CHAPTER

14 Designs for and Analyses of Response Time Experiments

Trisha Van Zandt *and* James T. Townsend

Abstract

This chapter provides historical background and a review of the design of response time experiments in psychology and human performance research. It also presents the most common techniques for the analysis of response time data, focusing in particular on parameter estimation and some "meta-theoretic" approaches for testing cognitive architecture.

Key Words: Response Time, Experimental Design, Cognitive Modeling, Data Analysis.

Response times, sometimes referred to as reaction times or latencies, are measured as the time elapsed between the onset of a stimulus and the response to that stimulus. Response times (RTs) are very widely used in the study of human performance. In cognitive psychology and neuroscience, RTs are used to develop and test models of cognitive processing and brain function (e.g., Ratcliff & Smith, 2004). In ergonomics and human factors, sometimes called engineering psychology, they are used to evaluate training regimens, user interface design, vehicle operation performance and to perform task analyses (e.g., Borowsky, Oron-Gilad, & Parmet, 2009; Stevens, Brennan, Petocz, & Howell, 2009; Sullivan, Tsimhoni, & Bogard, 2008). In clinical psychology, psychiatry, and education, they are used to evaluate medical conditions and assist in diagnoses of such conditions as schizophrenia, learning disorders, and other psychological disorders (e.g., Heiervang & Hugdahl, 2003; Querne & Berquin, 2009). Modeling the processes that give rise to RT data forms the foundation for much work in cognitive psychology (Luce, 1986; Townsend & Ashby, 1983).

This chapter discusses the design of RT experiments and how RTs may be analyzed. It should be noted not only that the vista of research involving RTs is vast but also that the design of any experiment depends less on the variable to be measured than on the question that experiment is intended to answer. Some questions need experimental designs in which RT is controlled. Others require designs where RT is the dependent measure. Therefore, we cannot hope to provide a comprehensive index of all issues and designs relevant to RT data, but we can provide a broad summary of the kinds of designs that are likely to be most useful in varying circumstances.

We begin this chapter with a history of RT measurements and the logic behind using RT to discover the structure of mental events. We then present the most common experimental designs, grouping them by the relationship between stimuli and responses. We will then discuss methods of data analysis, including parameter estimation and how RTs are used to test hypotheses about the structure of a cognitive task.

Historically, research that uses RTs can be roughly divided into two major and often overlapping realms. First, RTs have been used to describe changes in performance under different experimental conditions, usually in applied situations. We will present some of these descriptive analyses in the first half of the *Analysis* section. However, the most influential use of RT data has been to answer a theoretical question or test a theoretical hypothesis—for example in determining characteristics of cognitive information processing systems. Theoretical approaches can themselves be classified into (1) verbally based models or theories, (2) models expressed as specific stochastic processes with psychologically meaningful parameters, and (3) "meta-theories" in which entire classes of models based on one or more psychological principle are tested via theory-driven experimental methodologies.

Throughout this chapter we will emphasize the use of RTs in evaluating theories of mental function. Our main focus will be on modeling and meta-theory, approaches that have been most useful in answering questions about how the mind works. Although the modeling approach is presently far more popular than the meta-theoretic approach, the meta-theoretic approach appeared first, and the modeling approach derived from it. Hence, we will discuss the development of the meta-theoretic approach in our brief history and expand on it later in the second half of the *Analysis* section. We conclude the chapter by outlining current approaches to RT data and developing methodology.

History: From Astronomy to the Arrangement of Mental Processes

Some of the earliest recorded attempts to evaluate task performance with response time were made by seventeenth-century astronomers. They referred to the *personal equation* to describe individual differences in the times taken by different observers to estimate the transit times of stars as they moved across the visual field. Exactly measuring the personal equation was important because astronomers hoped to calibrate their equipment to cancel out the effects of these individual differences and so arrive at more accurate measurements of the stars (Duncombe, 1945). Astronomer and mathematician Fredrich Bessel (1784–1846) was even more interested in why there should be such a personal equation. Using what we would now recognize as a psychological approach, he formulated a hypothesis about the interactions between the visual and the auditory systems that we now refer to as the *doctrine of prior entry* (Shore & Spence, 2005).

Over the most recent century, time has been used to explore hypotheses about the architecture of mental processing. Such uses of RT are sometimes called *mental chronometry*, but it is important to recognize that mental chronometry not only includes the measurement of RTs but also the careful control of stimulus exposure times, which determines how much information gets into the perceptual system.

The first application of mental chronometry was performed by Donders, who proposed the *method of subtraction* (Donders, 1868/1969). Donders was inspired by von Helmholtz's (1850) work demonstrating that the time taken by a neuron to transmit information could be measured. If it takes time for nerves to send information around the body, then perhaps it might be possible to estimate the time taken by different components of a mental task. These components are called *stages*, and the tasks he considered are now called simple reactions, go/no-go reactions, and choice reactions.

In a simple reaction an observer responds as soon as he sees any stimulus at all (such as red and green lights) with a single response, such as depressing a telegraph key. Donders reasoned that simple responses require only a perceptual encoding stage, where the perceptual system apprehends the presentation of a stimulus, and a response execution stage, where the key is depressed. By contrast, in a go/no-go reaction, an observer responds to only one of two possible stimuli, pressing the key only when, say, the green light is presented. This task requires perceptual encoding and response execution like the simple reaction and an additional stage of stimulus identification in which observers determine the color of the presented stimulus. For a choice reaction, an observer makes one keypress to a red stimulus but a different keypress to a green stimulus. This task requires all the stages of a go/no-go reaction plus a response selection stage in which the observer chooses the key appropriate to the color of the light presented (*see* Fig. 14.1).

To apply the method of subtraction, Donders had to make three important assumptions. The first was that the different stages of each task were arranged serially. That is, only one stage could be operating at any time, and each stage had to be completed before the next could begin (*see* Fig. 14.1, top panel). The second assumption is independence of stages. For example, however long the perceptual

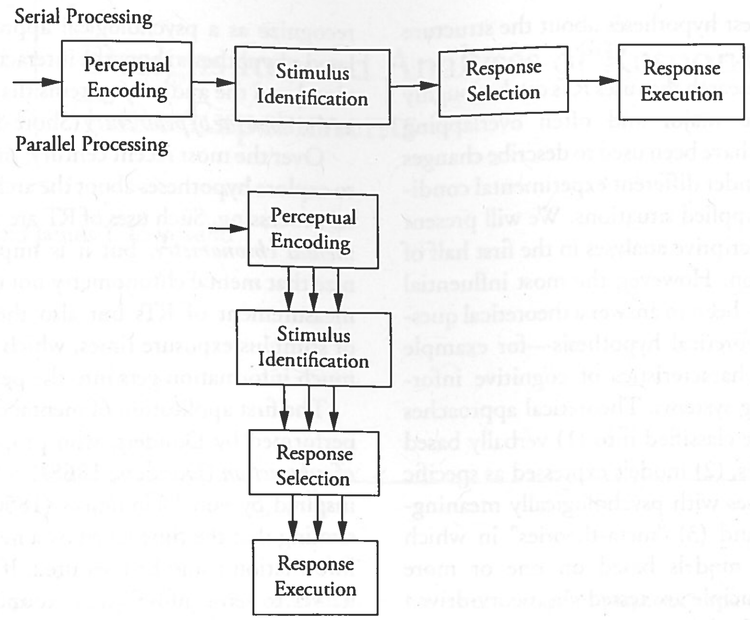

Figure 14.1 Serial processing (top panel) and parallel processing (bottom panel) of stages in Donders' choice reaction task.

encoding stage, the response selection stage will not be affected. The third assumption was that the stages in each task did not change as other stages were added. So, for example, the time required for response execution was the same whether or not there was a stage of response selection. This third assumption is sometimes called *pure insertion.* Given these three assumptions, because the three tasks differ only by a single processing stage, the time required for stimulus identification should equal the go/no-go RT minus the simple RT. Similarly, the time for response selection should equal the choice RT minus the go/no-go RT.

Donders' method of subtraction is the earliest example of a meta-theoretical approach in experimental psychology, and many variants of it are still in use today. The method of subtraction provides a statement about an entire class of models: models with serial processing stages, independence of those processing stages, and invariance of processing with the addition of stages (pure insertion). Furthermore, the assumption that a mental process could be added or subtracted by the experimenter led to many of the experimental designs discussed in the next section.

Very little was done with Donders' idea until the 1960s when Sternberg published two very significant papers (Sternberg, 1966, 1969). The most influential article, which appeared roughly 100 years after Donders' historic studies, introduced the *additive factors method* (Sternberg, 1969). This method, like Donders', proposed that cognitive architecture could be examined by looking at the difference between RTs in different experimental conditions. In particular, the method required that the experimenter identify experimental factors that *selectively* influenced different stages of processing. For example, consider a short-term memory search task in which observers are required to determine whether a target stimulus (e.g., a numeral) is one of a previously memorized set of stimuli called the *search set.*

The plot of mean RT as a function of search set size is called the search set function. The search set function is usually a linear increasing function of search set size, as if the addition of each search set member increases the number of comparisons between the search set and the target. Sternberg (1966) therefore proposed (not on the basis of the additive factors method) that this task is accomplished by a serial process in which observers compare the target to each member of the memorized search set one at a time.

Sternberg (1969) applied the additive factors method to examine all the stages of processing in the memory search task. He reasoned that the overall task could be broken into at least two serial processing stages. The first was a perceptual encoding stage and the second was the search stage in which the target was compared to each of the stimuli held in memory. His idea, like Donders', was to prolong each of these stages and examine the increases in mean RT. Making the target difficult to see would prolong the encoding stage but (he assumed) not influence the search stage. Similarly, increasing the memory load, the size of the search set, would prolong the search stage but not influence the encoding stage.

Sternberg (1969) asked his subjects to make memory decisions in a factorial design that varied both memory load and stimulus visibility. He then implemented the additive factors method by plotting the search set functions separately for each stimulus visibility condition. The search set functions under different visibility conditions were parallel; there was no interaction between visibility and memory load. That is, the effects of visibility and search set size were additive: The extent to which RT was prolonged by poor visibility was the same regardless of the size of the search set. Similarly, the extent to which RT was prolonged by increasing the size of the search set was the same regardless of the visibility condition. This noninteraction, this additive effect of the two factors (and not the linearity of the mean RTs) supported the notion of serially organized and selectively influenced stages of processing: encoding, followed by search.

The additive factors method had a huge influence across the field of experimental psychology. Later work generalized the additive factors method from the simple serial structures of Donders and Sternberg to other kinds of cognitive architectures (e.g., Schweickert, 1978; Schweickert & Townsend, 1989; Townsend, 1984; Townsend & Wenger, 2004b). One alternative architecture is parallel processing, where all stages operate simultaneously (*see* Fig. 14.1, bottom panel). Parallel processing has always been considered the antithesis of serial processing and, along with serial processing, has been the most investigated.

Most applications of additive factors logic use measurements of mean RT. More modern treatments of additive factore, including *factorial* methodologies, make use of the detailed RT probability distributions (Ashby & Townsend, 1980; Balakrishnan, 1994; Roberts & Sternberg, 1992; Townsend & Nozawa, 1995; Schweickert, Giorgini, & Dzhafarov, 2000) Factorial approaches are meta-theoretic in that they attempt to rule out entire classes of models with a set of data—for example, all parallel models for a certain kind of task. Another line of RT modeling is less focused on questions of mental architecture but, rather, on theoretically motivated models of the cognitive system. These models, called *sequential sampling models*, can also predict the entire distribution of RTs, as well as the accuracy of responding (e.g., Ratcliff & Smith, 2004).

Many modern experiments are designed to test predictions generated by the sequential sampling models. Thus, there are many aspects of experimental design and RT analysis that have been derived from consideration of this sort of data-generating mechanism. Other work has followed the additive factors tradition of testing more general cognitive architectures without being very specific about the kinds of mechanisms that might give rise to different processing stage durations. In the rest of this chapter we will discuss both approaches, emphasizing how RT experiments are designed from both theoretical perspectives and touching briefly on how RTs from such experiments are analyzed.

Design

The design of an RT experiment can be classified according to the degree of *information compression* between the number of possible stimuli that can be presented and the number of possible responses that can be made. Simple response tasks, which have only a single response for a potentially very large number of stimuli, have the highest degree of information compression. Identification tasks, which have *N* different responses for each of *N* different stimuli, have no information compression. In this section we will discuss the different kinds of RT experiments, outlining for each the major variables that influence RT and potentially confounding variables.

There are some variables that influence RT that we will not discuss. These include level of arousal or fatigue, extent of practice, gender, handedness, intelligence, the effect of drugs, and presentation to different brain hemispheres. The interested reader may consult Welford and Brebner (1980) for older but still accurate and comprehensive reviews of some of these additional variables.

Simple Reaction Tasks

Simple RT designs are characterized by having only a single response option, although many different stimuli may be presented. For example, in Donders' simple RT experiment, there were two stimuli: a red light and a green light. However, there was only one response, which was to press a key when a light had appeared, regardless of its color. Simple RT tasks are sometimes called detection tasks, because the observer's job is simply to detect the presence of a stimulus no matter what it is. A number of variables influence simple RT, most importantly the stimulus modality (the sensory system that encodes the stimulus), the intensity of the stimuli, and the temporal structure of the trials.

STIMULUS MODALITY AND INTENSITY

Stimuli presented auditorily elicit significantly faster responses than stimuli presented visually (Woodworth & Schlosberg, 1954, p. 16), but this difference decreases as the intensities (detectabilities) of the stimuli increase (Kohfeld, 1971). The difference attenuates because simple RT decreases rapidly overall as intensity increases, attaining a minimum simple RT for both visual and auditory stimuli somewhere between 150 and 200 milliseconds. This decrease is so reliable that, for intensity defined on a physical scale (e.g., amplitude of a tone), it can be captured by a relationship known as *Piéron's law* (Piéron, 1920). Piéron's law states that mean RT is equal to $a + bI^{-c}$, where a, b, and c are parameters, all greater than zero, to be estimated from the data.

We can think of simple RT as being influenced more generally by stimulus energy. Energy is computed as the intensity I of the stimulus multiplied by its duration t. Most studies have found that as the energy in a display increases, mean RT decreases (Teichner & Krebs, 1972; Ueno, 1978). For visual signals of very short duration ($t < 20$ *ms*, approximately), mean RT is approximately equal to $a+b(It)^{-c}$, but for longer durations, Piéron's law holds (Mansfield, 1973). The point to remember is that for a range of stimulus durations, intensity I can be traded for increases in duration t (or vice versa) to produce the same effect on RTs.

If the stimulus is presented for a fixed duration (say, 50 ms), the total amount of energy presented to the observer is also fixed (at $50I$). However, if the stimulus remains on until a response, the amount of energy continues to increase over time until the response is executed. As energy increases, the observer will eventually be able to see the stimulus, a process referred to as "summation," which is closely related to the evidence accumulation models we will discuss below. Thus response-terminated stimuli introduce a confound into the design: longer RTs mean that some stimuli have been presented for longer durations. Presentations with longer RTs have higher energy displays, so stimulus energy is no longer a controlled, independent variable. This may place important restrictions on the kinds of conclusions that can be drawn from the data.

Constant energy displays introduce a not-insignificant problem in the design of simple RT experiments. Consider, for example, what could happen when a low-intensity stimulus is presented for a fixed duration. This low-energy display may be undetectable by an observer, and so he will not make a response (if he is performing the task correctly). If the experimenter has not designed the experimental trials in light of this possibility, then the experiment will stop at this point: The observer will wait indefinitely for a stimulus that has already been presented, and the next experimental trial can't begin until he responds that he has seen the stimulus that he couldn't see.

For this reason, simple RT designs may use one of several strategies to ensure the experiment will continue. In addition to using stimuli that are response-terminated, this includes using a stimulus stream that continues even if a response is not made, presenting stimuli at fixed, predictable points in time, and/or using warning signals. All these possibilities are considerations for the temporal structure of the simple RT task.

TEMPORAL STRUCTURE

A stimulus stream that continues even in the absence of a response will be either random, with stimuli occurring at unpredictable times, or nonrandom. A nonrandom stream presents signals at the end of fixed time intervals, such as every 500 milliseconds or every 3 seconds. The difference between the onset of a signal and the onset of the next signal is sometimes called the *stimulus onset asynchrony*. A random stream uses stimulus onset asynchronies drawn at random on each trial.

A nonrandom stream implicitly informs the observer about when stimuli are presented by creating a rhythmic context for the task. Such rhythms create a temporal expectation about when the next signal will be presented (Large & Jones, 1999). There are concerns, however, that this fundamentally changes the nature of the task from one of detection to one of timing, in which observers tend to respond by rhythmic tapping. To prevent this, researchers can introduce "catch trials" on which no signal is presented. The number of times observers respond on catch trials (the number of anticipations or false alarms) can be used as an indication of the extent to which they are timing their responses rather than responding to a detected signal. Even with catch trials, however, RTs are susceptible to the rate at which rhythmic stimuli are presented (Van Zandt & Jones, 2012), which means that nonrandom streams may confound the influence of other variables on simple RT. Because of concerns like this, most simple RT designs use random streams.

A random stream with stimuli of fixed duration is called a *vigilance* task. In a vigilance task, one important dependent variable is the number of misses the observer makes as a function of the amount of time he or she has been performing the task. If the stimulus onset asynchronies are quite large, resulting in "rare" stimulus events, a vigilance task can be quite tiring. The number of misses an observer makes will increase as the task duration increases, an effect called the *vigilance decrement*. For shorter stimulus onset asynchronies, the vigilance decrement is not as severe, presumably because the higher event rate results in a higher level of arousal in the observer.

The distribution of the stimulus onset asynchrony also affects RT. The most common methods of selecting the stimulus onset asynchrony are to select at random from a small set of durations or to generate a random duration from a uniform or exponential distribution. The reason why the choice of distribution is important is that the length of the stimulus onset asynchrony can provide information about when the signal will occur. For example, if the stimulus onset asynchrony is selected at random from any distribution on a fixed interval (e.g., from 0 to 1000 ms) if the observer has waited for 800 milliseconds, then she knows the stimulus must appear within the next 200 milliseconds. This will result in a decrease in RTs to longer stimulus onset asynchronies, a decrease that will be especially pronounced if only a few possible stimulus onset asynchronies are used (e.g., 200 ms, 400 ms, 600 ms, 800 ms, and 1000 ms; Klemmer, 1956).

To eliminate this problem, which is usually attributed to increased response preparation or increased attention as uncertainty about target presentation time decreases, some researchers have used stimulus onset asynchronies drawn from an exponential distribution (e.g., Green & Luce, 1971). The exponential distribution has the peculiar statistical property that, regardless of the amount of time the observer has waited for a signal, the likelihood that the signal will appear in the next instant is constant. There is no way, then, to predict the onset of the signal from the amount of time that has elapsed. This kind of structure eliminates the problem of varying RTs caused by stimulus timing or anticipation, although RT still varies with the length of the stimulus onset ansynchrony (e.g., Green & Luce, 1971). One drawback is that the exponential distribution introduces the likelihood that some trials can be delayed by (rare) very long stimulus onset asynchronies.

WARNING SIGNALS

Perhaps the most popular way of informing the subject that a trial has ended or begun is to use a separate, easily-detectable warning signal to which a response is not required. Warning signals have at least two benefits: first, they provide a salient point at which the trial begins, and second, they provide a way to identify anticipatory responses, which are not made in response to any signal. In a vigilance task, any response could be to an earlier signal or a false alarm to a signal that the observer thought he saw. It is impossible, then, to determine what kind of response it is. With warning signals, any response made during the time between the warning signal and the target signal (i.e., the foreperiod) is an anticipatory response. Therefore, this procedure eliminates the need for catch trials.

The same issues arising with stimulus onset asynchrony arise again with foreperiod durations. However, there is a large literature on foreperiod designs concentrated on the effects of attention in motor learning. Like stimulus onset asynchrony, foreperiods can either be fixed or random. For fixed foreperiods, RT increases as foreperiod increases. For random foreperiods, the reverse is true. A number of explanations have been proposed for this strange pattern of effects, and the most likely seems to involve uncertainty (*see* Ellis & Jones, 2010; Niemi & Näätänen, 1981, for reviews). As we will discuss later, increased uncertainty, whether about what is going to happen or when it is going to happen, will increase RT. For the fixed foreperiod design, longer foreperiods lead to more uncertainty about when the signal will be presented, perhaps because longer intervals are more difficult to estimate, and this greater uncertainty results in longer RTs for the longer foreperiods. For the random foreperiod designs, there is also uncertainty about what foreperiod will be presented. However, as the foreperiod increases, this uncertainty decreases, resulting in faster RTs to the longer foreperiods.

An alternative to presenting a warning signal is to allow the observer to initiate the beginning of the trial with a keypress. This is referred to as a self-paced design. In a self-paced design, the foreperiod is measured from the keypress (the observer's signal to begin) to the onset of the signal. The rate of stimulus presentations in a self-paced design is, of course, determined by the observer, and so there is the risk that some observers will pace themselves very slowly, a pace that will likely be associated with slower RTs. Also, the experimenter has less control over intra trial variables, such as stimulus order, which may or may not be important.

One last important issue remains, and that concerns the response-stimulus interval, which is important regardless of whether a warning signal is used. The response-stimulus interval is the time between the observer's response and the next stimulus presented (which may be either a warning signal or the next target signal). It is difficult to simultaneously control both stimulus-onset asynchrony and the response-stimulus interval. Most designs control only the response-stimulus interval. In studies without warning signals, the response-stimulus interval is considered equivalent to a foreperiod, and indeed, the same general effects on RT are observed for increasing and decreasing response-stimulus intervals. If the response-stimulus interval is fixed, observers may use the resulting predictability of the stimulus onset to time their responses. Similarly, increasing the response-stimulus interval may result in increased temporal uncertainty, which can produce slower RTs. Conversely, increasing the response-stimulus interval through a limited range of interval durations can decrease temporal uncertainty for longer response-stimulus intervals, which may speed RTs.

IS THE SIMPLE RT TASK TOO SIMPLE TO BE INTERESTING?

In many ways, the simple RT task serves as a point of connection between work in psychophysics, which focuses on lower-level perceptual mechanisms, and work in simple choice, which we discuss in the next section. Whereas psychophysical experiments usually concentrate on changes in accuracy with changes in stimulus conditions, choice RT experiments are frequently concerned with simultaneous changes in accuracy and RT. The RTs measured in a simple RT task vary with the same variables that influence accuracy in psychophysical tasks, and many of the variables influencing RT in the simple RT task also influence RT in the choice RT task. Smith (1995) has provided an excellent review of the link between these different areas as well as a model of the simple RT tasks that explains many of the effects we have presented in this section.

Although the simple RT design is one of the, well, simplest kinds of RT experiment, it is widely used to study highly complex perceptual phenomena. We have barely skimmed the surface of this literature in this brief review. For example, we have focused in this section primarily on data from keypress responses, but in fact almost any overt motor action can be the basis of a simple RT. This includes, for example, eye movements, measured with eye-tracking equipment, or vocal responses (e.g., Diederich & Colonius, 2008). The stimuli can be very complex, including words, pictures, or a combination of sensory modalities. Of course, each of these stimulus types will influence overall RT.

Choice Reaction Tasks

If the number of stimuli presented is greater than the number of responses permitted, then the task is either a go/no-go task or an n-choice task, where n is the number of possible responses.

Considering first the n-choice task, we can conceive of the cognitive process as one of classifying the signals into one of n possible categories. In many RT experiments, $n = 2$ and the observer is asked to determine, for possibly very many different stimuli, whether the signal is an "A"-type or a "B"-type. For example, given a burst of white noise in which a pure tone may or may not be embedded, an observer may be asked to say whether a signal is present or whether it is pure noise—a signal detection task. Given an object like a letter, numeral, word, or picture, the observer may be asked to determine whether the object was encountered previously in the experiment ("old") or not ("new")—a recognition task.

There are also n-choice tasks, which can arise in studies of categorization. A subject may be presented with a stimulus defined as a location in r-dimensional space, where r is the number of unique and not-necessarily-orthogonal features of the stimulus. A geometric shape, for example, could be defined by the number of its sides and its convexity, size and, color. "A"-shapes may tend to be pinkish, have small numbers of sides, and be large and convex. "B"-shapes may be similar but tend to be greenish and are not always convex. "C"-shapes are pinkish but small and have more sides, and so forth. The observer's job, given a stimulus, is to say whether it is of type "A," "B," or "C."

A go/no-go task has at least one (but possibly more) more stimulus than responses, and that one stimulus is the one that requires withholding a response. The simple RT task with catch trials can be considered a go/no-go task, if one defines the absence of a signal as being a different sort of signal. We will discuss the go/no-go task in more detail in a later section.

There are several important design considerations in constructing a choice task. Many of the issues that arise in the design of a simple RT task will still be important, such as the use of warning signals, fixed versus variable foreperiods, and the response-stimulus interval. In addition, we must consider the facts that RT is going to increase as both the number of stimuli and responses increases and that RT is correlated with response accuracy.

TRANSMITTED INFORMATION

Response time is a linear function of the amount of uncertainty in the task (Hick, 1952; Hyman, 1953). This effect is so robust that it is referred to as the *Hick-Hyman Law* of mean RT. We mentioned uncertainty in somewhat vague terms earlier in our discussion of the role that temporal structures play in simple RT task performance. Now we formalize this idea.

Uncertainty is a dimensionless quantity that depends on the number of possible outcomes in an experiment and their probability. It can be used to describe many things, but in this context it refers to the amount of information provided by the occurrence of an event. For example, if there is only one possible response, observing that response does not change the amount of information about what response the observer is going to make. However, if there are n equally likely responses, observing one of them changes the amount of information about the observer a lot.

In choice RT, the event for which we measure uncertainty is a particular stimulus–response combination. Suppose that a set of stimuli $\{S_1, S_2, \ldots, S_m\}$ may be presented, and to each the observer can select one response from a set of responses $\{R_1, R_2, \ldots, R_n\}$, where $n \le m$.[1] Let the probability that stimulus S_i is presented be p_i, and let the probability that response R_j is made be q_j. Also let the probability that response R_j is made to stimulus S_i be r_{ij}, which will equal $p_i q_j$ only when the responses are independent from the stimuli presented.

We define the amount of stimulus information to be

$$H(S) = -\sum_{i=1}^{m} p_i \log p_i,$$

where log is taken to the base 2. Similarly,

$$H(R) = -\sum_{i=1}^{n} q_i \log q_i$$

is the amount of response information. Information is measured in bits, which is the fewest number of binary questions that would be required to uniquely identify the event that occurred. Joint information is measured over the collection of stimulus–response pairs. It is given by

$$H(S,R) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} \log r_{ij}.$$

For any set of n events $\{X_1, X_2, \ldots, X_n\}$ that occur with probabilities $\{p_1, p_2, \ldots, p_n\}$, if $p_i = 1/n$ then the amount of information in the set is $\log n$, which is also the maximum amount of information possible.

Transmitted information determines the speed of responding. Transmitted information is given by

$$T(S,R) = H(S) + H(R) - H(S,R).$$

Note that the information measure does not depend on how accurate the observer is but only on how consistent he is. Transmitted information is at the highest possible level when $r_{ij} = 1$ for some $i = k \in [1, n]$ and 0 for all other $i \neq k$. Transmitted information is 0 when $r_{ij} = p_i q_j$—when the response is statistically independent from the stimulus.

Hick (1952) and Hyman (1953) both showed that

$$E[RT] = a + bT(S,R):$$

mean RT is a linear function of transmitted information. The coefficient b is called the channel capacity of the observer, and it reflects how quickly information is processed (in time units per bit). The importance of this law for the design of reaction time experiments is that as the number of possible stimulus–response combinations increases, mean RT will also increase.

As an historical aside, recall our earlier discussion of Sternberg's (1966) paper in which he examined mean RT in a memory-search task. Observers were shown a search set of digits that they had to remember and then were shown a target digit. They were asked to determine whether the target digit was present in the search set. Sternberg showed that the mean RT to respond "yes" or "no" increased linearly as a function of search set size. According to the Hick-Hyman Law, we would have expected such an increase in mean RT only because of the change in the amount of information transmitted as the size of the search set increased. However, one underappreciated feature of Sternberg's design is that he very carefully matched search set sizes with stimulus probabilities to keep the amount of transmitted information constant across changes in set size. Thus, the effect he observed had to result from changes in set size alone.

Information theory and the Hick-Hyman Law do not provide a good theory of how choice RTs are generated. More recent work by Usher and McClelland (2001) has suggested that Hick-Hyman Law behavior could be produced by a model of stochastic information accumulation. According to this sort of model, the increase in RT with increases in information transmitted arises from observers' increasing the amount of information necessary for a response to avoid making errors, the speed–accuracy tradeoff.

SPEED–ACCURACY TRADEOFF

One difficulty in RT experiments is the fact that accuracy is correlated with RT: The faster an observer responds, the more errors she makes. Earlier studies such as those of Sternberg (1966, 1969) and Donders (1868/1969) assumed (implicitly or explicitly) that if the error rate were small enough, then error responses could be safely ignored. Although in general this is true, it is also true that very small changes in error rate may reflect a very large change in processing strategy and hence RT.

It is straightforward but somewhat tedious to try and keep subjects at a constant accuracy so that their RTs can be measured at a single point on the speed–accuracy tradeoff curve (e.g., Santee & Egeth, 1982). This kind of design requires trial-by-trial manipulation of an independent variable that can control accuracy, such as stimulus presentation time or contrast. Using psychophysical procedures (adaptive staircase methods; Garcia-Perez, 1998), the independent variable is adjusted upward or downward on each trial depending on the accuracy of the previous response. Such procedures may indeed control for the speed–accuracy tradeoff but do not provide any explanation for it.

Explanations of the speed–accuracy tradeoff in RT experiments are provided by sequential sampling models, discussed briefly above. These models, perhaps the most successful models of RT, produce the speed–accuracy tradeoff naturally as subjects raise and lower the amounts of information (thresholds) necessary to select a particular response on each trial (*see* Fig. 14.2). If thresholds are low then less evidence will be required and therefore less time will elapse before a threshold is reached. However, it will be easier for an inappropriate response to accumulate enough evidence to reach a lower threshold. If the thresholds are higher, then RTs will be slower, but

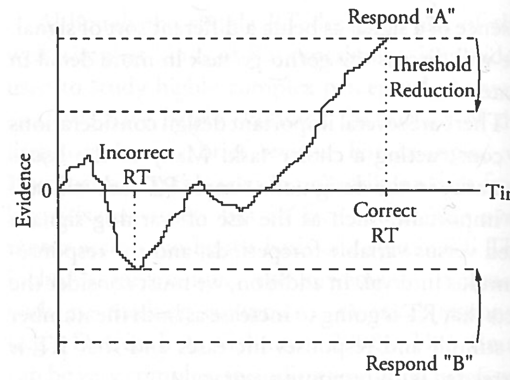

Figure 14.2 The speed-accuracy tradeoff in a sequential sampling model. At the presentation of target *A* at time 0, there is 0 evidence toward either of the two responses (*A* or *B*). Information accumulates randomly over time, reaching one of the two thresholds to determine the response. If the thresholds are reduced, then spurious evidence toward response *B* results in an incorrect decision.

inappropriate responses will be less likely to reach the threshold. Thus, faster RTs will be associated with lower accuracy levels and slower RTs will be associated with higher accuracy levels.

The sequential sampling models make predictions about the state of the cognitive processing system over time. Although this system produces as output an RT and a response, researchers have tried to peer inside the system to verify these predictions. Of course, researchers can't watch the process unfold, but if they assume that the accumulation process is not influenced by where the thresholds are placed, then they can try and move those thresholds up and down and look for predicted changes in RT and response probability. This desire to look into the heart of the information accumulation process led to the development of deadline and response-signal designs.

Deadline and response-signal experiments attempt to tell people what their RTs should be. Simple deadlines tell subjects "Too slow" (or "Too fast") after the response and could potentially penalize them in some way by taking away points or repeating the trial later in the session (e.g., Pachella & Pew, 1968). More severe deadlines can time-out the trial if the subject hasn't responded. By contrast, response-signal experiments ask subjects to make their responses when they see a signal presented after the target stimulus (e.g., Reed, 1973). Some response signals are presented with a very short foreperiod, and others may be quite long.

Researchers assume that under deadlines subjects move their thresholds down or up to permit faster or slower responding (*see* Fig. 14.2). Deadlines are usually fixed from trial to trial so that subjects can move their thresholds to reliably produce the desired RT. However, under the response-signal paradigm, the thresholds are irrelevant and the response will be based on the level of information accumulated at the time of the response signal.

The problem with the deadline design is that the RTs produced may not follow the shape predicted by the model, because they are truncated at the deadline boundaries, distorting their shape (e.g., Van Zandt, Colonius, & Proctor, 2000). However, if the purpose of the experiment doesn't depend on distribution shape, then deadlines are an excellent way to produce a high proportion of RTs within a particular time window. The dependent variables are choice accuracy and the number of responses executed on time.

The response-signal design takes the thresholds out of the response process. Because the assumption is that the response will be made based on the amount of evidence accumulated at the point in time at which the response signal appears, the dependent variable is then the change in accuracy as a function of RT (or response signal time).

Number of Stimuli Equal to the Number of Responses

Designs in which each stimulus presented requires its own unique response are sometimes called identification or absolute identification tasks. For example, four-letter stimuli ("A," "B," "C," and "D") may be presented one at a time, and the subject may be asked to press one of four buttons numbered 1, 2, 3, and 4 to each. Response data from this kind of experiment may be arranged in a *confusion matrix*, which indicates the fidelity of the assigned responses to their stimuli. Response times in identification tasks are influenced by the same variables that influence simple and choice RT, such as information transmitted and the speed–accuracy tradoff.

One important limitation in identification performance was identified by George Miller in his famous (1956) paper, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." Miller argued, after reviewing a wide range of literature, that people could efficiently transmit around 2.5 bits of information (approximately 7 equiprobable stimuli) but not much more without decrements in performance. This means that RT will increase with increased number of stimuli to be identified and that these increases in RT will occur with a concurrent decrease in accuracy.

Another important factor in identification RT is stimulus–response compatibility. Stimulus–response compatibility is a term that describes the extent to which features of the stimulus set (which may or may not be relevant to the responses required to them) overlap or are similar to features of the response set. Although stimulus–response compatibility may influence RTs in any choice-RT design, it is most commonly studied using identification tasks.

Compatibility experiments have often focused on the spatial features of stimuli and responses, or where the stimuli appear relative to the location of the responses to be made to them. Highly compatible spatial relationships (e.g., responding with a right button to the stimulus that appears on the right) result in faster RTs than less compatible spatial relationships. Early experiments by Fitts and his colleagues (e.g., Fitts & Seeger, 1953) demonstrated that when stimuli were characterized by different spatial configurations of lights, RTs were fastest when the responses matched those spatial configurations, even when the number of stimuli and responses are the same (holding information constant).

Compatibility effects can also occur when the spatial stimulus dimension is task-irrelevant. For example, the Simon effect occurs when two stimuli, say red and green lights, are mapped to two different responses, say a left or right button-press (Simon, 1969). Assume that the observer is to respond with a left button-press to the red light and a right button-press to the green light. Response times will be faster if the red light appears to the left of center than if the red light appears to the right of center. Compatibility effects even arise for nonspatial stimulus dimensions such as positive–negative affect of stimuli and verbal responses. (*See* Proctor & Vu, 2006, for a thorough review of this and other compatibility effects.)

Stop Signal, Dual-Task, and Task-Switching Designs

Donders' go/no-go task can be viewed as a choice-RT task where one of the possible responses is not to respond at all. A paradigm closely related to the go/no-go task is the stop-signal task. A stop-signal task is a choice-RT tasks where, for some trials, a stop signal is presented at some time (usually hundreds of milliseconds' delay) after the stimulus, indicating that the observer should inhibit the response to the stimulus. Thus, the go/no-go task is a stop-signal task where one of the stimuli is the stop signal, and the stop signal is always presented with 0 milliseconds delay. Stop-signal tasks are used to explore the dynamics of response preparation. As the stop-signal delay increases, observers are less able to inhibit their responses: The function relating the probability of successfully inhibiting a response to stop-signal delay is a smooth, S-shaped curve. It is not a step-function, where there is a delay below which the responses are never inhibited but above which they are always inhibited. This suggests that the choice process has components that are gradual and build up over time (e.g., Logan, Cowan, & Davis, 1984).

Changes in mean RT with changes in stop-signal delay are consistent with the idea that the choice process unfolds over time and also that the stimulus is processed at the same time as the stop signal. As the stop-signal delay decreases, RTs on stop-signal trials (responses that the observer failed to inhibit) become faster. This decrease in RT may arise from a "race" between the processing of the stimulus and the processing of a stop signal. For stop signals that appear very close to the stimulus onset, a response will be made only on those trials where stimulus processing was fast enough to beat the processing of the stop signal, resulting in mean RTs that increase with increasing delay.

Research using the stop-signal task tries to answer questions about "executive control" of action. That is, how do people start and stop their behaviors at appropriate times? In these kinds of problems, we have to appreciate that people may be performing many tasks at once. The stop-signal task is a relatively simple task in which observers do two things at the same time: select a response to a stimulus and also prepare to inhibit that response. Another kind of task, the dual task, asks people to make two (possibly different) responses to two (possibly different) stimuli at the same time.

The dual-task design, like the stop-signal paradigm, varies the delay between the onsets of the first and second stimuli. This design was used in some early studies of attention (Telford, 1931; Welford, 1952). These studies demonstrated that RT to the second stimulus decreased as the delay between stimulus onsets increased, suggesting that there was only one "channel" through which the stimuli could be processed, and that this channel could only accomodate one stimulus at a time. This interference between the processing of the first and second stimulus is sometimes called the psychological refractory period, and it is interesting that there is apparently no such interference for the very similar stop-signal task.

Another related paradigm is the task-switching paradigm. In task switching, people are asked to do different things on different trials, and their performance on "switch" trials, in which the task changes from the previous trial, is then compared to their performance on "repetition" trials, in which the task does not change. Typically RTs show a "switch cost": RTs are slower after a switch to a new task than when the task is repeated (Logan & Gordon, 2001). These switch costs are used to measure the time required by executive processes to move between different tasks.

The similarity between the stop-signal, dual-task, and task-switching paradigms was explored by Logan and Burkell (1986). They used two stimuli, a letter followed by a tone. The letter required one response whereas the tone required another. In stop-signal conditions, the tone was the stop signal indicating that the response to the letter should be inhibited. In dual-task conditions, both the letter and the tone required responses. In the task-switching conditions, the tone was both a stop signal to the letter and required its own response, so observers had to stop processing the letter and switch to the tone. The critical comparison was between RTs to the tone when the letter response had been either inhibited or uninhibited.

The switch costs for trials in which the tone failed to inhibit the response to the letter were similar in magnitude to the interference in the dual-task conditions. There was little or no switch cost for trials in which the tone successfully inhibited the response to the letter, although the extent of interference should have been approximately the same as in the dual-task conditions. Logan and Burkell (1986) argued that this finding supports the idea that the difference between RTs in stop-signal and dual-task paradigms results from interference between responses and not competition for processing resources.

The stop-signal, dual-task and task-switching paradigms have been used to explore mechanisms of response inhibition and automaticity of processing, and more generally to understand executive processing, or how people are able to control their behaviors, and also the factors that contribute to uncontrollable (automatic) behaviors. Interested readers should consult Verbruggen and Logan (2009) for a recent review of the literature in this area.

Analysis of Response Time Data

Response time data, regardless of the experiment in which they were collected, have the following characteristics. First, the data may be assumed to be a mixture from the process under study and a number of contaminant or outlier observations arising from equipment failures, attentional lapses, subject perversity, and so forth. Second, RTs are samples from positively skewed distributions; the longest RTs may span a very wide range, whereas the shortest RTs are usually concentrated around some modal value. Third, RTs are sequential data and are not, generally speaking, independent from trial to trial. Finally, RTs are subject to a wide range of individual differences, so collapsing data across subjects is not usually a good idea. Analysis of RTs, if done well, takes all these characteristics into account. Unfortunately, there are not many canned procedures that have the ability to accommodate mixtures, skewness, sequential dependencies and individual differences simultaneously, so most analyses compromise on one or more of these issues.

The most common approach to the analysis of RT data is to compute the mean RT for each subject's responses in each experimental condition. For example, an experiment might ask subjects to make choice responses in four conditions. They may be asked to make their responses as quickly as possible, sacrificing accuracy if necessary, or they may be asked go as slowly as necessary to to maintain a high level of accuracy. Within each of these conditions, subjects may see both high-energy and low-energy stimuli. After completing some number of trials in each condition, each subject will have four mean RTs (fast or accurate responding by high or low energy). These means would then be subjected to a repeated-measures factorial analysis of variance to test for effects of instructions and stimulus energy.

There are a number of unsatisfactory features of this approach. First, the analysis of variance assumes a model in which the relationship between mean RT and the effect of the independent variables is linear. There is no theoretical basis for this assumption. Second, compressing every subject's data into means discards a great deal of information that may be useful for determining how the RTs were generated, such as skewness and sequential effects, which is the ultimate goal of experimentation. Third, the assumptions required for analysis of variance, such as normally distributed residuals, independence of observations, and homogeneity of variance across conditions, are routinely violated in RT data.

In this section we present a number of methods for analyzing RT data, including the use of mean RT, estimating the parameters of models for RT data, and using RT for testing cognitive architecture. For a more thorough treatment of RT analysis, interested readers can consult Van Zandt (2002). For a more general treatment of modeling and parameter estimation issues, interested readers can consult Busemeyer and Diederich (2010).

Analyses of Mean Response Time

Many hypotheses about cognitive processing are formulated to provide predictions about mean RT. For example, the additive factors method looks for interactive effects of variables on mean RT. The typical procedure involves fitting the general linear model (most commonly the analysis of variance model) to the mean RTs computed for each subject and condition and relying on variance accounted for to argue for effects of different independent variables on performance.

The general linear model is unsatisfactory as an inferential instrument for RTs. The assumptions necessary for application of the general linear model are generally not met in RT data, even in mean RT data. These assumptions include normal or symmetric distributions, independence of observations, and homogeneity of variance across conditions.

To understand why the assumptions of normality and symmetry are violated, consider how the distributions of RTs from individual subjects are distributed. Response time distributions are positively skewed and hence asymmetric, and RTs are highly correlated across trials and conditions, showing evidence of autocorrelation structure and dependence on previous stimuli and responses. The degree of asymmetry and autocorrelation varies across subjects—that is, each subject's RTs come from a different distribution. Therefore, although the mean RT from a single subject may be approximately normally distributed via the Central Limit Theorem, the mean RTs across different subjects come from different normal distributions with different means and variances. This means that the mean RTs, the dependent variables, are drawn from a mixture of normal distributions, which is unlikely to be normally distributed itself and probably not symmetric.

Homogeneity of variance is violated in mean RT data not only because individual subjects have different mean RT distributions but also because mean RT and RT variance are correlated such that as the mean increases so does the variance. The coefficient of variation (the standard deviation divided by the mean) of RT data is approximately constant (e.g., Luce, 1986, p. 64), implying that the standard deviation is a nearly linear function of the mean. This fact provides strong evidence for the kinds of distributions that best describe RT data and hence the classes of models that best explain performance in RT tasks (Wagenmakers & Brown, 2007). That is, we should not choose to model RT using, say, a normal distribution, because the variance of the normal distribution does not increase as its mean increases. However, the variance of the gamma distribution increases linearly with its mean and has a constant coefficient of variation. Therefore, the gamma distribution is a better choice for a model of RT data.

Finally, the linear model itself, which is the basis of procedures like regression and the analysis of variance, relates mean RTs to a linear function of the independent variables. This linear relationship is not an accurate representation of the influences of the independent variables on RTs, which are generated by a highly nonlinear dynamic system.

Apart from the marginal benefits of a linear modeling approach, there are other issues that arise when collapsing across observations to compute a summary statistic for RT. The most perennial of these problems arises from the skewness of RT data. The large upper tail in the RT distribution has the effect of creating "outliers," RTs that are much longer than the bulk of the RTs observed for an individual. Outliers are a problem in all areas of statistical analysis, but the unusual aspect of outliers in RT data is that they potentially derive from the process of interest. That is, they are not necessarily outliers in the sense of contaminations of the data. There is a problem, then, in deciding which observations are contaminants and which are not.

Every experimenter has a preferred method for cleaning their RT data, and these methods are usually based on personal preference rather than statistical necessity. One of the authors of this chapter (TVZ), for example, routinely discards choice RT observations faster than 200 milliseconds and greater than 3.5 standard deviations above the mean. Ratcliff (1993) performed an extensive Monte Carlo study of different methods of RT outlier treatment and their effects on inferential tests on the mean. Some of the methods he examined used cutoff values such as those of TVZ, as well as common data transformations such as the inverse and logarithm. For each of these methods, he computed power and the probability of Type I errors for analyses of variance under different levels of outlier contamination. He demonstrated that the choice of outlier treatment had no influence on the rate of Type I errors. However, the different methods had strong effects on power.

Cutoff methods that use standard deviations can reduce power. Fixed cutoffs that did not depend on sample statistics maintained the highest power. Unfortunately, a fixed cutoff is difficult to apply across all subjects and conditions in an experiment. A cutoff that seems appropriate for one condition (e.g., 5000 ms) might not be appropriate for another condition, especially because the usual purpose of the different conditions of an RT experiment is to observe increases or decreases in mean RT. In addition, slower subjects will have more RTs eliminated as outliers, which will have implications for evaluating mean differences over conditions and may lead to statistical artifacts such as truncation or floor and ceiling effects.

A statistical artifact that arises from cutoffs is estimation bias, which is the extent to which a statistic like the sample mean fails as an estimate of a population parameter. Ulrich and Miller (1994) showed that cutoffs can introduce bias into estimates of the mean, median, and higher moments of the RT distribution and that these effects could be larger than the experimental effects of interest. Van Selst and Jolicoeur (1994) also showed that this bias is influenced by sample size: Smaller sample sizes result in the elimination of fewer high RTs.

One way to avoid the problems associated with cutoffs is to use a data transformation. Both the log (log(RT)) and the inverse (1/RT) transformations have the effect of "squeezing" the distribution and reducing skew. Ratcliff (1993) showed that the inverse transformation had better power than the log transformation, almost as high as that of fixed cutoffs. One important benefit of data transformations is that they do not require the researcher to discard data, which is always risky if the researcher is not absolutely certain that an observation is a contaminant.

Another way to handle the outlier problem is not to use moment-based statistics like the mean and standard deviation at all. Rather, the researcher may turn to robust statistical methods that are based on the median and interquartile range (*see* Erceg-Hurn, Wilcox, & Keselman, Chapter 19, Volume 1). The median and interquartile range statistics are called robust because they change very little in the presence of outliers and skew. Their use is uncommon in RT analysis (and most other areas in psychology) because they are mathematically more difficult to work with and their standard errors are larger than those of their moment-based equivalents. The sample sizes required to attain approximate normality of their sampling distributions are much larger (Stuart & Ord, 1999).

Analysis of mean RT is therefore not as simple as it appears. However, there are a few rules of thumb that can be followed. First, the researcher must recognize that the linear model does not portray the data-generating mechanism accurately and consider using more sophisticated modeling schemes such as those presented below. Second, if the researcher decides that she must collapse across individual observations into a summary statistic, then she must pay close attention to outliers. If outliers seem to be a problem, then she can use a data transformation method instead of discarding data or use the median instead. Third, the researcher should perform the analysis in several ways (with and without the outlier treatment, or on both the means and medians) to make sure statistical artifacts have not been introduced. Finally, the researcher should be aware that collapsing across individual observations and using a linear modeling scheme may hide important effects in the data. If the effects of independent variables are strong, then the researcher may see them in the means and a regression may easily detect their presence. However, the exact nature of that effect may be obscured, as will the effect itself if it is at all subtle.

Time Series Analysis

An increasingly popular way to analyze RT data is to treat them as time series. A time series is a sequence of measurements with a time index, such as the level of the Dow Jones index at the end of every trading day (*see* Wei, Chapter 22, Volume 2). For RT data, although the measurements are of time, the index is the trial, which may or may not occur at fixed points in time, depending on the design of the experiment in which the RTs were collected. Despite this deviation from a true time series, treating RT data over a sequence of trials as a time series has a number of benefits.

Most important of these benefits is the fact that RTs are, as we noted earlier, correlated across trials. Not only do RTs vary as a function of the previous stimuli and responses (Kirby, 1980; Laming, 1968, 1979; Remington, 1969), but they are autocorrelated, usually positively, so that fast responses tend to follow fast responses and slow responses tend to follow slow responses (Wagenmakers, Farrell, & Ratcliff, 2004). Time series approaches attempt to model directly these correlations, although there are several pitfalls to doing so.

A general time series model for a measurement $\{T_1, T_2, \ldots, T_t\}$ at time t is a (possibly nonlinear) function of three things: the values of the measurement $\{T_1, T_2, \ldots, T_{t-1}\}$ up to time t, a trend component $\{\mu_1, \mu_2, \ldots, \mu_t\}$ that does not depend on any T_i, and a random noise process $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_t\}$. Perhaps the simplest, but a quite powerful, time series model is the autoregressive process of order 1 or AR(1) model, which is written

$$T_t = \phi T_{t-1} + \mu + \epsilon_t,$$

where the coefficient ϕ is a constant with absolute value less than 1 and the trend μ is constant across trials. For the AR(1) model, ϕ determines the extent to which the observation at time t is correlated with the observation at time $t-1$. The constant trend is the overall mean of the process, and ϵ_t is a white noise process, an independent sample from a normal distribution with mean 0 and variance σ^2.

It is the assumption of white noise that poses the first problem for time series analysis of RTs. Almost all applications of time series models in psychology, including autoregressive and moving average models, as well as integrated moving average models, assume white noise. To understand why this is problematic, consider the simple AR(1) model. Under the white noise assumption, the marginal distribution of measurements T should be Gaussian with mean μ and variance σ^2. However, RTs are not normally distributed. This means that using the AR(1) to estimate, for example, the magnitude of the autocorrelation coefficient for an RT series will not produce accurate results.

A second problem is how to identify trend and isolate it from the process generating the RTs. There are many reasons why mean RT might fluctuate over time. One commonly observed trend is a decrease in RT with practice, which occurs even for simple RT. This trend may be completely separate from the mechanism that produces the RTs or it may be an integral part of it. If trend is separate from the data-generating mechanism, then how can we accurately estimate and remove it so that we may estimate the other important features of the process? If trend is not properly removed, then it will distort the impression of autocorrelation. If trend is an integral part of the data-generating mechanism, then how do we explain it and how it contributes to the autocorrelation structure?

Much current interest in time series analysis of RT data has been spurred by work of Gilden (1997,

2001) and others (Holden, Van Orden, & Turvey, 2009; Kello, Anderson, Holden, & Van Orden, 2008), who have argued that RT variance shows evidence of "$1/f$ noise" or long-range dependence. Long-range dependence means that the RT on trial t is influenced not just by the RTs on trials $t-2$ and $t-1$ but by all the RTs up to that point $(RT_1, RT_2, \ldots, RT_{t-1})$. Long-range dependence is characteristic of a number of natural processes (such as heart rhythms) and is associated with system complexity and fractal structures. Fractal structures are usually formed by simple iterative processes that produce regular patterns at arbitrarily small scales of measurement (*see*, for example, the Mandelbrot set; "Mandelbrot Set," 2010). The recurrence of these patterns over different measurement scales is called self-similarity. Self-similar processes can exhibit long-range dependence. For RTs, this may imply scale invariance: RTs may behave the same way whether measured in milliseconds, seconds, minutes, and so forth. However, much of the work exploring long-range dependence uses techniques appropriate only to Gaussian processes and does not adequately deal with trend, which means that measurements of long-range dependence may be distorted.

Many demonstrations of long-range dependency have focused on the power spectrum of RT series. There are several nonparametric approaches to spectral density estimation that can be used to support the notion of long-range dependence and fractal structure in RT data. Holden (2005) advocates the use of nonparametric dispersion analysis together with classic parametric estimation methods. Dispersion analysis provides an estimate of fractal dimension of the series, which in turn can provide evidence of long-range dependence (Van Orden, Holden, & Turvey, 2003).

The problem of separating trend or experimental effects from dependence is a difficult one (e.g., Peruggia, Van Zandt, & Chen, 2002). Trend that has not been removed from the analyzed series will inflate the perception of long-range dependency. One simple way to detrend a series, which also permits the use of Gaussian process techniques, is to "normalize" the log RTs by passing them through the inverse normal cumulative probability function—that is convert the RT quantiles to normal scores. These scores can then be detrended using a number of different techniques and the detrended scores then passed back through the normal probability function and converted to the original RT scale. Craigmile, Peruggia, and Van Zandt (2010b) have showed how this procedure can quite accurately recover even very complicated patterns of trend.

Another more complex way to separate trend from dependence is to explicitly model both the trend and the dependence structure from theoretical principles. For example, Craigmile, Peruggia, and Van Zandt (2010a) constructed a Bayesian model within which they estimated the parameters of both the trend and a realistic RT-generating mechanism. This approach is computationally quite expensive, although it yields information about effects on RTs that are not at all obvious when the RTs are treated as independent samples.

Model Fitting

To this point we have discussed analysis of mean RT data and RT time series. The analysis of mean RT is popular for empirical evaluation of non-mathematical hypothesis of cognitive performance (e.g., the stimulus–response compatilibility effects described above). The treatment of RT as time-series data is primarily descriptive, without focused theories to explain trend or the dependencies in the data. By contrast to these approaches, most RT researchers test hypotheses about RT generated by a proposed model of the phenomenon of interest. Most modern models of mental mechanisms make predictions about the shape of the RT distributions. That is, a hypothetical process may dictate that RTs follow some distribution F conditioned on a set of psychologically important parameters θ.[2]

The most common techniques of analysis in RT research are concerned with fits of a proposed model to the data. Fitting a model involves estimating the parameters θ that result in the closest agreement between the hypothesized distribution F and the data. The procedures we describe in this section are very general and apply to any data set and any distribution F.

Once a model is fit to the data, arguments about whether the model is a good one (or better than some other model) usually rely on measures of goodness of fit, such as χ^2 statistics, percentage of variance accounted for, or one of several possible information criteria. In the Meta-Theoretic Model Testing section, we will discuss an alternative to this kind of argumentation. The meta-theoretic approach to model testing is diagnostic in that it restricts models from consideration based on the qualitative characteristics of the RT distribution or other measures rather than goodness-of-fit statistics.

What we present here is not intended to be a "how-to" guide for model fitting; books have been written on these techniques. We wish only to provide an overview of model fitting with enough information that a researcher might decide which technique best suits his needs, so he may then seek out the appropriate comprehensive tutorial (*see*, e.g., Yuan & Schuster, Chapter 18 and Cavagnaro, Myung, & Pitt, Chapter 21, Volume 1).

PARAMETER ESTIMATION

Fitting a model to data is the process by which the model's parameters are estimated. A compelling model has parameters with clear psychological interpretation, and so in addition to being constrained by the observed data, the parameters are constrained by the experimental conditions. If, for example, the influence of stimulus intensity is represented by a parameter a and response bias by parameter b, then the model should fit well over changes in stimulus intensity by changing only parameter a and leaving b constant (*see*, e.g., Donkin, Brown, & Heathcote, 2011).

There are many ways to estimate parameter values, and the most effective methods will depend on the characteristics of the model. We can divide these methods roughly into linear and nonlinear approaches. Linear approaches rely on the concept of least squares: The goal is to estimate parameters by choosing those that minimize the sum of squared error between the observed measurements and those predicted by the model. Response time data, however, usually require nonlinear approaches such as maximum likelihood estimation, nonlinear least squares, and Bayesian methods. Linear approaches are much easier, because there are closed-form solutions for the best-fitting parameters, but models of RT are usually not linear.

Whether a linear model exists (or is reasonable) may depend on the level at which predictions are to be made: Do predictions concern summary statistics such as mean RT or does the model dictate more fine-grained measurement behavior such as how the RTs are to be distributed? Model fitting to mean RTs often relies on linear approaches and, even with nonlinear methods, can sometimes lead to closed-form solutions for parameter estimates, depending on the method.

For example, the *method of moments* can sometimes provide an easy set of equations to solve to obtain parameter estimates. The method of moments is a simple technique in which the mean and higher moments predicted by a model are equated to the sample mean and higher moments of the data. For example, an ex-Gaussian distribution (the sum of independent normal and exponential variables, *see* p. 276) has three parameters, the mean μ and standard deviation σ of the normal component and the mean τ of the exponential component. To solve for three parameters, we will need the first three moments of the distribution. The mean of the ex-Gaussian distribution is $\mu + \tau$, its variance is $\sigma^2 + \tau^2$, and its skewness is $2\tau^3$. Fitting the ex-Gaussian, then, requires setting these moments equal to the sample mean ($\overline{X}$), variance (s^2) and skew (Sk) and solving for the parameter estimates to obtain $\hat{\tau} = (Sk/2)^{1/3}$, $\hat{\sigma}^2 = s^2 - (Sk/2)^{2/3}$ and $\hat{\mu} = \overline{X} - (Sk/2)^{1/3}$.

Unfortunately, the method of moments sometimes yields unsatisfactory results. For example, there is nothing in the ex-Gaussian method of moments estimates that prevents $\hat{\sigma}^2$ from being negative. Method of moments, however, can be very useful for providing starting values for other methods, such as maximum likelihood or nonlinear least squares estimation, in which some objective function is optimized to be as large (or small) as possible by iterative updating of the parameter values.

METHODS OF LEAST SQUARES

Methods of least squares are designed to minimize the error between the observed values of the measurements from an experiment and a model's predicted values. Consider the observed RTs $\{T_1, T_2, \ldots, T_n\}$ and a set of independent variables $\{X_1, X_2, \ldots, X_m\}$. A model's predictions for observation i can be written as $p(X_i, \theta)$, and the sum of squared errors or residuals is

$$SSE = \sum_{i=1}^{n} \left(T_i - p_i, \theta)\right)^2 .$$

We could also consider a set of mean RTs $\{\overline{T}_{ij}\}$ for subjects $i = 1, \ldots, n$ and experimental conditions $j = 1, \ldots J$. If the model's predictions $p(X_{ij}, \theta)$ are targeted at the means, then

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{J} \left(\overline{T}_{ij} - p(X_{ij}, \theta)\right)^2 ,$$

where the independent variable X_{ij} is taken as the ij^{th} element in the $n \times J$ design matrix X.

If the predictions p are a linear function of the parameters θ, then the least-squares method is linear and the solution for the estimates of θ is of closed form and well known. Otherwise, the method is nonlinear. Nonlinear least squares is more tricky than linear least squares but is nonetheless straightforward (Gallant, 1987; Seber & Wild, 2003). Nonlinear least squares does not have closed form solutions and requires iterative algorithms to find θ to minimize SSE. In some circumstances, a nonlinear model can be made linear by a transformation of variables. For example, if a model predicts that mean RT is a power function of the number of trials in an experiment (Logan, 1988), then the log mean RTs can be compared to the (linear) log power function, and the estimation of parameters can proceed using standard regression methods.

For RT data, many researchers have applied a least-squares approach to fitting the RT distribution predicted by a model. In such applications the usual least-squares logic applies, except that the prediction $p(X, \theta)$ is a probability density or cumulative distribution $p(t|X, \theta)$ defined over time t. The data are then summarized as an estimate of that probability density or cumulative distribution. These estimates may be obtained by either parametric or nonparametric methods. For example, a histogram estimate of the empirical RT density may be computed for some fixed number of points along the time axis, and the heights of the histogram bars at those points could be compared to the density function of the model. It turns out that least-squared fits to the empirical probability density do not generally recover accurate values of the parameters (Van Zandt, 2002). However, least squares fits of distribution quantiles or the cumulative distribution function can be as accurate as maximum likelihood estimates.

For example, consider a set of RTs $\{T_1, T_2, \ldots, T_n\}$ from an individual subject. The empirical distribution function $\hat{F}(t)$ is defined as

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i < t),$$

where $I(x)$ is an indicator function that equals 1 if the statement x is true and 0 otherwise. If the model states that the RTs should follow a distribution with cumulative distribution function $F(t, \theta)$, then the parameters θ can be estimated by minimizing

$$SSE = \sum_{i=1}^{m}\left(\hat{F}(t_i) - F(t_i, \theta)\right)^2,$$

for an appropriately selected set of points $\{t_1, t_2, \ldots, t_m\}$.

Another nearly equivalent procedure involves setting the points $\{t_1, t_2, \ldots, t_m\}$ to be the quantiles of the sample $\{T_1, T_2, \ldots, T_n\}$ and using these points as bin boundaries for a χ^2 statistic. If, for example, the t_is are selected to be the sample deciles, then 10% of the sample falls in each of 10 bins defined as the intervals from t_i to t_{i+1} (where $i = 0, \ldots, 9$, $t_0 = 0$ and $t_{10} = \infty$). The theoretical cumulative distribution F dictated by the model gives the predicted proportion of observations between t_i and t_{i+1} as $F(t_{i+1}, \theta) - F(t_i, \theta)$. Letting $O_i = 0.1n$ and $E_i = n\,(F(t_{i+1}, \theta) - F(t_i, \theta))$, we may adjust θ to minimize

$$\chi^2 = \sum_{i=0}^{9}(O_i - E_i)^2/E_i$$

(e.g., Smith & Vickers, 1988).

One nice feature of χ^2 minimization is that the minimized value of χ^2 may also serve as a goodness-of-fit measure. If χ^2 is sufficiently small given the degrees of freedom in the model, then we can argue that the model fits well. However, large values of χ^2 do not necessarily indicate an incorrect or misspecified model. The χ^2 statistic is very sensitive to sample size and frequently can be "significantly" large even when the model is correct (Van Zandt, 2000).

MAXIMUM LIKELIHOOD

Maximum likelihood is a powerful estimation method that produces fits to a model that makes predictions about the distribution from which data were sampled. For example, a model might state that RTs are normally distributed with mean μ and standard deviation σ (*see*, e.g., the AR(1) model presented earlier). For a single RT observed to be t, the *likelihood* of the value t is given by the height of the normal density with mean μ and standard deviation σ at time t, or $\phi\,((t - \mu)/\sigma)$, where ϕ is the standard normal density. For a discrete random variable, the likelihood is interpreted as the probability of observing the measured value of the variable given the parameter values θ. One way of thinking about maximum likelihood, then, is that we choose parameters that give the highest possible probability of having observed the data we obtained.

Suppose that a model states that the RT distribution has a probability density function $f(t|\theta)$, where t is a possible value for an observation and θ is the vector of parameters for the distribution (like μ, σ and τ for the ex-Gaussian distribution introduced earlier). We typically assume that the data from an experiment $\{T_1, T_2, \ldots, T_n\}$ form a set of independent and identically distributed observations from the distribution described by $f(t|\theta)$, so that the joint probability of having observed the data is given by

$$f(T_1|\theta)f(T_2|\theta)\cdots f(T_n|\theta) = \prod_{i=1}^{n} f(T_i|\theta).$$

The likelihood is then defined as

$$L(\theta|\mathbf{T}) = \prod_{i=1}^{n} f(T_i|\theta),$$

a function of θ given fixed values for $\mathbf{T}$.

We want to choose $\hat{\theta}$ so that the likelihood of the data we obtained is as high as possible, or L reaches a global maximum at $\hat{\theta}$. Sometimes closed-form expressions for the maximum likelihood estimates of θ are obtainable by methods of calculus. For example, the maximum likelihood estimate of a shift parameter (sometimes referred to as peripheral processing time in RT data) is given by the smallest observation in the sample. However, it is rarely possible to find closed-form expressions for the parameters for real problems. Rather, we program the likelihood function (if a canned routine is not easily obtainable) and pass it to an optimization algorithm that attempts to find the maximum, just as for nonlinear least squares minimization. Numerically it is usually easier to work with the log likelihood function; because the relationship between L and $\log L$ is monotonic, maximizing $\log L$ also maximizes L.

As with every estimation method, maximum likelihood has some drawbacks. First, parameter estimates that maximize likelihood may not exist. Second, if they do exist, then they may not be unique. That is, there may be another, completely different set of parameter values that work equally well. A third and related problem is that once a set of estimates has been found, it is difficult to determine if the value of the likelihood is a global maximum or only a local maximum. Fourth, sometimes the maximum likelihood estimate will be found at the extremes of the boundaries for the parameters. Proportions, for example, are bounded between 0 and 1, and the maximum likelihood estimate may be 1. The shift parameter for RT data is another example where the maximum likelihood estimate is equal to the value of the smallest observation. When the best estimates are at the extreme ends of a scale, this will frequently influence the estimates of other parameters. Maximum likelihood estimates may also be biased—for example, the maximum likelihood estimate of the shift parameter is too large and consistently overestimates the true shift. Sometimes the modeler will need to make some arbitrary decisions about parameter limits to move the estimates back to a reasonable value.

Maximum likelihood estimates also have many good qualities. Under general conditions, maximum likelihood estimates converge in probability to the true parameter value, they are asymptotically normal, and they have the smallest possible variance. Also, under general conditions, the maximum likelihood estimates minimize the sum of squared error.

A related method is quantile maximum likelihood estimation, which transforms the data into quantiles and then maximizes a likelihood based on the predicted proportion of observations falling between the quantiles (Heathcote, Brown, & Mewhort, 2002). This method is especially useful when the probability density function misbehaves for some parameter values (e.g., when singularities arise or when the function becomes sharply peaked) and when outliers result in likelihoods equal to zero. Brown and Heathcote (2003) have provided software for quantile maximum likelihood estimation of the ex-Gaussian distribution—software that can be modified to accommodate other RT distributions.

THE EX-GAUSSIAN DISTRIBUTION

A popular way to characterize RT data is to use a parametric description of the sample that provides a summary of the shape of the empirical distribution. Although several candidate distributions exist, the most popular is the ex-Gaussian distribution, which is the distribution of the sum of a Gaussian variable (with mean μ and standard deviation σ) and an exponential variable (with mean τ). This distribution, although atheoretical, is very flexible and can capture a wide variety of postively skewed distributions. Therefore, many have found it very convenient to summarize RT data with estimates of μ, σ, and τ (Ratcliff, 1979; Ratcliff & Murdock, 1976; Heathcote, Popiel, & Mewhort, 1991).

The ex-Gaussian estimates can be obtained in a number of ways, the most reliable being maximum likelihood or nonlinear least-squares fits to the cumulative distribution functions. Several routines are publicly available to assist in performing these computations (Cousineau & Larochelle, 1997; Dawson, 1988; Heathcote, 1996).

The ex-Gaussian characterization of RTs is useful for estimating the shape of the RT distribution (Heathcote et al., 1991). It is less useful as a tool for inference, or trying to determine whether experimental manipulations had different effects on the

RT distributions. For example, a researcher may be concerned that one variable influenced only the slow RTs (an effect that might show up in τ) and another variable influenced only the fast RTs (an effect that might show up in μ). However, the distributions of the estimates of μ, σ, and τ are unknown; they depend on the underlying (and unknown) RT distribution. It is difficult, therefore, to determine the error in the estimates of μ, σ, and τ. More crucially, the parameter estimates are highly correlated. Because the sample mean must approximately equal $\mu + \tau$ (*see* p. 274), as μ increases τ must decrease to fit the data. It may not be possible, therefore, to argue conclusively about the effects that experimental manipulations have on these parameters.

It must also be noted that despite the ability of the ex-Gaussian to fit RT data, there are a number of features of RT data we can point to that rule out the ex-Gaussian as a model for RT data (Burbeck & Luce, 1982; Luce, 1986; Van Zandt, 2000). This fact, together with the understanding that the ex-Gaussian is an atheoretical model of the RT distribution and, conditioned on the data, the parameters are strongly correlated, makes it difficult to interpret psychologically the changes in the different parameters across experimental conditions.

Meta-Theoretic Model Testing

Although model fitting is primarily an exercise in parameter estimation, model testing is more concerned with discriminating between different potential data-generating mechanisms on the basis of qualitative characteristics of the data. This is a problem that arises in many scientific endeavors, but in psychology it relies quite heavily on RT data. We now have a range of theoretical tools that can be applied to such data to try and discriminate among different kinds of cognitive architectures.

The question of whether people can perceive or process a set of visual objects immediately and simultaneously (i.e, in parallel) or whether attention must be switched to each object in succession extends back more than a hundred years (e.g., Hamilton, n.d.). We discussed already how in the 1960s, this question was re-opened by Sternberg (1966) as the serial versus parallel processing issue. He and others reasoned that serial processing should produce mean RTs that increase linearly with the number of items to be searched, whereas parallel processing should produce increasing but negatively accelerated mean RTs. Townsend (1972, 1976) demonstrated, however, that the behavior of mean RT as a function of items to be processed was determined more by the capacity of the process than whether the architecture of the process was serial or parallel. Increasing mean RT indicates that the system slows down as the load increases, and certain parallel models with limited capacity could generate RTs distributed in exactly the same way as serial models.

Over the past several years, Townsend and his colleagues (e.g., Townsend & Nozawa, 1995; Townsend & Wenger, 2004a) have proposed a methodology to separate capacity from architectural issues such as serial versus parallel processing and dependence versus independence of processing channels using factorial methods and redundant targets. These methods rely on experimental designs in which stimuli vary on at least two orthogonal dimensions, such as intensity and location. One of the stimulus dimensions (location) can be reasonably assumed to correspond to different processing channels or pathways. Townsend Nozawa have referred to these kinds of experiments as "double factorial designs." The logic of Sternberg's (1969) additive factors method rests on such a design, although the procedures we describe here extend to the RT distributions and do not depend on the mean RTs (cf. Roberts & Sternberg, 1992).

The factorial methods proposed by Townsend and colleagues depend on the variables in the experimental design having *selective influence*—that is, a variable influences one and only one subprocess of the task. Dzhafarov and colleagues have worked extensively on the question of selective influence and how it can be used to learn about the smaller components that make up a more complex task (Dzhafarov, 2003; Dzhafarov & Cortese, 1996; Dzhafarov & Gluhovsky, 2006; Dzhafarov & Schweickert, 1995; Kujala & Dzhafarov, 2008). Interested readers may consult these papers or Van Zandt (2002) for a brief overview of selective influence.

FACTORIAL METHODS

Consider an experiment where more than one stimulus can be presented at one time. To investigate questions of process architecture, we can assume that each distinct stimulus is processed through a separate processing channel. It is easiest to think about stimuli that differ in spatial location in a visual array, but we can also consider auditory stimuli presented to different ears, tactile stimuli presented at different locations on the body, or even visual spatial gratings presented in the same location but with different frequencies, such frequencies being generally thought to require different processing locations in visual cortex. The general idea is to set up a scenario where more than one stimulus could possibly be processed at the same time.

We now do a simple RT experiment where people respond to stimuli that vary according to some feature (such as intensity), crossed factorially with processing channels (like location). In the simplest design, consider stimuli with two levels of intensity (on or off) presented in two locations (left or right). Are the left and right channels independent of each other? What is their capacity? Does one slow down when the other is working? Can it work at all or must it wait until the other is finished? Can information be shared across channels?

A parallel channel model is a broad class of models that assume information flows through more than one pathway toward the execution of a response (*see also* Fig. 14.1, bottom panel). These models are often conceptualized as races, where a response is made as soon as any channel is finished (*see* Fig. 14.3, top panel). Response times for race models are therefore distributed as the minimum of the processing times for all of the channels. Such race models are often called "self-terminating," or OR, models because processing ends as soon as one or the other channel is finished. By contrast, an "exhaustive," or AND, model requires that all channels complete processing before a response is made (*see* Fig. 14.3, bottom panel), and RTs are distributed as the maximum of the processing times for all of the channels.

Imagine a factorial design presenting observers with two lights colored red and green in different locations. The lights may be of different colors or they may be the same. We might assume that the simple RT task with these stimuli, using instructions stating, "Respond as soon as you see a light," encourages OR processing. Similarly, we might assume that the go/no-go task using instructions stating, "Respond only if the two lights are the same color," encourages AND processing. However, we can't be certain that this is actually what people do. People could use either OR or AND processing under either set of instructions, which brings up the issue of the *decisional stopping rule* in any task requiring the processing of more than one stimulus. The stopping rule is not something that can be deduced from the task itself. For example, a lazy system might process fewer items than required in a task that encourages AND processing. Townsend and Colonius (1997) and Van Zandt and Townsend (1993) have explored ways for determining the stopping rule in visual display and memory search experiments, but the factorial methods we present here can be applied to either stopping rule. The researcher must be aware that the stopping rule interacts with other aspects of cognitive architecture, most importantly those of processing channel independence and capacity.

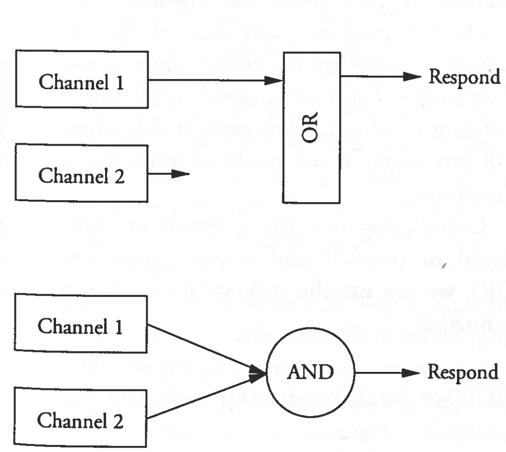

Figure 14.3 Two parallel architectures with different processing schemes. The top panel shows a self-terminating process, where the response can be made at the end of processing on either Channel 1 or Channel 2. The bottom panel shows an exhaustive process, where the response can be made only after processing is completed for both Channels 1 and 2.

CHANNEL INDEPENDENCE AND CAPACITY

Consider the simple factorial design with two stimulus locations (A and B) and two light intensities (on or off). Assume that a response can be made as soon as any channel signals the presence of a stimulus (an OR stopping rule). This kind of experiment is often called a redundant targets design. In such a design, RTs are faster when two targets are present (both are on) than when only one is present (one is off). If responses are generated by a race between independent processing channels, a decrease in RT when both targets are present is expected because the minimum of two random variables will have a smaller mean than the means of either of the two random variables alone; this is sometimes called a *statistical advantage*. How fast does the redundant target RT have to be before we can say that something more than a statistical advantage is occurring—that there is information being shared between the channels? How slow does it have to be before we can say that the two channels interfere with each other?

The race inequality (J. Miller, 1982) is an empirical relationship between the RT cumulative distribution functions that must hold for the redundant target conditions and the single target conditions if a parallel race model is generating the RTs. The cumulative distribution function gives the proportion of observations that fall below a value t, or the probability that an RT will be less than t. If the two channels are A and B (corresponding to stimulus location in our design, but the channels could correspond to some other stimulus dimension), let $F(t|A,B)$ is the cumulative distribution of the RTs for the redundant condition and $F(t|A)$ and $F(t|B)$ are the cumulative distributions for the RTs in the A-only and B-only conditions, respectively. Then the race inequality states that

$$F(t|A,B) \leq F(t|A) + F(t|B).$$

If this inequality is violated for any value of t, then according to Miller, a large set of parallel channel race models are falsified.

The Grice inequality (Grice, Canham, & Gwynne, 1984) provides the upper bound on the RT distribution for the redundant targets condition, or how slow the redundant target RTs can be before we have to conclude that the two channels interfere with each other (Colonius, 1990). The redundant targets RT distribution must satisfy

$$\max\{F(t|A), F(t|B)\} \leq F(t|A,B)$$

if a parallel channel race model is generating the RTs.

Townsend and Nozawa (1995) noted that the boundaries on the RT distribution in the redundant targets paradigm imposed by the race and Grice inequalities are determined by the *workload capacity* of a parallel process and not the parallel or serial architecture itself. If the data satisfy both inequalities, then RTs neither speed up enough to conclude that processing is facilitated across channels (what Miller, 1982, called *coactive* processing) nor slow down enough to conclude that a load in one channel reduces the efficiency of the other. Increasing the amount of information to be processed by moving from a single target to a redundant target stimulus does not influence the efficiency of the channels. The fact that these bounds are capacity limits and not limits imposed by architecture means that they may be violated by a parallel race process of either limited or "super" capacity.

SEPARATING CAPACITY FROM ARCHITECTURE

The Miller and Grice inequalities apply only to processes with OR stopping rules. Townsend and Wenger (2004a) have reviewed much of the literature on RT-based tests of process structure and generalized this kind of thinking to processes with AND stopping rules as well. They have emphasized the relationships between channel independence, the stopping rule required by the process, process architecture, and capacity. An important approach to identifying these different aspects of a cognitive task was presented by Townsend and Nozawa (1995), who investigated the Miller and Grice inequalities in the context of their *systems factorial technology*.

Consider again the design of the redundant targets task, but add a third level of stimulus intensity so that a stimulus may be off (·) or of low (L) or high (H) intensity. The low-intensity stimulus slows the channel processing that stimulus. Lowering the intensity of the stimulus in one channel does not influence the processing speed in the other channel—that is, the intensity *selectively influences* the processing times in each channel. Townsend and Nozawa (1995) contributed two tools for analyzing data in such tasks: the *interaction contrast* $SIC(t)$ and the *capacity coefficient* $C(t)$ for OR tasks. Townsend and Wenger (Townsend & Wenger, 2004b) later expanded the capacity coefficient to AND tasks.

Recall from above that the RT cumulative distribution function $F(t)$ gives the proportion of RTs that are less than some value t. It is the probability of observing an RT faster than time t. The survivor function is $S(t) = 1 - F(t)$, or the probability that an RT is slower than time t. We can subscript these functions to indicate the different conditions in the double factorial experiment, so $F_{ij}(t)$ is the cumulative distribution function when stimulus $i = \cdot, L$, or H (absent, low intensity, or high intensity) is presented in the left channel and stimulus $j = \cdot, L$, or H is presented in the right channel, so both high- and low-intensity stimuli can be processed in either channel. We can characterize capacity by the relationships between the distribution $F_{ij}(t)$ or survivor $S_{ij}(t)$ functions in different conditions. For experiments encouraging OR processing, it is convenient to use the survivor functions.

Considering first the question of architecture (serial or parallel) and stopping rule (AND or OR), we can use the survivor interaction contrast defined as

$$SIC(t) = [S_{LL}(t) - S_{LH}(t)] - [S_{HL}(t) - S_{HH}(t)]. \quad (1)$$

Notice that the interaction contrast relies only on those (redundant target) conditions where a stimulus is presented in both the left and the right locations—the contrast is constructed using only the functions $F(t|A,B)$ over the different stimulus

Table 14.1. Interaction contrast predictions for different cognitive architectures and stopping rules. The time t^* is a constant that is not necessarily the same for all models; it only marks the point at which the behavior of the function changes.

Serial OR	$SIC(t) = 0$ for all t
Serial AND	$SIC(t) < 0$ for $t < t^*$ and $SIC(t) > 0$ for $t > t^*$
Parallel OR	$SIC(t) > 0$ for all t
Parallel AND	$SIC(t) < 0$ for all t
Coactivation	$SIC(t) < 0$ for $t < t^*$ and $SIC(t) > 0$ for $t > t^*$

conditions. This means that tests of processing architecture using $SIC(t)$ are not confounded by changes in the number of items to be processed (workload), as are tests that rely on changes in mean RT with changes in workload. Table 14.1 shows behavior of the survivor interaction contrast for different architectures and stopping rules. The qualitative behavior of serial AND and coactivation models shown in Table 14.1 appears to be the same, but the serial AND models predict equal positive and negative areas under the $SIC(t)$ curve, whereas coactivation models predict small negative and large positive areas under the curve, thus providing for experimental discrimination of these models.

We now turn to the issue of *system capacity*, or the efficiency of each processing channel or processing stage under changes in the workload of the system. The interaction contrast $SIC(t)$ function is measured for a constant workload and is used to assess architecture and stopping rules. Capacity is logically independent of the system's architecture—for example, whether the system is serial or parallel—although serial systems are frequently assumed to be of limited capacity and parallel systems are assumed to be of unlimited capacity. To measure system capacity, we must be able to assess the efficiency of the system, irrespective of architecture, under changes in workload.

Measures of capacity must take into account the fact that the stopping rule will affect overall processing time. When all processes must be completed (for AND processing), RTs will generally be slower than when only a single process must be completed (for OR processing). Therefore, the capacity coefficient $C(t)$ takes different forms for the two stopping rules. For the OR task,

$$C_{OR}(t) = \frac{-\ln S(t|A,B)}{-\ln S(t|A) - \ln S(t|B)}, \quad (2)$$

and for the AND task

$$C_{AND}(t) = \frac{\ln F(t|A) + \ln F(t|B)}{\ln F(t|A,B)}. \quad (3)$$

Note that we are suppressing the notation associated with stimulus intensity for the capacity coefficient and have returned to the notation specifying the activities in the processing channels A and B. Thus, in contrast to the interaction contrast $SIC(t)$, the capacity coefficient makes use of the stimuli in the single-target conditions and examines only one stimulus intensity i, which can be either high or low. If either $C_{OR}(t)$ or $C_{AND}(t)$ is greater than 1 for any t, then the process is "super" capacity or coactive. If either $C_{OR}(t)$ or $C_{AND}(t)$ is less than 1 for any t, then the process is limited capacity. If $C_{OR}(t)$ or $C_{AND}(t)$ equals 1 for all t, then the process is unlimited in capacity.

Townsend and Nozawa (1995) estimated the interaction contrast and the capacity coefficient functions from data from a double-factorial simple RT design. The behavior of $SIC(t)$ and $C_{OR}(t)$ suggested super-capacity parallel processing, with an OR stopping rule in one experiment, and limited capacity parallel processing, with an OR stopping rule in another experiment. More recently, Townsend and Eidels (2011) have showed how the race and Grice inequalities for AND and OR tasks could be expressed in terms of limits on the capacity coefficient. This allows the inequalites to be examined simultaneously with the capacity coefficient to allow greater insight into the capacity characteristics of factorial systems.

Together, the use of the interaction contrast and capacity coefficient provide initial insights on the fundamental structure and mechanisms of the investigated system, insights that can then be explored in additional experiments. Interested readers should consult Townsend and colleagues' work (1995; 2004a; 2011) for more details on these tests. Van Zandt (2002) has provided guidelines for how these tests can be applied to data and some examples.

Summary

In this section we discussed the analyses of RT data. There are different analyses for different purposes. Most commonly, we attempt to estimate parameters of cognitive models from RT data, or we test different classes of models in an attempt to discover the fundamentals of cognitive structure.

There are many good references that researchers intending to perform RT analyses should consult before proceeding; we have been able only to provide a brief overview of these techniques here. For more information on model fitting and parameter estimation, an excellent reference is Busemeyer and Diederich (2010). An excellent discussion of metatheoretic tests and the philosophy behind them can be found in Townsend and Wenger (2004a).

One issue we have not touched on is that of model comparison. That is, when two or more models fit the data or satisfy the constraints of the data, how do we choose between them? This important and difficult question, which faces all areas of quantitative research and not just RT experiments, has been extensively addressed by Myung, Pitt, and colleagues (e.g., Cavagnaro, Myung, Pitt, & Kujala, 2010; Navarro, Pitt, & Myung, 2004; Pitt, Kim, & Myung, 2003) and is summarized in Chapter 21 (Cavagnaro, Myung, & Pitt, Chapter 21, Volume 1).

Another closely related issue involves determining statistical significance of model tests. That is, the estimates of the RT distributions are random and subject to sampling error. Therefore, we might expect poor fits or violations of expected behavior (such as an interaction contrast everywhere positive) by chance alone. Determining whether violations are statistically significant is not trivial: The points on each curve are not independent from each other, and this dependence will increase the likelihood that spurious differences will be statistically significant. A number of strategies have been proposed to test these relationships, and the best approach so far is that of Houpt and Townsend (2010).

Conclusion

This chapter has outlined the kinds of experimental designs most commonly used in RT experiments and then the most popular methods of RT analysis. Each of the subsections in this chapter, however briefly presented here, has been the topic of many papers and chapters elsewhere and are necessarily very broad overviews of quite complex topics. We have tried to provide the best references to the work in these areas so that interested readers can find the help they need at a more detailed level.

To the reader who asks, "What kind of experiment should I do and how should I analyze my data?" we respond: It depends. It depends on the question you are asking, the hypothesis you are trying to test. It is never a good idea to shoehorn a general method to fit a specific problem. Although this chapter gives some guidance in experimental design and analysis, the beauty of the RT experiment is in its flexibility: It may be a simple as measuring a single keypress, or it could measure the times between notes executed by a concert pianist (e.g., Goebl & Palmer, 2009). We hope this chapter provides enough background that the reader feels more confident in creating the unique approach appropriate for his or her unique problem.

Future Directions

Perhaps the most important new technique for data analysis in RT studies is being provided by the application of Bayesian statistical techniques. These techniques allow the data to be analyzed within a theoretically motivated framework, one in which the likelihood of the data is provided by the model of interest. We can contrast that to more traditional methods, such as analysis of variance applied to mean RT data, where the model being fit is known to be false and is therefore of no interest at all.

There are now statistical packages (such as JAGS and WinBUGS) that will assist in the development of Bayesian models that will run on any desktop computer. Unfortunately, these packages still do not handle well the kinds of models typically explored for RT data, so Bayesian modeling of RTs is still restricted to a small group of quantitative researchers with mathematical and programming expertise. Specialized routines to assist in this kind of modeling will soon become available, opening this avenue to everyone.

Author Note

This work was supported by the National Science Foundation under grants no. BCS-0738059, and SES-1024709 the National Institutes of Mental Health under grant no. 57717-04A1, and the Air Force Office of Special Research grant no. FA9550-07-1-0078.

Notes

1. The restriction that $n \leq m$ is not required for the definition of information but is required for an n-choice task.

2. Psychologists and statisticians use the word "model" in slightly different ways. Psychologists refer to hypothetical cognitive mechanisms as models that then dictate the probability distributions that data will follow. Statisticians refer to the probability distributions themselves as models without as much consideration of the mechanisms that dictate those distributions. We see these two points of view as interchangeable for the purposes of this section, but the reader should be cautious. Different cognitive mechanisms may dictate that data follow the same distribution, and the same cognitive mechanism may dictate that data follow different distributions depending on the assumptions made to implement the model.

References

Ashby, F. G., & Townsend, J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, *21*, 93–123.

Balakrishnan, J. D. (1994). Simple additivity of stochastic psychological processes: Tests and measures. *Psychometrika*, *59*, 217–240.

Borowsky, A., Oron-Gilad, T., & Parmet, Y. (2009). Age and skill differences in classifying hazardous traffic scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*, 277–287.

Brown, S. D., & Heathcote, A. (2003). QMLE: Fast, robust, and efficient estimation of distribution functions based on quantiles. *Behavioral Research Methods, Instruments, & Computers*, *35*, 485–492.

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception and Psychophysics*, *32*, 117–132.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Thousand Oaks, CA: Sage Publications.

Cavagnaro, D. R., Myung, J. I., & Pitt, M. A. (2012). Mathematical modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (Vol. 1, pp. 438–453). New York: Oxford University Press.

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, *22*, 887–905.

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology*, *34*, 253–275.

Cousineau, D., & Larochelle, S. (1997). PASTIS: A program for curve and distribution analyses. *Behavioral Research Methods, Instruments, & Computers*, *29*, 542–548.

Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2010a). Hierarchical Bayes models for response time data. *Psychometrika*, *75(4)*, 613–632.

Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2010b). Detrending response time series. In S.-M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (Vol. 4, pp. 213–240). New York: Taylor and Francis.

Dawson, M. R. (1988). Fitting the ex-gaussian equation to reaction time distributions. *Behavioral Research Methods, Instruments, & Computers*, *20*, 54–57.

Diederich, A., & Colonius, H. (2008). Crossmodal interaction in saccadic reaction time: Separating multisensory from warning effects in the time window of integration model. *Experimental Brain Research*, *186*, 1–22.

Donders, F. C. (1868/1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. (Translated by W. G. Koster)

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, *55(2)*, 140–151.

Duncombe, R. L. (1945). Personal equation in astronomy. *Popular Astronomy*, *53*, 2–13, 63–76, 110–121.

Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika*, *68*, 7–26.

Dzhafarov, E. N., & Cortese, J. M. (1996). Empirical recovery of response time decomposition rules I. Sample-level decomposition tests. *Journal of Mathematical Psychology*, *40*, 185–202.

Dzhafarov, E. N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *Journal of Mathematical Psychology*, *50*, 390–401.

Dzhafarov, E. N., & Schweickert, R. (1995). Decompositions of response times: An almost general theory. *Journal of Mathematical Psychology*, *39*, 285–314.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics*, *72(8)*, 2274–2288.

Erceg-Hurn, D. M., Wilcox, R. R., & Keselman, H. H. (2012). Robust statistical estimation. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (Vol. 1, pp. 388–406). New York: Oxford University Press.

Fitts, P. M., & Seeger, M. (1953). S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*, 199–210.

Gallant, A. R. (1987). *Nonlinear statistical models*. New York: Wiley Press.

Garcia-Perez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, *38*, 1861–1881.

Gilden, D. L. (1997). Fluctuations in the time required for elementary decisions. *Psychological Science*, *8*, 296–301.

Gilden, D. L. (2001). Cognitive emissions of $1/f$ noise. *Psychological Review*, *108*, 33–56.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*, 427–438.

Green, D. M., & Luce, R. D. (1971). Detection of auditory signals presented at random times: III. *Perception and Psychophysics*, *9*, 257–268.

Grice, G. R., Canham, L., & Gwynne, J. W. (1984). Absence of a redundant-signals effect in a reaction time task with divided attention. *Perception and Psychophysics*, *36*, 565–570.

Hamilton, S. W. (n.d.). Lectures on metaphysics and logic. In (Vol. I, pp. 154–170). Boston: Gould and Lincoln.

Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavioral Research Methods, Instruments, & Computers*, *28*, 427–445.

Heathcote, A., Brown, S. D., & Mewhort, D. J. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin and Review*, *9*, 394–401.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the stroop task. *Psychological Bulletin*, *109*, 340–347.

Heiervang, E., & Hugdahl, K. (2003). Impaired visual attention in children with dyslexia. *Journal of Learning Disabilities*, *36*, 68–73.

Helmholtz, H. (1850). Vorläufiger Bericht über die Fortpflanzungsgeschwindigkeit der nervenreizung. *Archiv für Anatomie, Physiologie und Wissenschaftliche Medizin*, 71–73.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26.

Holden, J. G. (2005). Gauging the fractal dimension of response times from cognitive tasks. In M. A. Riley & G. C. V. Orden (Eds.), *Tutorials in contemporary nonlinear methods for behavioral scientists*. Arlington, VA: National Science Foundation. Retrieved September 2010 from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp. Last accessed May 7, 2012

Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review, 116*, 318–342.
Houpt, J. W., & Townsend, J. T. (2010). The statistical properties of the Survivor Interaction Contrast. *Journal of Mathematical Psychology, 54(5)*, 446–453. doi:10.1016/j.jmp.2010.06.006
Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*, 188–196.
Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cognitive Science: A Multidisciplinary Journal, 32*, 1217–1231.
Kirby, N. H. (1980). Sequential effects in choice reaction time. In A. T. Welford & J. M. T. Brebner (Eds.), *Reaction times* (pp. 129–172). New York: Academic Press.
Klemmer, E. T. (1956). Time uncertainty in simple reaction time. *Journal of Experimental Psychology, 51*, 179–184.
Kohfeld, D. (1971). Simple reaction time as a function of stimulus intensity in decibels of light and sound. *Journal of Experimental Psychology, 88*, 251–257.
Kujala, J. V., & Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *Journal of Mathematical Psychology, 52*, 128–144.
Laming, D. R. (1968). *Information theory of choice reaction time.* New York: Wiley Press.
Laming, D. R. (1979). Autocorrelation of choice-reaction times. *Acta Psychologica, 43*, 381–412.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 119–159.
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.
Logan, G. D., & Burkell, J. (1986). Dependence and independence in responding to double stimulation: A comparison of stop, change and dual-task paradigms. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 549–563.
Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: A model and a method. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 276–291.
Logan, G. D., & Gordon, R. D. (2001). Execution control of visual attention in dual-task situations. *Psychological Review, 108*, 393–434.
Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.
Mandelbrot set. (2010). In E. W. Weisstein (Ed.), *Mathworld – a Wolfram web resource.* Retrieved December 14, 2010, from http://mathworld.wolfram.com/MandelbrotSet.html.
Mansfield, R. J. (1973). Latency functions in human vision. *Vision Research, 13*, 2219–2234.
Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–87.
Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247–279.
Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology, 49*, 47–84.
Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89*, 133–162.
Pachella, R. G., & Pew, W. (1968). Speed-accuracy tradeoff in reaction time: Effect of discrete criterion times. *Journal of Experimental Psychology, 76*, 19–24.
Peruggia, M., Van Zandt, T., & Chen, M. (2002). Was it a car or a cat i saw? An analysis of response times for word recognition. *Case Studies in Bayesian Statistics, VI*, 319–334.
Piéron, H. (1920). Nouvelles recherches sur l'analyse du temps de latence sensorielle et sur la loi qui relie ce temps à l'intensité de l'excitation (New research on the analysis of sensory latency and the law relating latency to the intensity of excitation). *L'Année Psychologique, 22*, 58–142.
Pitt, M. A., Kim, W., & Myung, I. J. (2003). Flexibility vs generalizability in model selection. *Psychonomic Bulletin and Review, 10*, 29–44.
Proctor, R. W., & Vu, K.-P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application.* Boca Raton, FL: CRC Press.
Querne, L., & Berquin, P. (2009). Distinct response time distributions in attention deficit hyperactivity disorder subtypes. *Journal of Attention Disorders, 13*, 66–77.
Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446–461.
Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.
Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190–214.
Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333–367.
Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science, 181*, 574–576.
Remington, R. J. (1969). Analysis of seequential effects in chice reaction times. *Journal of Experimental Psychology, 82*, 250–257.
Roberts, S., & Sternberg, S. (1992). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 611–654). Cambridge, MA: MIT Press.
Santee, J. L., & Egeth, H. E. (1982). Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception and Performance, 8*, 489–501.
Schweickert, R. (1978). A critical path generalization of the additive factor methods analysis of a Stroop task. *Journal of Mathematical Psychology, 18*, 105–139.
Schweickert, R., Giorgini, M., & Dzhafarov, E. N. (2000). Selective influence and response time cumulative distribution functions in serial-parallel networks. *Journal of Mathematical Psychology, 44*, 504–535.
Schweickert, R., & Townsend, J. T. (1989). A trichotomy method: Interactions of factors prolonging sequential and concurrent mental processes in the stochastic PERT networks. *Journal of Mathematical Psychology, 33*, 328–348.
Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear regression.* Hoboken, NJ: John Wiley & Sons.
Shore, D. I., & Spence, C. (2005). Prior entry. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 89–95). New York: Elsevier.

Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81*, 174–176.
Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review, 102*, 567–593.
Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32*, 135–168.
Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652–654.
Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North-Holland.
Stevens, C. J., Brennan, D., Petocz, A., & Howell, C. (2009). Designing informative warning signals: Effects of indicator type, modality, and task demand on recognition speed and accuracy. *Advances in Cognitive Psychology, 5*, 42–48.
Stuart, A., & Ord, J. K. (1999). *Kendall's advanced theory of statistics* (6th ed., Vol. 1). London: Edward Arnold.
Sullivan, J. M., Tsimhoni, O., & Bogard, S. (2008). Warning reliability and driver performance in naturalistic driving. *Human Factors, 50*, 845–852.
Teichner, W. H., & Krebs, M. J. (1972). Laws of the simple visual reaction time. *Psychological Review, 79*, 344–358.
Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14*, 1–36.
Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology, 25*, 168–199.
Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology, 14*, 219–238.
Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology, 28*, 363–400.
Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* New York: Cambridge University Press.
Townsend, J. T., & Colonius, H. (1997). Parallel processing response times and experimental determination of the stopping rule. *Journal of Mathematical Psychology, 41*, 392–397.
Townsend, J. T., & Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic Bulletin and Review, 18*, 659–681.
Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*, 321–359.
Townsend, J. T., & Wenger, M. J. (2004a). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin and Review, 11*, 391–418.
Townsend, J. T., & Wenger, M. J. (2004b). A theory of interactive parallel processing: New capacity measures and predictions of a response time inequality series. *Psychological Review, 111*, 1003–1035.
Ueno, T. (1978). Temporal summation in human vision: Simple reaction time measurements. *Perception and Psychophysics, 23*, 43–50.
Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*, 34–80.
Usher, M., & McClelland, J. L. (2001). On the time course of pereptual choice: The leaky competing accumulator model. *Psychological Review, 108*, 550–592.
Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331–350.
Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology, 47*, 631–650.
Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin and Review, 7*, 424–465.
Van Zandt, T. (2002). Analysis of response time distributions. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., pp. 461–516). New York: John Wiley & Sons.
Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin and Review, 7*, 208–256.
Van Zandt, T., & Jones, M. R. (2012). *Stimulus rhythm and choice performance.* (Submitted)
Van Zandt, T., & Townsend, J. T. (1993). Self-terminating and exhaustive processes in rapid visual and memory search: An evaluative review. *Perception and Psychophysics, 53*, 563–580.
Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience and Biobehavioral Reviews, 33*, 647–661.
Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114*, 830–841.
Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition. *Psychonomic Bulletin and Review, 11*, 579–615.
Wei, W. W. S. (2012). Time series analysis. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (Vol. 2, pp. 458–485). New York: Oxford University Press.
Welford, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance – a review and a theory. *British Journal of Psychology, 43*, 2–19.
Welford, A. T., & Brebner, J. (Eds.). (1980). *Reaction times.* New York: Academic Press.
Woodworth, R., & Schlosberg, H. (1954). *Experimental psychology.* New York: Holt.
Yuan, K-H., & Schuster, C. (2012). Overview of statistical estimation methods. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (Vol. 1, pp. 360–386). New York: Oxford University Press.