended as a complete survey of the literature. The field is vast and continually growing, and thus any survey would most likely be both incomplete and too lengthy to include here. For instance, we will regretably not discuss system theoretic models based on neuronal fundamentals (Anderson 1973; Grossberg 1969), nor will we have space to examine trace strength models (Corballis, Kirby, & Miller 1972; Nickerson 1972; Baddeley & Ecob 1973). Trace strength models assume that when an item is presented, the memory location where it should be stored is immediately and instantaneously accessed. The task of the processor is to decide if the trace strength associated with that position exceeds some preset criterion. If it does, a "yes" response is given, and if not, a "no" response is made.

Trace strength models bear similarities to the parallel models we consider throughout this book and to certain more highly structured models that are capable of making both speed and accuracy predictions. In particular, we have in mind here the counting models to be discussed in Chapter 9. Thus, although no specific discussion of trace strength models will be made, models that are structurally similar will be considered in detail.

Of the three models we have chosen to discuss in this section, one is serial, one is parallel, and one could be formulated as either. The models were chosen because, first of all, they fall within the same general schema as the other latency models we discuss in this book, and secondly, because within that class the three models display rather marked differences. This serves to illustrate the variety of different processing mechanisms that can lead to the same sorts of mean RT predictions.

The very powerful *pushdown stack model,* which we discuss first, is a serial self-terminating model capable of predicting a very large set of observed empirical phenomena. Secondly, the *capacity reallocation model* is a parallel model based on the intuitively appealing notion that capacity that is freed when an item is completed can be reassigned to hasten the processing of yet uncompleted items. Finally, the *non-Donderian response bias model* is a self-terminating model that predicts parallel target-present and absent curves by postulating a temporal overlap of the comparison and response selection stages.

The pushdown stack model

Among the best-known serial self-terminating models is the pushdown stack model developed by John Theios and his colleagues (Falmagne & Theios 1969; Theios & Smith 1972; Theios et al. 1973; Theios 1973). This model not only predicts serial position effects and the characteristic decrease in RT that accompanies the appearance of redundant targets in the search list, but it also predicts parallel and linear target-present and absent curves. In addition, a guiding principle behind the development of the model was to have as many results as possible predicted by its natural internal structure rather than by appended post hoc assumptions. In particular, this was meant to include stimulus probability and sequential effects that serial exhaustive models, for example, must relegate to a response selection or stimulus encoding stage. It is a well-established empirical fact that RT is shorter the more probable the stimulus (Hyman 1943; Falmagne 1965) and that it is also shorter on those trials for which the target item is the same as it was on the preceding trial (Bertelson 1961, 1965), at least for fairly short response–stimulus intervals (see, e.g., Kirby 1980 for effects of longer intervals). The pushdown stack model nicely predicts both of these effects without having to rely on any post hoc assumptions.

The model derives its name by assuming that memory is arranged as a pushdown stack, that is, that there is a hierarchical ordering to memory with the more frequently presented items tending toward the top of the stack. The popular analogy here is to the almost magical stack of lunch trays so often encountered in cafeterias. We pick a tray up and the rest of the stack rises until its place is taken by the newly exposed tray. Similarly, when a tray is replaced on top of the stack the whole stack settles down somewhat so that the topmost tray is almost always at the same level. In a pushdown stack, trays are accessed on a last-in, first-out basis.

Search through the pushdown memory stack postulated by the model is always from top to bottom, serial, and self-terminating. The lunch tray

analogy breaks down somewhat because when an item is presented that is already represented in memory, its representation does not automatically move to the top of the stack. It may stay where it is or it may move up to some higher level; however, it never moves downward. A jump to a higher level moves all intervening representations down one level.

Memory representations are assumed to consist of a stimulus–response pair rather than the more traditional stimulus alone. Notice that this obviates the need for a response selection stage, since stimulus identification and response selection now both take place during the comparison process. The number of levels or positions in short-term memory is a parameter of the model. Exactly one representation can be stored in each of these positions. Long-term memory is assumed to occupy the first level of the memory stack *below* all of the short-term levels. Many representations can be stored in the long-term memory level of the stack, but long-term memory is searched only after all levels of the short-term memory have been.

To facilitate discussion of the model imagine an LT or memory-scanning experiment where the stimulus ensemble consists of 10 items, half of which never occur in the search set. This is precisely the experiment of Theios et al. (1973). The model assumes that all 10 items are represented in memory along with the response they are associated with. In this case, when a target item is presented, the observer begins a sequential search down through the memory set for the target item's representation. When the representation is encountered, the response associated with that item is revealed. This will be true even if the target item is not contained in the memory set. Thus *both* target-absent and target-present trials are self-terminating and this is how the model accounts for parallel target-present and absent curves.

Serial position effects arise when an item's position in the search set influences the position of its representation within the memory stack. As for multiple target effects, the more replicas of a given stimulus the search set contains, the higher in the stack its memory representation will most likely be stored. Then, when that item is presented as a target, RT will tend to be shorter, because, on the average, fewer levels of the stack will have to be searched. Both sequential effects and stimulus probability effects are also predicted from the general characteristic of the memory stack that the more frequently items are presented, the higher in the stack the corresponding representations are stored.

The model very successfully predicts the results of the experiment outlined above. It easily outperforms the standard serial exhaustive model. The linear parallel target-present and target-absent curves are economically predicted (parameterwise) from the clever and intuitively reasonable assumption that search is self-terminating on both target-present and target-absent trials.

If the model can be faulted, it is in its lack of generalization to other experimental paradigms. This is a general problem that tends to increase with model complexity, and the pushdown stack model is surely more complex than, say, the standard serial exhaustive model. The added complexity comes primarily from its detailed assumptions about the memory stack, and in particular in the way the model intertwines stimulus identification and response selection. The standard serial exhaustive model is currently mute about such details, as is illustrated by the fact that it needs more structure to predict stimulus probability effects. It is surely insufficient in the long run to simply say that the locus of such effects must be in response selection or stimulus encoding.

The pushdown stack model loses some of its intuitive appeal when we consider results of experiments in which the memory set is a small subset of some large class of alphanumeric items and that it changes its membership from trial to trial by drawing items from the entire stimulus set, as in Townsend and Roos (1973), for example. The most rational strategy here appears to be to associate a "no" response with all stimulus items. When the memory set is now presented to the subject, each of the corresponding memory representations can be located within the memory stack and reassociated with a "yes" response. At the same time, these items tend to move vertically up through the stack. This state of affairs should cause no problem on target-present trials but might pose some difficulties predicting results on target-absent trials. With a large stimulus set, such as the letters of the English alphabet, many (most) of the memory representations will be stored in the lowest level of the stack, the so-called long-term level. The model typically assumes that for the kinds of stimuli used in these experiments the mean time to retrieve an element from long-term memory, t_L, is the same for all stimulus representations residing there and that the individual item mean memory comparison time t_c is the same for all items in the short-term memory levels of the stack (Theios et al. 1973). It is thus hard to see how target-absent RT can increase with search set size, as we know it must.

Providing an appropriate generalization is difficult due to the mathematical intractability of the model. Past applications have relied on computer simulations to remedy this problem. It turns out that the model does predict a tendency for target-absent RT to increase with search set size, but in general only at a very slow rate. The increase in RT occurs because larger search set sizes will, on the average, cause fewer items associated with "no" responses to be represented in short-term memory since they will tend to be bumped out by the search set items. This means that as the search set size increases, it becomes more and more likely that on target-absent trials the target item representation will be stored in the long-term level of the stack, since on these trials it is associated with a "no" response. The longest possible RTs occur when the desired representation is in the lowest level of the stack, the long-term level. Thus, as search set size increases, the proportion of long target-absent RTs will increase and hence the target-absent curve will increase.

Let us try to get some rough idea of how sharp this increase might be. Assume the stimulus set is the English alphabet, that search set size ranges from 1 to 5, and that the short-term memory stack has 7 positions or levels.

We will also assume that presentation of the search set always causes the representations of the search set items to be stored in the topmost positions in the stack. These assumptions allow us to write the expected target-present RT when the search set size is n as $E(\mathbf{RT}_n^+) = [(n+1)/2]t_c + t_+$, where t_+ is the mean base time and t_c is the mean time to search one level of the short-term memory stack. The predictions are a little more complicated on target-absent trials:

$$E(\mathbf{RT}_4^-) = \tfrac{1}{22}(5t_c) + \tfrac{1}{22}(6t_c) + \tfrac{1}{22}(7t_c) + \tfrac{19}{22}(7t_c + t_L) + t_-$$

The first term reflects the comparison time on those trials on which the target representation is located in the short-term stack directly below the 4 search set item representations. Thus a total of 5 positions of the stack are searched before the target is discovered. Hence the mean comparison time is $5t_c$. The probability that the target will be found in this position is $\frac{1}{22}$, since a search set size of 4 means 22 letters of the alphabet will be associated with a "no" response. The second term in the above expression concerns the case when the target is in position 6 of the stack, the third term when it is in position 7, and the fourth term when it is in long-term memory. The probability the target representation is stored in long-term memory is $\frac{19}{22}$ or 1 minus the probability it is stored in the short-term stack [i.e., $1 - (\frac{1}{22} + \frac{1}{22} + \frac{1}{22})$]. Similarly, when the search set size is 5,

$$E(\mathbf{RT}_5^-) = \tfrac{1}{21}(6t_c) + \tfrac{1}{21}(7t_c) + \tfrac{19}{21}(7t_c + t_L) + t_-$$

Simplifying these two expressions gives

$$E(\mathbf{RT}_4^-) = 6.864t_c + .864t_L + t_-$$

and

$$E(\mathbf{RT}_5^-) = 6.952t_c + .905t_L + t_-$$

so the target-absent RT curve does increase. However, if we compare the rate of increase to the target-present curve, we find

$$E(\mathbf{RT}_5^+) - E(\mathbf{RT}_4^+) = .5t_c$$

whereas

$$E(\mathbf{RT}_5^-) - E(\mathbf{RT}_4^-) = .088t_c + .041t_L$$

The two curves can increase at the same rate only if t_L is quite large. For example, if t_c equals a reasonable 30 msec, then for the two curves to increase at the same rate the time to retrieve an element from long-term memory, t_L, must be over 3 sec, an absurdly long value considering that the to-be-recalled items are letters of the English alphabet. It thus appears that the pushdown stack model must add some special assumptions if it hopes to predict results of experiments in which search set size is much less than the size of the entire stimulus alphabet.

The capacity reallocation model

We turn back now to a parallel model that we briefly mentioned earlier in this chapter, that is, the capacity reallocation model proposed by Townsend (1969, 1974b) and independently by Atkinson et al. (1969). (See also Chapter 4, "The capacity issue.") When the model was originally developed it was one of the first parallel models capable of predicting linear and parallel target-present and target-absent curves. Indeed, this was the major impetus for its development.

The model is predicated on the assumption that when an item completes processing, the capacity that was allocated to it is suddenly freed. The second assumption is that this suddenly available capacity can be instantaneously diverted or reallocated to aid the processing of the remaining uncompleted items. The most widely known, and the original, version of the model postulated exponentially distributed intercompletion times. Under this distributional assumption, the capacity reallocation property is equivalent to introducing a constraint on the processing rates that says that at any time t during processing, the sum of the rates is always constant. These assumptions yield a parallel model that is identical to the standard serial model on mean (RT as well as intercompletion time) statistics, as we will shortly see. Therefore, if exhaustive processing is assumed, there is no way to discriminate the standard serial exhaustive model from the exhaustive reallocation parallel model on the basis of mean RTs.

If there are n items in the search set and a fixed amount of capacity is always distributed among the items, then the processing time of the first item completed will be the minimum of n exponentially distributed random times where the sum of the n rates is v. From Proposition 3.9 the expected processing time of the first item completed in this parallel model is therefore $1/v$.

The capacity freed by the completion of this item is now reallocated to the remaining items, and so the sum of the processing rates during the second stage is still v and thus the expected duration of the second intercompletion time is also $1/v$. This state of affairs is repeated until processing is completed. The expected target-present and target-absent processing times are therefore $E(\mathbf{RT}_n^+)=n/v+t_+$ and $E(\mathbf{RT}_n^-)=n/v+t_-$. These are exactly the same predictions given by the standard serial exhaustive model.

It might be noted that the model makes no assumptions about how the capacity is divided up among the items. It may or may not be spread evenly. The mean RT predictions of the model are the same no matter how it is divided up.

As mentioned earlier, Sternberg (1966) argued that linear target-absent curves falsify independent parallel models because these are constrained to predict negatively accelerated RT functions, and thus any increase they predict must be bounded above by some linear function. Unlike the majority of parallel exponential models we have encountered, however, because of its capacity reallocation property this model does not predict independence of

the individual item completion times and thus does not contradict Sternberg's assertion. However, the parallel, self-terminating model we considered earlier (in Proposition 6.2) that predicts linear target-present and target-absent curves is an independent parallel model of limited capacity. Therefore, it is the capacity and not the independence issue that is critical to the prediction of linear target-present and target-absent curves. Independent parallel models appear to be more versatile than was once thought.

The capacity reallocation model with exhaustive processing is open to the same criticism as the standard serial exhaustive model because it is equivalent to it on mean statistics. For instance, in general the model predicts no serial position effects. We therefore cannot discriminate between these different classes of models given the data collected in typical ET and LT studies. For this reason the capacity reallocation parallel model should not be viewed as superior to the standard serial model. However, it should be viewed as a viable alternative since all data supporting the serial model also support the parallel model.

The idea of capacity reallocation is an intuitively compelling notion. It smacks of optimality. The system uses the energy or attention available to it to the fullest. When some is freed upon the completion of a given task, it is immediately diverted elsewhere and not allowed to languish about. One might imagine such plasticity evolving over the eons.

It is possible to generalize this idea of optimal capacity reallocation to other nonstandard search tasks. For instance, consider a visual search (ET) task in which the a priori probability that a target appears in a given location is different for the various display locations. Then, although capacity reallocation is still optimal, a strictly parallel search is not. Marilyn Shaw (Shaw & Shaw 1977; Shaw 1978) generalized the capacity reallocation model to just this type of task. Her work serves to illustrate the potential versatility of capacity reallocation assumptions and represents, we feel, a promising alternative to the more traditional memory and display search modeling attempts.

An optimal search model

The capacity allocation model of Shaw assumes that the observer is capable of and actually attempts to optimize his or her search performance in the sense of maximizing the probability that a target will be detected, if one exists, for any given search time. The success of optimal models of the human observer or operator (Green & Swets 1966; Sheridan & Ferrel 1974) suggests that an optimal search model might provide a good description of experienced human search performance.

Shaw's model is based on the mathematical theory of optimal search, which was developed by Koopman (1956a, 1956b, 1957) during World War II in an attempt to solve the problem of how best to search the ocean for enemy submarines. More specifically, the problem can be stated as follows: Given a limited amount of resources available to conduct a search, what is the optimal allocation of these resources that maximizes the probability of detecting the target within some specific cost limit? In the case of memory and display search a good candidate for the limited set of resources might be a limited capacity source or some finite amount of available attention.

Shaw had to make several assumptions to adapt optimal search theory to memory and display search tasks. First, she assumed that attention is available and expended at a constant rate v over time and thus that the total attention expended through time t, $A(t)$, is $A(t)=vt$. This assumption is fairly common in psychological theorizing (Townsend & Ashby 1978).

Now let $a(j,t)$ be the amount of attention allocated to position j through time t and let $v(j,t)$ be the instantaneous attentional output to position j at time t. Here (in the Townsend and Ashby terms) $a(j,t)$ is in energy dimensions whereas $v(j,t)$ is in power terms. Thus

$$a(j,t)=\int_0^t v(j,x)\,dx$$

The second assumption of the model is that the total available attention v is divided among the n possible positions in such a way that none is wasted; that is

$$v=\sum_{j=1}^{n} v(j,t) \quad \text{for all } t>0$$

Finally, define $F[j,a(j,t)]$ as the probability of detecting a target that is in position j given that the total amount of attention allocated to that position through time t is $a(j,t)$. $F[j,a(j,t)]$ is very much like a cumulative distribution function and in fact becomes one if we assume, as Shaw does, that a target is certain to be eventually detected (i.e., so that $F[j,a(j,\infty)]=1$), at least when it always has some instantaneous attention allocated to it. Shaw's third assumption is that

$$F[j,a(j,t)]=1-\exp[-a(j,t)]=1-\exp\left[-\int_0^t v(j,x)\,dx\right] \tag{6.5}$$

She then invokes a theorem due to Stone (1975) that says that if $F[j,a(j,t)]$ is concave (down) and continuous in t, the allocation strategy a that minimizes mean search time for a given position j is exactly the one that maximizes the probability of finding the target for every time t, which we decided is a condition for optimality. This theorem is very handy. It tells us that solving the easier problem of finding the allocation strategy that minimizes mean search time is the same as solving for the allocation strategy that maximizes the probability of target detection after any given search time t. Shaw argues that because of her third assumption the conditions of this theorem are met in the present model so that we can use it to solve for the optimal allocation strategies.

Contrary to Shaw, however, Eq. 6.5 does not imply concavity and therefore her third assumption is not strong enough to satisfy the conditions of Stone's theorem (compare also Eq. 3.7 of Chapter 3). To see this, note that (assuming the relevant functions are all differentiable)

.125 each

\$ ___ ___ ___ ___ \$

.25 each

\$ ___ ___ \$

Fig. 6.4. One of the display configurations used by Shaw (1978). A target appeared on every trial. The probability it occurred in each of the four upper left-hand positions was .125 and the probability it appeared in either of the lower right-hand positions was .25.

$$\frac{d}{dt}F[j,a(j,t)]=\frac{d}{dt}\left\{1-\exp\left[-\int_0^t v(j,x)\,dx\right]\right\}$$
$$=v(j,t)\exp\left[-\int_0^t v(j,x)\,dx\right]$$

Therefore, an equivalent way of stating Shaw's third assumption is as[1]

$$v(j,t)=\frac{(d/dt)F[j,a(j,t)]}{1-F[j,a(j,t)]}$$

Written in this fashion it is easy to see that $F[j,a(j,t)]$ need not be concave. Any nondecreasing function whose range is confined to the real interval $[0,1)$ defines a possible $v(j,t)$, and therefore some further restrictions on either $F[j,a(j,t)]$ or $v(j,t)$ are required.

One extra restriction that does guarantee the concavity of $F[j,a(j,t)]$ and the one implicitly assumed by Shaw is that $v(j,t)=v_j$ for all $t>0$, and is therefore a constant (perhaps piecewise) over time. In this case,

$$F[j,a(j,t)]=1-\exp(-v_j t)$$

To test the model Shaw (1978) conducted a visual search task where she varied the probability distribution of the target location in a nonlinear array such as is in Fig. 6.4. The target, either an F or a Z, appeared in each of the four upper left positions with probability .125 and appeared in each of the lower right positions with probability .25. A target thus appeared on every trial. The observer's task was to move a lever to the right if the presented target was Z and to the left when the target was F. The other five positions were always filled with distractor letters, and so there were always a total of six letters presented on every trial.

To understand the optimal allocation strategy in this experiment let us sim-

[1] To see that the implication goes in both directions, substitute this expression into the right-hand side of Eq. 6.5 and perform the integration.

plify the situation somewhat by assuming there are only two locations and that the probability the target is in position 1, $P(1)$, is greater than the corresponding probability for position 2, $P(2)$. Then the optimal strategy under the above formulation (which includes the assumption that $F[j, a(j, t)]$ is exponential) allocates all capacity to position 1 until time t_0, when the posterior probability that the target is there equals the posterior probability that the target is in position 2. After time t_0 the optimal strategy is to divide capacity evenly among the two positions.

Now the posterior probability that the target is in position 1 after time t has been spent searching that position, when $t \leqslant t_0$, is

$$P(\text{target is in position 1} \mid \text{target has not been found by time } t)$$

$$= \frac{P(1)P(\text{target has not been found by time } t \mid \text{target is in position 1})}{P(\text{target has not been found by time } t)}$$

$$= \frac{P(1)P(\text{target has not been found by time } t \mid \text{target is in position 1})}{\sum_{i=1}^{2} P(i)P(\text{target has not been found by time } t \mid \text{target is in position } i)}$$

$$= \frac{P(1)e^{-vt}}{P(1)e^{-vt}+P(2)}$$

The idea here is that up until time t_0 all capacity is allocated to position 1 and so the probability that a target in position 1 has not yet been found is

$$1-F[1, a(1,t)] = 1-(1-e^{-vt}) = e^{-vt}$$

On the other hand, if the target is in position 2, this probability will be 1 since the target cannot be found until some capacity is allocated to position 2.

Similarly, the posterior probability that the target is in position 2 after time t has been unsuccessfully spent searching position 1 is

$$\frac{P(2)}{P(1)e^{-vt}+P(2)}$$

Note that this posterior probability will always be greater than $P(2)$, the a priori probability the target is in position 2. This is because an unsuccessful search for the target in position 1 increases the chances a target will be found in position 2.

The optimal strategy, therefore, is to allocate all capacity to position 1 until

$$\frac{P(1)\exp(-vt_0)}{P(1)\exp(-vt_0)+P(2)} = \frac{P(2)}{P(1)\exp(-vt_0)+P(2)}$$

that is, up to time

$$t_0 = \left(\frac{1}{v}\right)\ln\frac{P(1)}{P(2)}$$

and thereafter to divide capacity equally among positions 1 and 2. In other words, to begin with, a serial strategy is employed, with all capacity being focused on a single position, and this is followed by an equal-attention parallel strategy. For the display of Fig. 6.4 the strategy then is to divide v equally among the locations having target probability .25 (so $v/2$ is assigned to each) until time $t_0 = (2/v)\ln 2$ and then to assign capacity $v/6$ to each location.

To test the model Shaw derived the predicted difference in expected RTs in the two cases when the target is in one of the upper left-hand positions of Fig. 6.4 and when it is in one of the lower right-hand positions and then compared this prediction with the observed difference in mean RTs. Let us denote the two kinds of target locations in Fig. 6.4 by a and b, where the probability that the target is in position i (for $i = a, b$) is $P(i)$ and n_i locations have this target probability. Then it can be shown (Shaw 1978) that the optimal search model predicts

$$E(\mathbf{RT}_b) - E(\mathbf{RT}_a) = \frac{1}{v}\left\{n_b\left[1-\frac{P(b)}{P(a)}\right] + n_a \ln\left[\frac{P(a)}{P(b)}\right]\right\} \tag{6.6}$$

A comparison of this prediction with the observed difference in mean RTs requires the estimation of only one parameter, that is, $1/v$. Note that if we solve the above equation for $1/v$ and replace expectations by observed means, we can estimate $1/v$ by

$$\frac{1}{v} = \frac{\overline{\mathrm{RT}}_b - \overline{\mathrm{RT}}_a}{n_b[1-P(b)/P(a)] + n_a \ln[P(a)/P(b)]} \tag{6.7}$$

If we now use the same set of data, for instance, the data collected by using the display of Fig. 6.4, to both estimate $1/v$ and to compare the observed and predicted mean RT difference, the model will appear to fit perfectly since the predicted difference in mean RT will exactly equal the observed difference, as can be seen by plugging the Eq. 6.7 estimate of $1/v$ into the Eq. 6.6 prediction. What is needed is an estimate of $1/v$ obtained from an independent set of data. In an effort to satisfy this requirement, Shaw incorporated a second display into the experiment, in which the target location probability distribution was different than in Fig. 6.4. An estimate of $1/v$ can be obtained from the data collected using either display. Shaw averages these two estimates and uses the result to estimate $1/v$ in Eq. 6.6. These two estimates should be the same if the model is correct. Unfortunately, Shaw does not report the individual $1/v$ estimates so that we can compare them.

In this rather limited test, Shaw (1978) found that the optimal search model fit the observed mean RT differences reasonably well for five of eight observers in one experiment and four of six in another. This hardly represents overwhelming evidence in favor of the model, but it is a start. More rigorous and comprehensive tests of the model are clearly called for.

Note that although the model is self-terminating, it is in general neither serial nor parallel but instead is hybrid. Even so, it clearly appears to be more closely aligned with parallel processing. For instance, with the typical memory-scanning or visual search display in which the a priori target probability is the same for all locations, the optimal allocation strategy is to divide the available attention v equally among all display locations. The result of this strategy is a fixed capacity, parallel, self-terminating process. Thus, parallel processes result for all positions having the same a priori target prob- ability. A serial strategy is thus seen to be inferior to a parallel one under these criteria. We shall consider the optimality of parallel vs. serial strategies again in Chapter 8.

When we first began discussing this model, we mentioned that it predicts linear target-present curves. We now verify this claim. Shaw (1978) shows that in the capacity allocation model

$$E(\mathbf{RT}^{+}_{j,n}) = \frac{1}{v}\left\{\sum_{i=1}^{j}\left[1-\ln\left(\frac{P(j)}{P(i)}\right)\right] + \sum_{i=j+1}^{n}\frac{P(i)}{P(j)}\right\}$$

where $\mathbf{RT}^{+}_{j,n}$ is the RT when the search set size is n and the target is in position j. In a standard memory or display search paradigm the a priori target proba- bility is the same for all display locations. With n items in the search set this probability is $1/n$. Thus in a memory or visual search task

$$E(\mathbf{RT}^{+}_{j,n}) = \frac{1}{v}\left\{\sum_{i=1}^{j}\left[1-\ln\frac{1/n}{1/n}\right] + \sum_{i=j+1}^{n}\frac{1/n}{1/n}\right\}$$

$$= \frac{1}{v}\left[\sum_{i=1}^{j}1 + \sum_{i=j+1}^{n}1\right] = \frac{n}{v}$$

Since this prediction does not depend on j, the expected RT is the same for all target locations, that is

$$E(\mathbf{RT}^{+}_{n}) = E(\mathbf{RT}^{+}_{j,n}) = \frac{n}{v} \quad \text{for all } j \leqslant n$$

and thus the target-present curves are linear functions of n. The model pre- dicts no serial position effects and thus is open to the same criticism as the standard serial exhaustive model on these grounds.

As it stands, the model does not have the structure necessary to predict target-absent results. This is because the model assumes that the observer always keeps searching for a target until one is found. There is no mechanism available for the observer to stop the search at some point in time and decide that a target does not exist in a given location. This clearly calls for modifica- tion. Perhaps the most obvious strategy is to terminate the search at a given location, with the conclusion that no target exists there, when the posterior probability that a target is there after time t has been spent searching for it first falls below some criterion value. This approach appears mathematically feasible, although it will complicate the above expressions. Analytically it is equivalent to assuming that when the posterior probability falls below a cri- terion level, all capacity is permanently reallocated away from that position. This means that under normal ET and LT experimental conditions the model is equivalent to the parallel capacity reallocation model we considered at the beginning of this section. In other words, Shaw's model can be viewed as a generalization of the capacity reallocation model to situations in which the a priori target probability varies across display positions.

The optimal search model developed by Shaw represents a promising alter- native to the standard parallel and serial search models. It is based on the plausible assumption that humans attempt to optimize the probability that a target will be discovered in any given search time t. The model, as it stands, needs much more extensive empirical testing, but even if it turns out that human search performance is nonoptimal (in the above sense) the contribu- tion of the model could still be significant, for knowledge of how and why human search performance is nonoptimal will be valuable in aiding our understanding of this important cognitive process.

A non-Donderian response bias model

The third general model we will consider is a self-terminating model capable of predicting linear, parallel target-present and absent curves. It postulates that the equal slopes are the result of two separate processing stages, namely comparison and response selection, rather than just one (i.e., comparison). In 1868 Donders conceived of the RT processing chain as a series of discrete nonoverlapping subsystems, an idea that has exerted pro- found influence over RT theory since that time. The present model is non- Donderian because it postulates a temporal overlap between the comparison and response selection processes (see also Chapter 12). The idea is based on the fact that as more and more nontarget items are completed, the probability that the ultimate correct response will be "no" increases. We can verify this fact rather easily. To do so, assume n items are in the search set and that the first $n-1$ items completed are all nonmatches. The a priori probability of a "no" response is $\frac{1}{2}$, and so if the probability of a "no" response is now greater than $\frac{1}{2}$, our statement is verified. We thus wish to evaluate

$$P(\text{"no"} \mid 1\text{st } n-1 \text{ items are nonmatches})$$

$$= \frac{P(\text{"no" \& 1st } n-1 \text{ items are nonmatches})}{P(1\text{st } n-1 \text{ items are nonmatches})}$$

$$= \frac{P(\text{all } n \text{ items are nonmatches})}{[P(1\text{st } n-1 \text{ items are nonmatches \& } n\text{th item is a nonmatch}) + P(1\text{st } n-1 \text{ items are nonmatches \& } n\text{th item is match})]}$$

$$= \frac{P(\text{target-absent trial})}{P(\text{target-absent trial}) + P(\text{target-present trial \& target is completed last})}$$

$$= \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}(1/n)} = \frac{n}{n+1}$$

which is substantially greater than $\frac{1}{2}$ for all $n > 1$.

Thus as more nontarget items are completed it becomes more likely that the correct response will be "no." The non-Donderian response bias model takes advantage of this fact by assuming that the response selection process begins to anticipate a "no" response as its probability increases. On target- absent trials this anticipation will tend to shorten response time, whereas on target-present trials it will actually slow things down since the inertia toward a "no" response has to be overcome. Even so, overall such an anticipation will tend to speed responding, since a "no" response (and hence a shortening of RT) is so much more likely than a "yes" response on trials when the first $n-1$ completed items are nontargets. At any rate, the anticipation of "no" responses will tend to make target-absent RTs faster and target-present RTs slower and could offset the 2:1 slope ratio of target-absent to target-present RT curves characteristic of standard serial self-terminating models.

A precursor of the present model is found in a guessing strategy proposed by Nickerson (1966), who postulated that after processing several items associated with the same response (e.g., nontargets) there was a certain probability that the observer would prematurely terminate processing and guess that response. However, the model in its present form was developed by Townsend (1973b, 1974b) and Taylor (1973, 1976c).

At this point in our development, we could assume that search is either serial or parallel; the equations are only slightly different in the two cases. Either assumption can lead to linear target-present and target-absent curves with the same slope. We will choose a serial interpretation, because the resulting equations are somewhat easier to manipulate.

We begin by assuming that capacity is unlimited at the individual item level for both targets and nontargets so that

$$\frac{1}{n}\sum_{i=1}^{n} E(\mathbf{T}_{i,n}^{+}) = C = \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{T}_{i,n}^{-}) \tag{6.8}$$

This does not mean that all target rates and all nontarget rates must be equal, but only that the average target rate is equal to the average nontarget rate. The model can (but does not have to) predict a large variety of serial position effects. At any rate, under the constraint of Eq. 6.8, the expected RT on target-absent trials when the search set size is n is given by

$$E(\mathbf{RT}_n^-) = nC + E(\mathbf{Y}) - E(\mathbf{Z}_{n-1}) + t_- \tag{6.9}$$

The first term is just the mean comparison time when search is serial, $\mathbf{Y}$ is the unbiased random response selection time, t_- is the expected base time component, which in this case includes all RT components other than comparison and response selection, and $\mathbf{Z}_{n-1}$ is the random amount of time that response selection is shortened because of the shift toward a "no" response caused by the processing of the first $n-1$ *non*targets. The reason that the subscript on $\mathbf{Z}$ is $n-1$ rather than n is because it is assumed that the final response decision would normally occur immediately after the completion of the nth item, so that the results of the nth comparison do not further bias the starting time of the response selection process.

Similarly, the self-terminating mean on target-present trials is given by

$$E(\mathbf{RT}_n^+) = \left(\frac{n+1}{2}\right)C + E(\mathbf{Y}) + \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{Z}_{i-1}) + t_+ \tag{6.10}$$

Notice that the same value $\mathbf{Z}_j$ is present here as was in the target-absent RT

prediction and that it is added here to the expected RT rather than subtracted from it. This means we are assuming that the savings in time afforded by the completion of j nontarget items on a target-absent trial is equal to the increase (or "waste") in RT caused by having to overcome the inertia toward a "no" response on target-present trials that results from processing j nontarget items before completing the target. Of course, this assumption need not be made. Instead we could postulate two different random variables $\mathbf{Z}_j^+$ and $\mathbf{Z}_j^-$, but for now the model will be left as it is.

The $1/n$ term in the $E(\mathbf{RT}_n^+)$ equation represents the probability that the target is the ith item completed when there are a total of n items in the search set. Thus when the target is the jth completed item the RT prolongation is given by $\mathbf{Z}_{j-1}$ since $j-1$ nontargets will already have been completed. Before processing begins, there is no predilection toward either response, and so $E(\mathbf{Z}_0)=0$.

What now of the promised linear and parallel target-present and target-absent curves? The following proposition addresses this goal.

Proposition 6.3: If the individual-item processing times are unlimited capacity as in Eq. 6.8, then the serial self-terminating non-Donderian response bias model predicts the target-present and target-absent mean RT curves to be linear functions of n with the same slope if and only if $E(\mathbf{Z}_j)=\frac{1}{3}Cj$.

Proof: First note from Eq. 6.9 that $E(\mathbf{RT}_n^-)$ is a linear function of n if and only if $E(\mathbf{Z}_{n-1})=an+b$. Recall now that $E(\mathbf{Z}_0)=0$, and so $b=-a$. Thus $E(\mathbf{Z}_{n-1})=a(n-1)$. We can now rewrite Eqs. 6.9 and 6.10 as

$$E(\mathbf{RT}_n^-) = nC + E(\mathbf{Y}) - a(n-1) + t_-$$

and

$$E(\mathbf{RT}_n^+) = \left(\frac{n+1}{2}\right)C + E(\mathbf{Y}) + \frac{1}{2}a(n-1) + t_+$$

These two functions rise with the same slope if and only if

$$E(\mathbf{RT}_n^-) - t_- = E(\mathbf{RT}_n^+) - t_+$$

This equality holds if and only if

$$nC - a(n-1) = \left(\frac{n+1}{2}\right)C + \frac{1}{2}a(n-1)$$

Solving for a yields $a=\frac{1}{3}C$ and thus

$$E(\mathbf{Z}_{n-1}) = \tfrac{1}{3}C(n-1) \quad \square$$

Thus the target-present and target-absent mean RT curves are linear functions of n with the same slope whenever $E(\mathbf{Z}_j)$ is a linear function of n with slope $C/3$ msec. In other words, the extra time saved in response selection on target-absent trials is the same for each successive nontarget item completed. This savings is roughly equal to $\frac{1}{3}$ of the average individual item processing time. Similarly, on target-present trials, response selection is slowed by an extra $C/3$ msec after each successive nontarget item is completed as the inertia toward a "no" response becomes more and more difficult to overcome.

The non-Donderian response bias model is a self-terminating model in which search can be either serial or parallel. It can predict target-present and target-absent curves that rise at the same rate, and because it is a self-terminating model, it can predict a great diversity of serial position effects. It therefore seems a good candidate for most ET and LT studies. Certainly the ease with which it produces serial position effects makes it a more viable model in many instances than the standard serial exhaustive model.

In this section we examined three models that are potentially competitive with the standard serial exhaustive model for typical ET and LT data. Two of these are self-terminating and one is exhaustive. Both self-terminating models rely on the interaction of at least two separate subsystems to predict the relative equality of target-absent and target-present RT functions and the exhaustive model does not easily predict serial position effects. Nevertheless, as we saw earlier in this chapter, there do exist self-terminating models that predict parallel target-present and absent RT functions entirely within the comparison system. Furthermore, we also saw exhaustive models that predict serial position effects within the comparison process. A great many studies have been run and conclusions drawn as if parallel target-present and absent mean RT functions logically implied an exhaustive search and as if serial position effects logically implied self-termination. The above results indicate the infirmitude of such conclusions.

This completes our very sporadic foray into the universe of viable alternatives to the standard serial exhaustive model. Our presentation was not meant to be complete, but to illustrate the diversity that competing models can take and still predict the major characteristics of RT data.

A class of models falsified by parallel target-present and target-absent curves

We have now seen many widely different models capable of predicting linear and parallel target-present and absent curves. One should not get the impression, however, that the large majority of models are able to predict such results. We already know, for example, that many parallel models, with unlimited capacity or supercapacity, predict negatively accelerating mean RT functions.

We also know that the standard serial self-terminating model, in which all individual item processing rates are identical, predicts a 2:1 slope ratio and therefore is falsified by parallel target-present and absent curves. In fact, it turns out that a very large number of serial self-terminating models cannot predict such curves and so are falsified when such data are found.

To support this contention, we present a large class of such serial self-terminating models originally isolated by Townsend and Roos (1973). To keep the class as large as possible we need to generalize our notation a bit.

Recall that we were letting $E(\mathbf{T}_{i,n}^{+})$ represent the expected processing time of the target in position i when the load is n items. We now need another subscript to denote processing order. Let $E_j(\mathbf{T}_{i,n}^{+})$ be the same expected processing time, except now assume that the jth processing path was taken through the n items, where j runs from 1 to $n!$. Finally let P_j be the probability that the jth path is chosen.

The models we will examine are defined by the following two assumptions:

$$\text{(A1)}\qquad E(\mathbf{T}_{\cdot n}^{+})=\sum_{j=1}^{n!} P_j\left[\frac{1}{n}\sum_{i=1}^{n} E_j(\mathbf{T}_{i,n}^{+})\right]=E(\mathbf{T}_{\cdot 1}^{+})$$

$$\text{(A2)}\qquad E(\mathbf{T}_{\cdot n}^{-})=\sum_{j=1}^{n!} P_j\left[\frac{1}{n}\sum_{i=1}^{n} E_j(\mathbf{T}_{i,n}^{-})\right]=E(\mathbf{T}_{\cdot 1}^{-})$$

The assumptions state that the target and nontarget processing times, respectively, remain constant over n, when averaged across all serial positions and processing orders.

The class of models satisfying these assumptions is quite large, as (A1) and (A2) still allow processing times to vary across serial position, to vary with processing load, to vary with element identity, and to vary with processing path. The requirement is that these differences average out to be the same for each search set size in the case of both targets and nontargets. To take just one well-known example, it should be clear that the standard serial self-terminating model, which postulates just one processing rate parameter, satisfies (A1) and (A2). In this model, $E_j(\mathbf{T}_{i,n}^{+})=E_j(\mathbf{T}_{i,n}^{-})=E(\mathbf{T})$ for all values of j, i, and n. If we now average over all serial positions and processing orders we see, for example, that

$$\begin{aligned}E(\mathbf{T}_{\cdot n}^{+})&=\sum_{j=1}^{n!} P_j\left[\frac{1}{n}\sum_{i=1}^{n} E(\mathbf{T})\right]\\&=E(\mathbf{T})\sum_{j=1}^{n!} P_j=E(\mathbf{T})=E(\mathbf{T}_{\cdot 1}^{+})\end{aligned}$$

and so assumption (A1) holds.

Of course, the standard serial self-terminating model is only one of the many models satisfying (A1) and (A2). The following result holds for every model in this class.

Proposition 6.4: No serial self-terminating models that satisfy assumptions (A1) and (A2) can predict parallel target-present and target-absent mean RT curves.

Proof: To show that this class of models is indeed falsified by parallel target-present and absent curves, it suffices to show that the functions are unable to maintain equal slopes even for search set sizes 1 and 2. First note that in the target-present case,

$$E(\mathbf{RT}_1^+)=E_1(\mathbf{T}_{1,1}^+)+t_+=E(\mathbf{T}_{\cdot 1}^+)+t_+$$

by assumption (A1). When the search set size is 2,

$$E(\mathbf{RT}_2^+)=\tfrac{1}{2}\{P_1E_1(\mathbf{T}_{1,2}^+)+P_2[E_2(\mathbf{T}_{2,2}^-)+E_2(\mathbf{T}_{1,2}^+)]\}$$
$$+\tfrac{1}{2}\{P_1[E_1(\mathbf{T}_{1,2}^-)+E_1(\mathbf{T}_{2,2}^+)]+P_2E_2(\mathbf{T}_{2,2}^+)\}+t_+$$

Position 1 is searched first with probability P_1, and with probability $P_2=1-P_1$ position 2 is checked first. The first term accounts for that half of the trials for which the target is in position 1, whereas the second term handles cases in which the target is in position 2. Using assumption (A1) again, we can rewrite this expression as

$$E(\mathbf{RT}_2^+)=E(\mathbf{T}_{\cdot 1}^+)+\tfrac{1}{2}[P_1E_1(\mathbf{T}_{1,2}^-)+P_2E_2(\mathbf{T}_{2,2}^-)]+t_+$$

Turning now to the target-absent expressions, we see that

$$E(\mathbf{RT}_1^-)=E_1(\mathbf{T}_{1,1}^-)+t_-=E(\mathbf{T}_{\cdot 1}^-)+t_-$$

this time by (A2). For search set size 2,

$$E(\mathbf{RT}_2^-)=P_1[E_1(\mathbf{T}_{1,2}^-)+E_1(\mathbf{T}_{2,2}^-)]+P_2[E_2(\mathbf{T}_{1,2}^-)+E_2(\mathbf{T}_{2,2}^-)]+t_-$$
$$=2E(\mathbf{T}_{\cdot 1}^-)+t_-$$

Our proof will be complete if we can show that $E(\mathbf{RT}_2^-)-E(\mathbf{RT}_2^+)\neq E(\mathbf{RT}_1^-)-E(\mathbf{RT}_1^+)$. First note that

$$\begin{aligned}E(\mathbf{RT}_2^-)-E(\mathbf{RT}_2^+)&=2E(\mathbf{T}_{\cdot 1}^-)-E(\mathbf{T}_{\cdot 1}^+)-\tfrac{1}{2}P_1E_1(\mathbf{T}_{1,2}^-)\\&\quad-\tfrac{1}{2}P_2E_2(\mathbf{T}_{2,2}^-)+(t_--t_+)\\&=[E(\mathbf{T}_{\cdot 1}^-)-\tfrac{1}{2}P_1E_1(\mathbf{T}_{1,2}^-)-\tfrac{1}{2}P_2E_2(\mathbf{T}_{2,2}^-)]\\&\quad+[E(\mathbf{T}_{\cdot 1}^-)-E(\mathbf{T}_{\cdot 1}^+)]+(t_--t_+)\\&=\tfrac{1}{2}[P_1E_1(\mathbf{T}_{2,2}^-)+P_2E_2(\mathbf{T}_{1,2}^-)]\\&\quad+[E(\mathbf{T}_{\cdot 1}^-)-E(\mathbf{T}_{\cdot 1}^+)]+(t_--t_+)\end{aligned}$$

Meanwhile,

$$E(\mathbf{RT}_1^-)-E(\mathbf{RT}_1^+)=[E(\mathbf{T}_{\cdot 1}^-)-E(\mathbf{T}_{\cdot 1}^+)]+(t_--t_+)$$

Therefore, parallel target-present and absent curves occur only if

$$\tfrac{1}{2}[P_1E_1(\mathbf{T}_{2,2}^-)+P_2E_2(\mathbf{T}_{1,2}^-)]=0$$

and this equality holds only if

$$E_1(\mathbf{T}_{2,2}^-)=E_2(\mathbf{T}_{1,2}^-)=0$$

In other words, this class of self-terminating models predicts that target-present and absent curves rise with the same slope only when any nontarget item finishing second is processed infinitely fast. We can immediately reject this possibility as absurd, however, leaving us with the conclusion that the above class cannot predict parallel curves. □

From the proof we see that the slopes of the target-present and target-absent curves differ by the amount

$$\tfrac{1}{2}[P_1E_1(\mathbf{T}_{2,2}^-)+P_2E_2(\mathbf{T}_{1,2}^-)]$$

This "error term" may be substantial. For example, suppose that

$$E_1(\mathbf{T}_{2,2}^-)=E_2(\mathbf{T}_{1,2}^-)=30 \text{ msec}$$

which seems reasonable in memory and display search studies. In this case, the difference between the slopes of the target-present and absent curves will be 15 msec, which should be fairly easy to detect in most experimental settings.

This concludes our discussion of the standard memory-scanning and visual search tasks. Before we conclude the chapter, however, we will briefly consider some closely related experimental paradigms.

Related paradigms; current and future directions

During the time the visual and memory search paradigms were being explored and models were being suggested to account for the major results, several other closely related experimental paradigms were also being studied. Although to some extent there did exist problems and controversies unique to these different paradigms, by and large the processes postulated were of much the same nature in them all. While the results of any one experiment might not prove conclusive in delineating the underlying processing structures, a study of results obtained from different experimental paradigms might provide enough converging evidence to, at least, make some tentative conclusions possible. With this in mind, we take a bit of time to consider some of these related paradigms, and as we do we will keep the fundamental principle of system observability as a guidelight to our endeavors.

Simultaneous memory and visual search

It has been pointed out (Wingfield & Bolt 1970) that at times, memory scanning and visual search, as we have defined them above, are difficult to distinguish. In fact, when the search set size is 1, the two paradigms are logically identical. This functional similarity can be used to generalize the two experimental designs in such a way that a new paradigm is constructed that contains memory scanning and visual search as special cases. In this new paradigm, the list presented first, the memory list, contains M items and the list presented second, the display list, contains D items. The observer is typically told to respond "yes" if the memory and display lists have any items in common.

The first modern application of this paradigm apparently was by Nickerson (1966), who found that for display lists containing more than one item (i.e., for $D>1$) "no" RTs increased more quickly than "yes" RTs as the memory set size increased. Soon thereafter, Sternberg (1967) found essentially the same results. From these findings Sternberg argued that search through the memory set is exhaustive but that search through the display list is self-terminating. Thus, on a given trial the observer is assumed to select an item from the display list and compare it exhaustively to the items in the memory set. If, after this comparison has been completed, a match is discovered to have occurred, then the search process is terminated and a positive response is given. If no match is detected, a second item is selected from the display list and the process is repeated. A negative response is given only after the last display item has been compared to all the memory items and no match is found.

The task facing the observer in this paradigm is much more complicated than in either an ET or an LT design. As such, it provides more observability in the sense that a wider and more detailed data structure is available, but the cost of this advantage is the increased complexity of the candidate descriptive models. For example, in place of the four possible combinations of the parallel–serial and self-terminating–exhaustive dimensions that are typically combined in ET and LT analysis, we now have 16 distinct possible combinations that must be checked. The typical strategy has been to ignore one of the two dimensions and concentrate on the other, thereby reducing the number of candidate models fourfold. For example, Howell and Stockdale (1975) assumed processsing was serial and then concerned themselves with the self-terminating vs. exhaustive issue. On the basis of their results they argued, as Sternberg (1967) did, for an exhaustive search through memory and a self-terminating visual search.

Another application of this general paradigm, which actually fit a number of mathematical models, was reported by Rossmeissl, Theios, and Krunnfusz (1979). They simultaneously allowed both the memory set and the display set to contain either 1, 2, 4, or 6 items and recorded both mean RT and RT standard deviation. Both were found to increase approximately linearly with the total number of comparisons possible (i.e., $M \cdot D$), with the negative curves increasing more rapidly than the positive curves. At least for mean RT this is the same result found by Sternberg (1967). Rossmeissl et al. worked out both the standard deviation and mean RT predictions for a large number of models and found that within the serial class the best-fitting model incorporated a self-terminating search through both the visual and memory displays and postulated the same processing time for every item. Within the parallel class, the best-fitting model was limited capacity with differential target and nontarget processing rates (i.e., targets faster). When the serial and parallel models were compared it was found that the parallel model fit the observed RT standard deviations better.

This study is especially important because of the authors' attempts to fit the RT standard deviations as well as the means. Such a strategy can only increase model identifiability, and so will probably become more and more prevalent in RT theorizing as time goes by. We will adopt it in the next chapter to aid us in discriminating self-terminating from exhaustive search strategies. Use of even higher moments of the RT distribution is at this time discouraged because of the large standard error associated with their estimates (Ratcliff 1979). On the other hand, interest in the RT density functions seems to be growing (Ashby & Townsend 1980; Bloxom 1979; Ashby 1982a). To a large extent this interest is made possible by the work on nonparametric probability density estimation (see Tapia & Thompson 1978 for a review of this work) during the last 15 years or so by mathematical statisticians. As a result of their labors, estimates far superior to those produced by the age-old histogram techniques are now available. The increase in identifiability that results from fitting RT standard deviations as well as means should be even more dramatic when the entire RT densities are utilized.

We now turn to another closely related paradigm.

Same–different paradigms

Same–different paradigms were born about the same time as the memory-scanning paradigm from a rather vigorous interest in the fundamental processes involved in the matching of external pattern information to internal. The idea was simple: Have the observer respond "same" as quickly as possible if all members of the display set match all of the memory set items; if there is any discrepancy between the two sets, the observer responds "different."

The paradigm has seen many slight variations on this theme. For instance, the two sets of items are sometimes all presented at the same time (Egeth 1966; Posner & Mitchell 1967). As another example, Taylor (1976b) included a *some same–all different* condition in which the observer is instructed to respond "same" if any items in the two sets match and "different" only if no matches are found. This type of design is very similar to the simultaneous visual and memory search paradigm we just considered; however, in most same–different tasks the order of items is crucial, whereas in simultaneous visual and memory search it is not. For instance, in Taylor's some same–all different condition the lists *ABCD* and *XAXX* require an "all different" response, whereas in simultaneous visual and memory search they require a "yes."

Posner and Mitchell (1967) also elaborated on the response instructions given the observers. In their level 1 instructions, observers responded "same" only if the items were physically identical (e.g., *A* and *A*). Under level 2 instructions, items were to be classified "same" so long as they had the same name, even if they were not physically identical (e.g., *a* and *A*), and in level 3 a "same" response was to be given if the items were both vowels or both consonants. The idea here is that as the response level increases, a deeper, more "profound" level of processing must occur for the observer to complete the task (Craik & Lockhart 1972). For instance, to determine that an item is a vowel one presumably must already have processed it to the level of its name.

This idea of response levels, though quite popular in the same–different task (Bamber 1972; Posner & Mitchell 1967), has probably not received enough attention in the ET and LT paradigms. For instance, we shall see later in Chapter 13 the important role target and nontarget rate differences play in enhancing parallel–serial discriminability in multisymbol comparison tasks. It might be possible to use the idea of levels of processing to ensure that targets and nontargets are processed at different rates. For example, suppose our search set consisted of the items *AbCD* and that the target item was *b*, but that the observer was told to respond "yes" if any items of the search set had the same name as the target (i.e., level 2 instructions). Since physically identical items have the same name, the match of *b* and *b* would occur very quickly, leading to a very fast target comparison rate. On the other hand, the nontarget items would have to be processed to the "deeper" level of name identity so that it should take a longer time for them to be identified as nontargets, thus resulting in a fairly large target–nontarget rate difference and so, hopefully, in increased parallel–serial discriminability.

The results that are typically obtained in same–different tasks are partly consistent with what we would expect based on our knowledge of ET and LT studies, but they also contain something of a surprise. In many cases, the data on "different" trials look as though they might have come from an ET or LT study. Mean RT is found to increase with the number of items in the mismatching sets and for a given set size is found to decrease with the number of discrepant items. Models of the type we have discussed in this chapter have proved successful in describing these results. For instance, Bamber (1969) very successfully fit a simple single-rate serial self-terminating model to the "different" RT data he collected.

On the other hand, the "same" RTs turn out to be somewhat anomalous. The simple search models that did such a good job predicting the "different" RT data fail miserably. For instance, Bamber's (1969) serial self-terminating model predicts a serial exhaustive scan on "same" trials since the only way a "same" response can be given is if a match occurs in *every* position of the two lists. If this strategy is actually used, then the "same" RTs should increase at a faster rate than the "different" RTs. Not only was this prediction not observed, but the exact opposite state of affairs was obtained. "Same" RTs increased with processing load much more slowly than "different" RTs.

Bamber (1969) accounted for these untoward results by postulating two fundamentally different processes that operate in parallel on same–different tasks. The first is the fairly slow but very familiar item-by-item self-terminating comparison process. The second is a much faster, more holistic "identity reporter" that very quickly signals the observer when two stimulus patterns are identical. Either of these processes can signal "same," but only the slower serialistic one can signal "different."

Although many researchers reported these fast "same" responses (Bamber 1969, 1972; Nickerson 1972; Taylor 1976b), not everyone agreed with Bamber's hypothesis of a dual process as the mediating mechanism.[2] A major criticism (Taylor 1976b) is that an "identity reporter," if it exists, ought to also make possible fast "different" responses. This is because the lack of a signal from the "identity reporter" should indicate some discrepancy between the two stimulus configurations and thus indicate that a "different" response is appropriate. Bamber's (1969) response to this criticism is that it is sometimes quicker to go ahead and perform the slower item-by-item comparison than to wait for the identity reporter. Another possibility is that it might be more accurate to perform the serial search than to respond "different" whenever the identity reporter is silent. For instance, suppose the identity reporter does not always respond when the two stimulus lists are identical. Fast "same" responses could still result if the proportion of times it did respond is high enough, but the observer could no longer rely on this subsystem for perfect performance and so would have to wait for the completion of the serial scan on trials on which the identity reporter is silent.

[2] Fast "same" responses are typically only found when the stimuli differ on several dimensions, as with letters of the English alphabet. When stimuli differ on a single dimension, "same" responses often taken longer than "different" responses. For instance, it generally takes longer to respond "same" to lines of equal length than it does to respond "different" to lines of unequal length (Bindra, Donderi, & Nishisato 1968).

Another process that has been invoked to explain the observed discrepancy between "same" and "different" RTs is rechecking on "different" trials (Bamber 1969; Tversky 1969; Howell & Stockdale 1975; Krueger 1978). The idea here is not so much that "same" responses are fast but that "different" responses are slowed by the observer's rechecking of the mismatching item. Krueger (1978) notes that such a strategy makes sense only if the internal item representations or the comparison process itself are known to be noisy (see also Townsend 1974b: 178–79, with regard to potential influences of noise in such contexts). If internal representations are always perfect and item comparison is always error-free, then rechecking is redundant and hence a waste of time. Krueger (1978) developed noisy operator theory as an elaboration of this idea.

The basic postulate of noisy operator theory is that internal noise may deform the encoded representation of a stimulus item (or the comparison process) so that a positive or "same" match may not yield a perfect congruence. Thus a "same" response will often be made on the basis of an imperfect match. Krueger sees the comparison process as one in which the number of perceived discrete differences (e.g., features) are counted or accumulated. If this number is below some criterion level, a "same" response is given, if it is above some other value a "different" response is made, but if an intermediary value is obtained, Krueger assumes rechecking occurs and that the new difference count is added to the old one. The observer is also assumed to adjust his criterion values (i.e., increase them) with each successive scan so that the whole process can be repeated over and over again until a response is made. The accumulating nature of the count process bears similarity to the random walk models and the counting models that we consider in detail in Chapters 9 and 10. Since noise is much more likely to increase a difference count than to decrease it, this model predicts there will be more false "different'' responses than false ''same'' responses, a finding that is often reported in the literature (Taylor 1976b; Beller 1970).

All in all, the model seems fairly successful in predicting the rudimentary results of both the accuracy and latency data. As such, it is one of a growing number of new and sophisticated models these paradigms are spawning. As the models become more sophisticated it becomes harder and harder, and at the same time less meaningful, to bifurcate them with respect to a single dimension of processing such as parallel vs. serial or self-terminating vs. exhaustive. This makes model testability difficult, for when a model becomes very complex the probability that every one of its assumptions is correct must become fairly small. Even if only one assumption is wrong, the model may not provide a completely adequate fit to the data, but to reject it on the basis of a poor fit alone is not fair to the remaining set of correct assumptions.

To be sure, perhaps a far greater problem is the acceptance of models or theories when they are incorrect. It has become apparent that investigators place much more trust in their data than is usually justified by the amount they have collected (Tversky & Kahneman 1973), probably especially when the data are supportive of a pet theory or model. When conjoined with the disposition of most publication outlets to theoretical confirmation, the direction of bias becomes clear.

Another factor supplementing this bias may be the relative insensitivity of some of our favored tests. For instance, despite its several advantages, the chi-square test is known not to be very powerful in rejecting alternative (and untrue) models (Massey 1951). The actual state of affairs of theory testing at any particular point in time appears to be a function of the developmental level of the discipline interacting with the attendant scientific mores of the contemporary body of investigators. Thus, there is a delicate balance maintained between running so many experimental trials that any current model is deftly ''falsified'' and running so few that almost any putative explanation can be ''verified.'' (Of course, such influences interact with the cost of acquiring data and the like.)

A partial remedy is to perform, where possible, more than one statistical test. When all tests favor (especially ''significantly'') one certain model, then at least we would seem to have discovered the best current explanation. However, when nature does not prove to be so magnanimous, then it is not always obvious what conclusions to draw.

There is another strategy that, to some extent, attempts to eradicate the errors of inaccurate acceptance as well as those of inept rejection. This approach is directed to testing only one or a small set of the assumptions at a time and to then constructing the complex model on the basis of these tests. We saw something like this in the simultaneous visual and memory search paradigm when Howell and Stockdale (1975) concentrated on the self-terminating vs. exhaustive issue after they had assumed a serial search. Of course, one must be sure that the assumptions that are made are at least approximately correct or that they do not much affect the processing dimension under investigation. In Howell and Stockdale's case these requirements were probably met since they found mean RT to increase linearly with the total number of comparisons, so that serial processing *could* have been in action. Since both parallel and serial models can predict this result, the hope is that even if, say, processing is really parallel, assuming a serial search will not greatly affect the outcome of a self-terminating vs. exhaustive test. It should be expected that this type of investigation will depend on the processing dimension involved and on the ability of one class of models to mimic another.

Another example that adopts the precept of ''divide and conquer'' follows.

Varying the expected number of items processed while holding search set size constant

Taylor, Townsend, and Sudevan (1978) were interested in testing between a simple (i.e., one-rate) serial self-terminating model and a simple parallel self-terminating model in a visual search task. They maximized the probability that search was self-terminating by always using a very large, nine-item search set and by varying the number of target items the search set contained.

Both the serial and parallel models that were considered assume all items are processed at the same rate. To keep our notation as uncluttered as possible, let us denote the mean processing time of an item in the serial model by $E_s(\mathbf{T})$ and in the parallel model by $E_p(\mathbf{T})$, just to emphasize that the processing times may be different in the two classes of models. Similarly, let $t^s_\cdot$ ($t^p_\cdot$) be the mean base time in the serial (parallel) model where the dot can be either $+$ or $-$, denoting a possibly distinct decision or response selection duration for positive and negative matches.

Now suppose that C of the nine items in the search set are targets. Then the serial model predicts that

$$E_s(\mathbf{RT}_9^+)=\frac{10}{C+1}E_s(\mathbf{T})+t^s_+ \quad \text{and} \quad E_s(\mathbf{RT}_9^-)=9E_s(\mathbf{T})+t^s_-$$

The term $10/(C+1)$ is the expected number of items that must be processed before the first target is completed. We will make use of this statistic extensively in the next chapter.

The parallel model assumes that capacity is spread evenly among the 9 positions of the search set and that the individual item processing times are independent and exponentially distributed. Thus

$$E_p(\mathbf{RT}_9^+)=\frac{9}{C}E_p(\mathbf{T})+t^p_+ \quad \text{and} \quad E_p(\mathbf{RT}_9^-)=\sum_{i=1}^{9}\frac{9}{i}E_p(\mathbf{T})+t^p_-$$

The term $(9/C)E_p(\mathbf{T})$ in the target-present RT expression is the mean time until the first completion on C independent, exponential, parallel channels, where the processing rate on each channel is $[9E_p(\mathbf{T})]^{-1}$. Note that from this information alone, we cannot tell if the system is unlimited or limited capacity. To do so we must examine the processing rates for different search set sizes, information that is not available (or pertinent) here. Thus, the design of Taylor et al. lessens the concern of whether search is self-terminating or exhaustive, but it also eliminates the need to consider capacity as a relevant processing dimension. It therefore allows us to concentrate on the parallel-serial issue.

The mean target-present RT curves when considered as functions of C, the number of targets in the search set, are similar enough in the parallel and serial case that on the basis of these predictions alone we might have some problems discriminating between the two classes of models. Taylor et al., however, had the idea of plotting mean RT as a function of the expected number of completions and then comparing the predictions of the two models with the resulting plot.

First, it should become clear with a little thought that the two models predict the same expected number of items to be completed for each value of C. We will see in the next chapter that this number is $10/(C+1)$. The reason that both models predict this value is, in the serial case, because the targets were randomly placed in the 3×3 matrix of stimulus items. This guarantees that the probability that any given uncompleted item is the next one selected for processing is the same for all items. For the parallel model, this probability is also the same for all items, except that here the reason is that all items are processed with the same rate.

If we let $N=10/(C+1)$ be the expected number of completions, then for the serial model

$$E_s(\mathbf{RT}_9^+ \mid N)=E_s(\mathbf{T})N+t_+^s \tag{6.11}$$

so that mean RT increases linearly with the expected number of completions with slope $E_s(\mathbf{T})$ and intercept t_+^s.

Now $N=10/(C+1)$ implies that $C=(10-N)/N$, so that in the parallel case we see from above that

$$E_p(\mathbf{RT}_9^+ \mid N=1)=E_p(\mathbf{T})+t_+^p$$

$$E_p(\mathbf{RT}_9^+ \mid N=2)=\tfrac{9}{4}E_p(\mathbf{T})+t_+^p=2.25E_p(\mathbf{T})+t_+^p$$

$$E_p(\mathbf{RT}_9^+ \mid N=3)=\tfrac{27}{7}E_p(\mathbf{T})+t_+^p=3.86E_p(\mathbf{T})+t_+^p$$

Similarly,

$$E_p(\mathbf{RT}_9^+ \mid N=4)=6E_p(\mathbf{T})+t_+^p$$

and

$$E_p(\mathbf{RT}_9^+ \mid N=5)=9E_p(\mathbf{T})+t_+^p$$

The increase in mean RT with the expected number of completions is faster than linear, which means that the serial and parallel models are mean-testable on this prediction of positive acceleration.

Before examining the data, let us analyze the curvature of a more general parallel model that encompasses the exponential as a special case. Assume now that the distribution of the individual element processing times is Weibull, with the following distribution function, survivor function, density function, and hazard function, respectively:

$$G(t)=1-\exp(-at^b)$$

$$\bar{G}(t)=\exp(-at^b)$$

$$g(t)=abt^{b-1}\exp(-at^b)$$

$$H(t)=\frac{g(t)}{\bar{G}(t)}=abt^{b-1}, \quad 0<a,b<+\infty$$

This is an extremely interesting distribution because, not only is it of an especially simple form, but the hazard function decreases, stays the same, or increases, depending on whether b is less than, equal to, or greater than 1. When it is equal to 1, we have the exponential distribution as a special case.

If we continue to assume that all items have the same processing time distribution, then the expected number of items completed when there are C items in the display is still $N=10/(C+1)$. Thus, to compare predictions with the exponential model, we need to first compute $E_p[\mathbf{RT}_9^+ \mid N=10/(C+1)]$ in the more general Weibull case.

Proposition 6.5: In the parallel Weibull model with self-terminating search,

$$E_p\left(\mathbf{RT}_9^+ \mid N=\frac{10}{C+1}\right)=\frac{\Gamma(1/b)}{(aC)^{1/b}}+t_+^p=\frac{N^{1/b}\Gamma(1/b)}{[a(10-N)]^{1/b}}+t_+^p$$

where $\Gamma(1/b)$ is the gamma function of $1/b$ [which is just $(k-1)!$ if $b=1/k$ and k=positive integer] and a and b are the Weibull parameters.

Proof: On target-present trials RT is determined by the first of the C targets to be completed. We will use the fact that the density function of the minimum of C identically distributed random variables is $Cg(t)[\bar{G}(t)]^{C-1}$. Thus,

$$E_p(\mathbf{RT}_9^+ \mid N)=\int_0^\infty t\left\{Cg(t)[\bar{G}(t)]^{C-1}\right\}dt+t_+^p$$

$$=\int_0^\infty tCabt^{b-1}\exp(-at^b)[\exp(-at^b)]^{C-1}\,dt+t_+^p$$

$$=\int_0^\infty abCt^b\exp(-aCt^b)\,dt+t_+^p$$

Letting

$$aCt^b=u, \quad t=\left(\frac{u}{aC}\right)^{1/b} \quad \text{and} \quad dt=\frac{u^{(1-b)/b}}{b(aC)^{1/b}}\,du$$

we get

$$E_p(\mathbf{RT}_9^+ \mid N) = \int_0^\infty bue^{-u} \frac{u^{(1-b)/b}}{b(aC)^{1/b}}\, du + t_+^p$$

$$= \frac{1}{(aC)^{1/b}} \int_0^\infty u^{1/b} e^{-u}\, du + t_+^p$$

which is the gamma function of $1/b$ [i.e., $\Gamma(1/b)$] multiplied by $1/(aC)^{1/b}$. Thus

$$E_p(\mathbf{RT}_9^+ \mid N) = \frac{\Gamma(1/b)}{(aC)^{1/b}} + t_+^p$$

Substituting in $C = (10-N)/N$ concludes the proof. □

We now wish to investigate the curvature of this expectation as a function of the average number of completions by examining its derivatives with respect to N (which we will treat as a continuous variable). The first derivative can be shown to be equal to

$$\frac{dE_p(\mathbf{RT}_9^+ \mid N)}{dN} = \frac{10\Gamma(1/b)N^{(1-b)/b}}{a^{1/b}b(10-N)^{(1+b)/b}}$$

Each of these terms is always positive since N is always less than 10, so that

$$\frac{dE_p(\mathbf{RT}_9^+ \mid N)}{dN} > 0 \quad \text{for all } N > 0$$

Therefore, mean RT always increases as a function of N, just as we would expect.

Curvature, however, is determined by the second derivative. If it is positive, the mean RT vs. N curve is positively accelerated, as with the exponential model. If the second derivative is always negative, the curve is negatively accelerated, and if it is always zero, the curve is linear, as with the serial model. After some simplification, the second derivative can be shown to be

$$\frac{d^2E_p(\mathbf{RT}_9^+ \mid N)}{dN^2} = \frac{10\Gamma(1/b)N^{(1-2b)/b}\{10[(1-b)/b]+2N\}}{a^{1/b}b(10-N)^{(1+2b)/b}}$$

This function is greater than or equal to zero whenever

$$10\left(\frac{1-b}{b}\right) + 2N \geqslant 0$$

Thus, when $b \leqslant 1$, the second derivative is positive, and so mean RT is positively accelerated at all values of N, mimicking the special exponential case. Recall that the parameter b determines the slope of the hazard function in the Weibull distribution. Specifically, when $b \leqslant 1$ the hazard function is nonincreasing. Thus we know that if the hazard function is constant (the exponential case) or decreasing (i.e., processing slows down as a function of time, for instance, by getting tired), then $E_p(\mathbf{RT}_9^+ \mid N)$ is positively accelerated.

Alternatively, if $b > 1$, the hazard function increases, as might be the case if there is some sort of warmup effect or if the individual items are comprised of components, so that less and less of the item remains to be processed as more of the components are completed. In this case, the sign of the second derivative of $E_p(\mathbf{RT}_9^+ \mid N)$ depends on the magnitude of b. Specifically, it is negative whenever

$$N < 5\left(\frac{b-1}{b}\right)$$

Since we have assumed $b > 1$, both sides of this inequality are positive. Thus there may be some smaller values of N for a given b such that

$$\frac{d^2E_p(\mathbf{RT}_9^+ \mid N)}{dN^2} < 0$$

For example, suppose $b = 2$; then if $N < \frac{5}{2}$, the curve is negatively accelerated. Further, when $b \to \infty$, the curve is convex or negatively accelerated for $N < 5$. With one target in a display of 9 items, the expected number of completions, N, is 5; and if there is more than one target, that expected number is less than 5. Thus, for all target-present data, as b becomes large the curve becomes convex.

This suggests at least a cursory test of the shape of the individual item-processing time hazard function. If the mean RT vs. expected number of completions curve is negatively accelerated, an increasing hazard function is suggested, whereas if a positively accelerated function results, a flat or decreasing hazard function is supported.

We might also ask if any set of parameter values can yield the linear function that we saw in Eq. 6.11 is characteristic of the serial model. A linear function results if and only if the second derivative is always zero. Thus we need to ask if any values of a and b exist for which

$$0 = \frac{d^2E_p(\mathbf{RT}_9^+ \mid N)}{dN^2} = \frac{10\Gamma(1/b)N^{(1-2b)/b}\{10[(1-b)/b]+2N\}}{a^{1/b}b(10-N)^{(1+2b)/b}}$$

Now $\Gamma(1/b) > 0$ for all $b > 0$. Further, $a > 0$ and $b < \infty$, whereas N is a variable, and so the only possibility for equality in the above expression is when $10[(1-b)/b] + 2N = 0$, which implies $b = 5/(5-N)$. However, b is a constant and so the equality cannot hold for all values of N. Thus a linear mean RT vs. N curve falsifies this much more general class of parallel models that includes the exponential as a special case.

Figure 6.5 illustrates the predictions of the simple serial and parallel models we first considered along with the mean RTs as reported by Taylor et al. (1978). In the parallel model $E_p(\mathbf{T})$ was set to $E_p(\mathbf{T}) = 7E_s(\mathbf{T})$ and t_+^p to $t_+^p = .3E_s(\mathbf{T}) + t_+^s$. Note that the data points conform very nicely to the serial model predictions and thus that a serial search (or parallel mimic) is clearly supported over an independent parallel search even when a more general

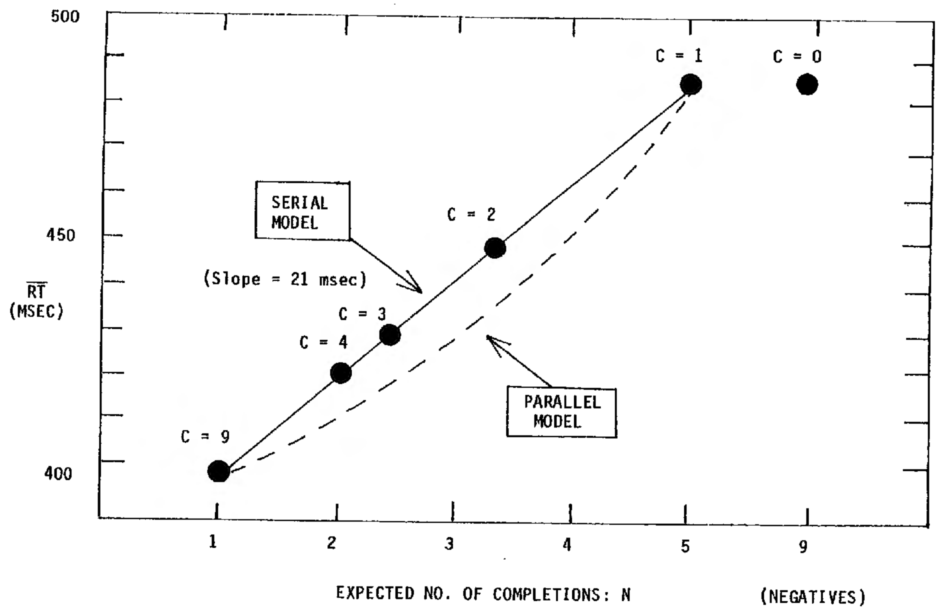


Fig. 6.5. Mean RT versus the expected number of items completed if search is self-terminating from the experiment of Taylor, Townsend, and Sudevan (1978). The solid line is the prediction of a serial model and the broken line is the prediction of a parallel exponential model. The circles give the obtained results. *C* is the number of targets in the display (of nine items).

individual item-processing time distribution is allowed. Observe that independent parallel models cannot seem to predict the constant average intercompletion times.

In a certain sense this strategy of selecting an experimental design that allows one to home in on and test some small set of processing assumptions plays a key role throughout this book. We shall see it in the next chapter where we focus on the self-terminating vs. exhaustive issue, and we shall see it in Chapter 13 where an experimental paradigm that has the potential to test among a rather large class of serial and parallel models is presented.

The parallel–serial testing (PST) paradigm of Chapter 13 accomplishes this goal by utilizing conditions from most of the experimental paradigms we have discussed in this chapter. This is, in general, a very powerful heuristic for maximizing model testability because of the fact that the classes of models mimicking each other in one paradigm are not necessarily the same set of models mimicking each other in a second paradigm. Thus, by incorporating conditions from different paradigms the number of possible mimicking models might be decreased.

This was the motivation behind the recent study of Snodgrass and Townsend (1980), which incorporated memory-scanning, visual search, same–different, and joint memory and visual scanning conditions into the same experiment. They first compared ordinal RT predictions from several large classes of models to the observed RT patterns and on the basis of these comparisons argued for a limited capacity self-terminating model. They then compared the quantitative predictions of serial and parallel models within this class. Although none of the tested models provided a completely adequate account of the observed variability in the data, the serial models provided a substantially better account than the parallel models. The best-fitting serial model, besides being self-terminating, predicted target comparison times to be consistently longer than nontarget comparison times.

In the next chapter we will again see some of these same paradigms as well as some new ones as we concentrate our attention on the self-terminating vs. exhaustive processing dimension.