HUMAN MEMORY AND COGNITIVE CAPABILITIES
Mechanisms and Performances
F. Klix and H. Hagendorf (Editors)

RESPONSE TIME DISTRIBUTIONS IN MEMORY SEARCH: A CAUTION

R.Duncan Luce

Harvard University
Cambridge, MA, 02138
USA

Distributions of response times include much useful information that is lost in summary statistics like the mean. Examples of their effective use in memory search are cited. However, absolute identification studies are cited in which pronounced sequential effects and local speed-accuracy tradeoffs are found that render observed distributions very misleading. It is urged that analogous studies be carried out for memory search to see whether or not the same phenomena exist there. The outcome is not certain since, unlike the identification designs, different stimulus sets are used on each trial. But until we know, caution in using distributions is advised.

EXAMPLES OF EMPIRICAL DISTRIBUTIONS

One characteristic of the modern studies of short term memory is a great dependence upon response times as a dependent variable. Indeed, experimental designs frequently are such that nothing else can be observed since it is arranged that the behavior is largely error free. Much of this work uses only estimates of the mean times, but occasionally one see studies in which other aspects of the response time distribution play a role. I wish to focus my attention on this and to suggest that a good deal of methodological work is needed.

Let me begin by outlining the major motivation for looking at distributions and cite the relatively few studies in which something other than the mean has been used. The major argument for studying distributions is that they include more information about the mental processing underlying the response than does just the mean. The actual times are, in some meaningful sense, a more discriminating measure than are statistics calculated from them. The operative phrase here is "in some meaningful sense" for otherwise the statement is tautologically true; however it is thought also to be true with the restriction added. Certainly, that is the case for simple reaction times where for a number of years effective use has been made of distributions (Chapter 4 of my book Response Times (Luce (1985) lists some 30 papers in which they are used). Distributions have been less widely reported in the study of both choice reaction times and the search procedures commonly used in the study of memory. Among the memory examples with which I am familiar are the following.

First, Schneider and Shiffrin (1977) argued against S. Sternberg's (1966) serial, exhaustive memory search model on the grounds that the model predicted a pattern for the variances not exhibited by the data.

Second, J. Theios and colleagues (Falmagne and Theios (1969), Theios (1973), Theios and Smith (1972), Theios, Smith, Haviland, Traupman and Moy (1973)) proposed a serial, self-terminating model, called the push-down stack, that accounted as well for Sternberg's mean data as did the serial, exhaustive search. Sternberg (1973) countered by developing two properties of distributions of search times that must hold in a broad class of self-terminating models. Denoting by $G_M(t)$ the probability of a response by time t when the memory set has M elements, then for $M < N$ he showed

$$(N/M)G_N(t) \geqslant G_M(t) \geqslant G_N.$$

He demonstrated violations of these inequalities in several bodies of data and so concluded that the push stack was not adequate.

The third example appears in tests of Ratcliff's (1978) memory retrieval model. This is a continuous random walk in which the diffusion rate is a Gaussian distributed random variable. The analytic development of this model is more complete than one might have anticipated given the difficulty statisticians have had in working out the response-time distributions for the better known discrete-time random walk (Laming (1968), Link (1975), Link and Heath (1975), Wald (1947)). In particular, explicit formulas are known for the distributions, and these have been fittet to the empirical distributions with considerable success as can be seen in the example shown in Figure 1.

SO, WHAT'S WRONG WITH EMPIRICAL DISTRIBUTIONS?

On the face of it, examples such as those just mentioned encourage one to the view that the distributions do indeed include a great deal of information beyond the means, and that it is essential for theoreticians to develop models adequate to account for the response-time distributions, as Ratcliff has done. And let me make clear that I firmly believe this to be true. Nevertheless, I fear that there may be serious problems that have not been fully confronted. Certain difficulties have been discussed rather fully in the literature on memory search, but others that may be at least as

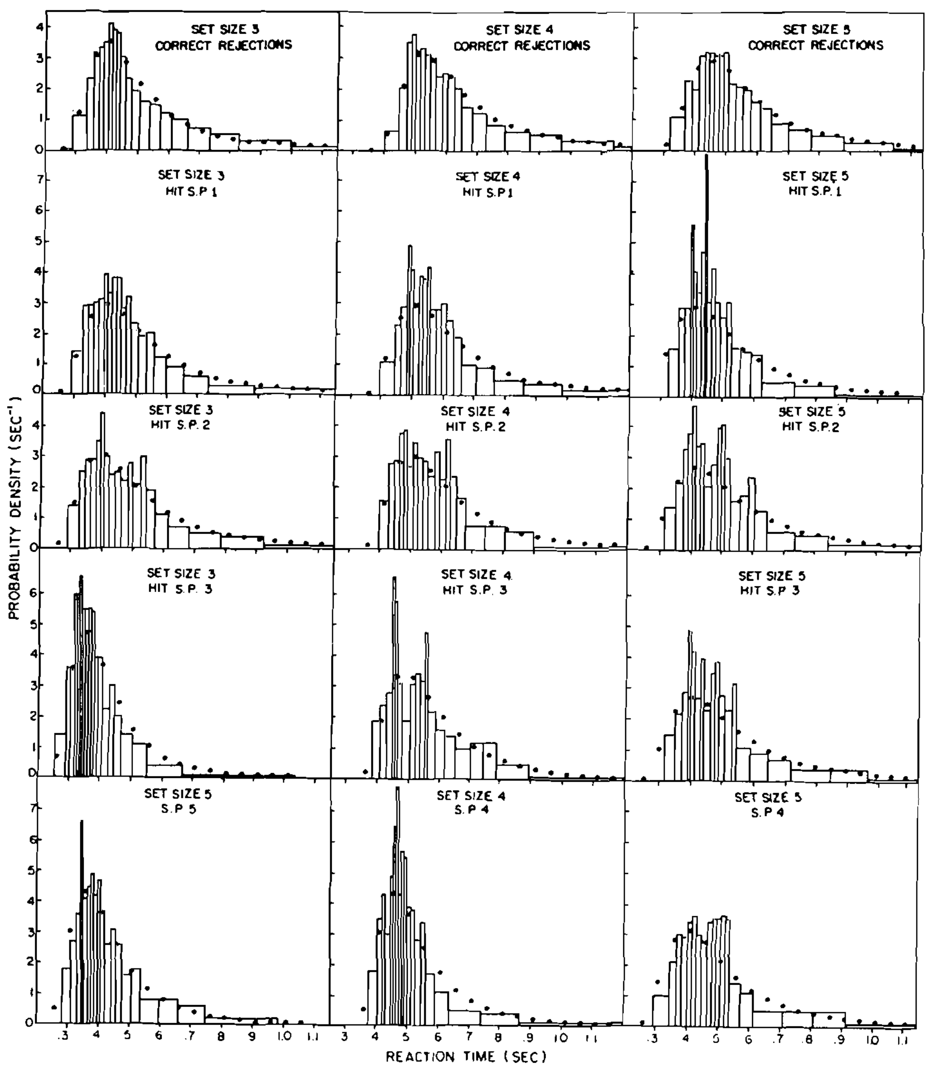

Figure 1

Histograms of response time distributions from a memory search experiment using three sizes for the memory set. The dots are theoretical predictions from Ratcliff's (1978) diffusion model. This is Figure 15 of Ratcliff (1978); reprinted by permission.

severe have received far less attention in that literature. Among the familiar criticisms, two are probably best known.

One is the fact that full knowledge of the distribution from one experimental paradigm does not uniquely implicate the process that gave rise to it. This possibility was perhaps first remarked on by Christie and Luce (1956), but the full development of the problem is due to J. Townsend (for a summary of the results, see his book with Ashby Stochastic Modeling of Elementary Psychological Processes (1983) or Chapter 12 of my book (Luce (1985)). These negative results do not, however, lessen the value of distributions in selecting the better of two somewhat specific models or even for firmly rejecting a particular model.

The other major criticism about the use of response times, even mean times, is the fact that there is some sort of tradeoff between speed and accuracy, and in experiments where several different conditions are run in blocked trials, the experimenter is in danger of being misled by just looking at response times. I do not have space to go into all of the issues involved, but I bring it up because I shall be concerned later about another aspect of the speed-accuracy tradeoff problem.

I take up a third concern that, despite the fact it has attracted rather less attention than the other two, I suspect is the most severe of the difficulties. It is the possibility that we cannot estimate the response time distribution because there isn't one! Rather, there are many and what we report as an estimate of a distribution really is an estimate of a probability mixture of many distributions. There are at least two, possibly related, sources to the mixture: sequential effects and local speed-accurary tradeoffs. Let me describe what we know about these problems.

SEQUENTIAL EFFECTS

To my knowledge, no one has really looked for and studied sequential effects in memory search designs. By contrast, they form a relatively major, if still ill-understood, topic in studies of absolute identification. Such studies are more psychophysical than cognitive in origin. In the out-and-out psychophysical literature on both absolute identification and the closely related magnitude estimation, careful attention has been given to sequential effects. These are observed by partitioning the responses to a given stimulus according to the stimulus-response histories leading up to trials on which that stimulus was presented. Whenever this has been done, sizable effects have been found. For example, in magnitude estimation we may plot the correlation between two successive responses as a function of signal separation on the corresponding trials, and typical results from Baird, Green, and Luce (1980) for estimates of area and loudness are shown in Figures 2 and 3. Other ways of studying the effect can be found in the work of G. Lockhead and his collegues (see Lock-

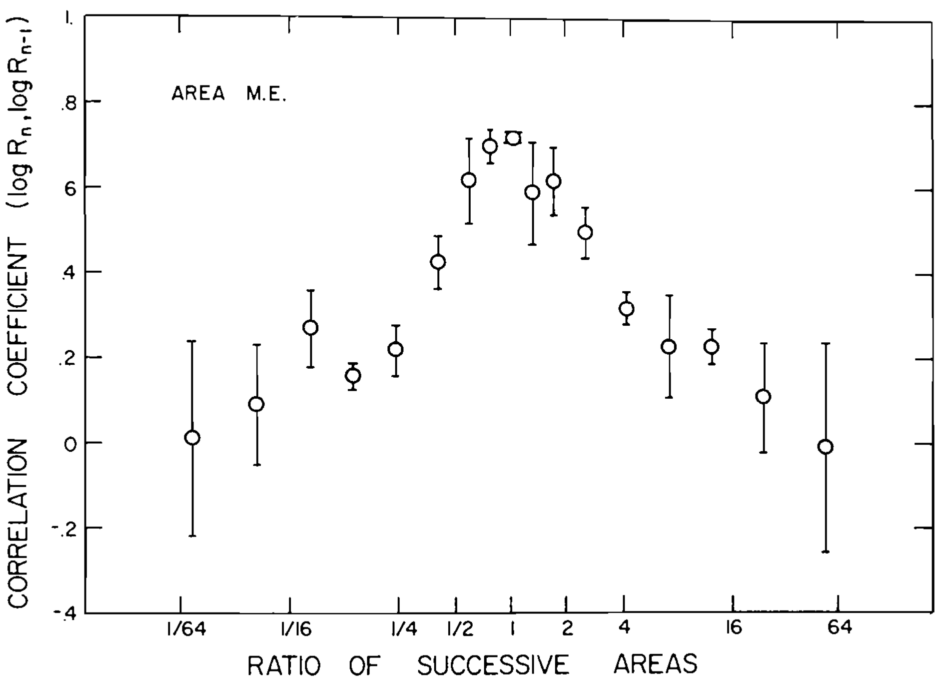

Figure 2

Correlations of the logarithm of magnitude estimates of area on successive trials as a function of the signal separation on these two trials. This is Figure 10 of Baird, Green, and Luce (1980); reprinted by permission.

head (1984) for a summary and the relevant references). More directly relevant, of course, are response time studies. A number of these are discussed by Kirby (1980). Probably the most visually clear demonstrations are those of mean response times as a function of previous histories, as reported by Laming (1968), Falmagne, Cohen and Dwivedi (1975), and Remington (1969). For example, Figure 4 shows the mean response time to one signal of a two-stimulus absolute identification experiment when the data are partitioned according to the previous stimulus history (but with the previous responses ignored). Note that a difference of 35 msec is found when histories of length four are taken into account;

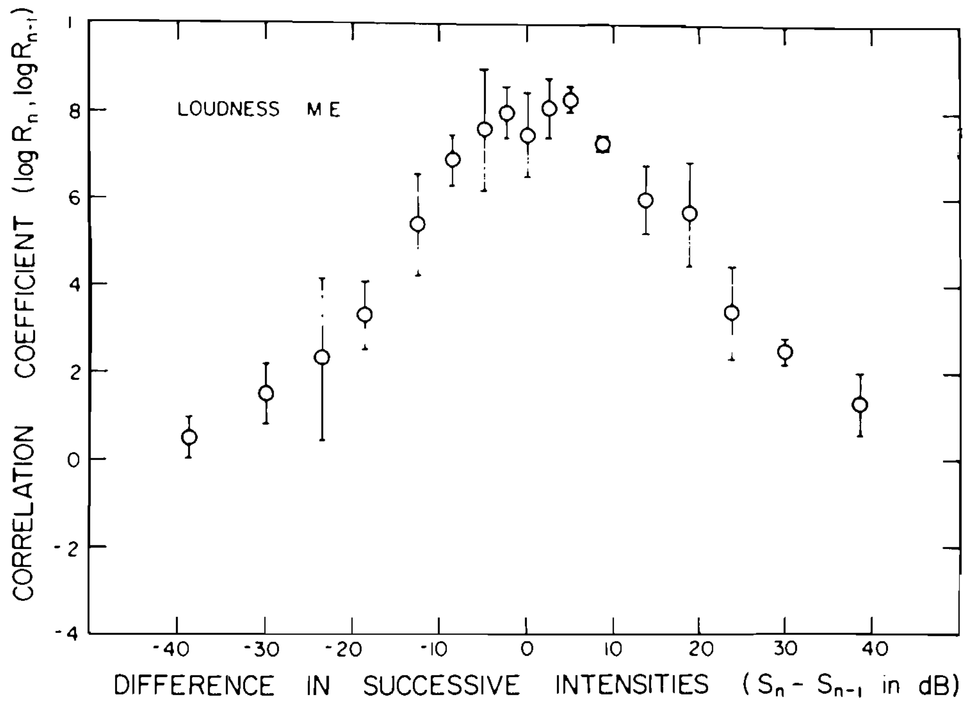

Figure 3

The same plot as Fig. 2 for loudness judgments. This is Figure 11 of Baird, Green, and Luce (1980), reprinted by permission.

this is approximately 12 % of the mean.

As Laming (1968) and Kirby (1976) and others have observed, there appear to be two quite distinct types of sequential effect depending upon the temporal spacing of the signals; this is discussed at some length by Kirby (1980) and in Section 6.5 of my book (Luce (1985)).

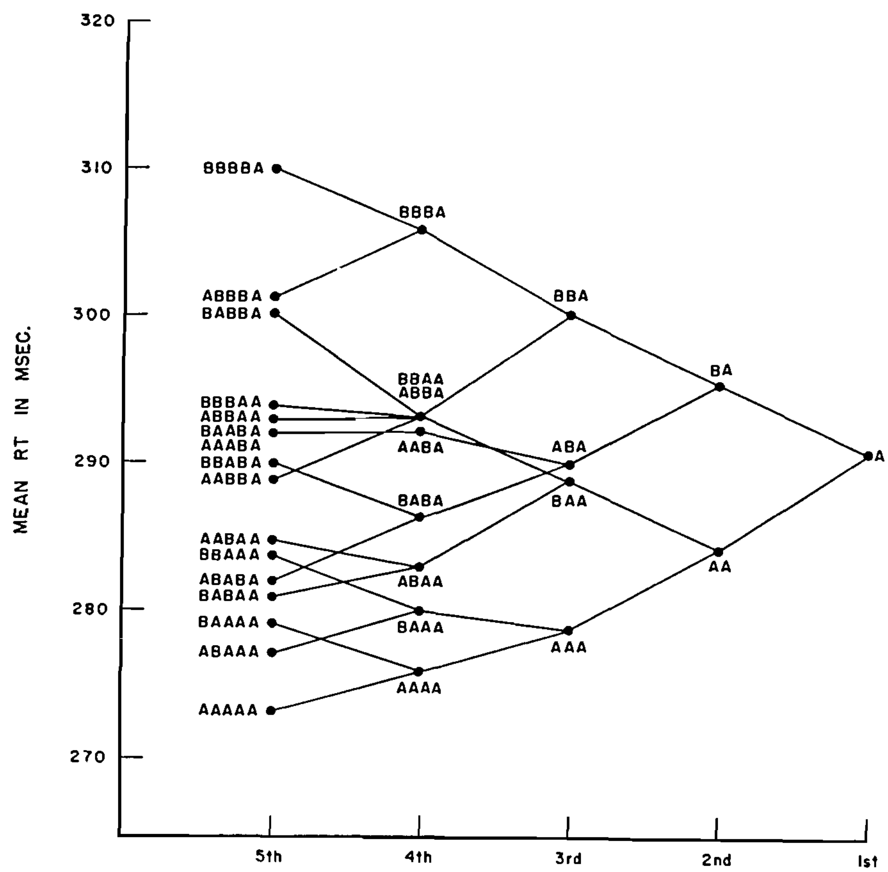

Figure 4

Mean response times to a particular signal, denoted A, in a two-stimulus, absolute identification experiment. The data are presented for several different partitions of the stimulus history. For example, the entry BBA means that it is the mean data for all trials on which A occurred and the other signal occurred on the two preceeding trials. This is Figure 2 of Remington (1969); reprinted by permission.

The main point that I wish to get across is that both in absolute identification and the closely related magnitude estimation studies, there are pronounced sequential effects. That fact implies that the observed distributions are, necessarily, mixtures of distributions and so are potentially misleading. For example, it is quite possible for the variance of each indicidual distribution to be smaller than the overall estimated variance. And certainly the shape of the mixture distribution need not have much relation to the shapes of the component distributions. The question we face is to understand what is causing these effects and either to model them or to perfect experimental designs in which the effects diappear. From a theorist's point of view, probably the latter is preferable, if possible, since trying to model the phenomenon is bound to create analytical problems, as is evident in the models so far proposed. However, as yet, no experimental procedure has been proposed for absolute identification that avoids sequential effects. Two attempts to model sequential effects in response times have been published as well as three others that ignore the times (the latter references are Luce, Bairs, Green and Smith (1980), Marley and Cook (1984), Treisman (1985); Treisman and Williams (1984)). One response-time model, due to Falmagne et al. (1975), is Markovian with internal states that correspond to two degrees of certainty about the stimulus. In the one case where they fit the model to a subject's data, the fit was excellent, but it has never been extensively tested, in part because it is complex. The other, due to Laming ((1968), (1978a), (1979b)) assumes that a discrete random walk model applies, but that the subject controls when information accumulation begins. If that starts prior to actual signal onset, then the initial "information" accumulated is from a noise source and so, in effect, it renders the starting point of the actual accumulation a random variable. The result is that errors are more likely and responses faster. Qualitatively, this model exhibits some features of the data that are otherwise difficult to understand.

All of these findings pertain to identification paradigms, and so they are not directly relevant to search designs. However, it seems to me that it is incumbent on those using distributional information from search designs to demonstrate empirically that there are no sequential effects. To my knowledge, no relevant data exist. That is a bit surprising. It probably is due in part to uncertainty about what at look for. In many search designs, each trial uses a new set of stimuli as the memory set, and so there is nothing as simple as repeated stimuli. We do not really know if actual stimulus repetition is a prerequisite for sequential effects. Other possibilities seem just as likely. For example, trials can be classed according to whether or not the target stimulus is or not in the memory set, and so data can be partitioned according to histories in terms of what was the appropriate response. It may be that repetition of a trial type results in faster responding to another occurrence of that type and slower to the other type. We simply do not know.

I turn now to the second source of difficulty with empirical distributions.

LOCAL SPEED-ACCURACY TRADEOFFS

Usually when a speed-accuracy tradeoff is mentioned, it refers to the fact that as the experimental conditions are changed, the subject may make adjustments in the variables under his control as so to alter the mix between speed and errors. But there is another possibility, namely, that within a single experimental condition, the subject may from time to time make adjustments in the speed-accuracy tradeoff. I am aware of only one type of study in which this has been demonstrated, namely, those with such severe time pressure that the subject is driven to fast guesses, i.e., on some trials the response is made to signal onset, as in simple reaction time, and the error pattern conform to change guessing. Swensson (1972) and Swensson and Edwards (1971) were able to partition their data into runs of fast guesses interleaved with runs of considerably slower trials exhibiting much more accurate judgments. This is shown in Figure 5. Whether something comparable happens under less time pressure, so that the subject shifts during an experimental run to various places on the speed-accuracy tradeoff function or region (see Pike and Dalgleish (1982), Weatherburn (1968), and Weatherburn and Grayson (1972)), has not been systematically investigated. One hopes that it does not occur, but I am not very optimistic.

The possibility of local changes along the speed-accuracy tradeoff raises the question whether the sequential effects are actually manifestations of such local changes, in which case there would be just one thing to be concerned about. Although some sequential effects may arise in this fashion, it is doubtful if it is the sole source. Laming (1979a) examined the behavior of subjects in absolute identification on the trials following an error. He found that mean response times often recovered to their average value after a single trial whereas errors were still not completely recovered after a lag of five trials.

Once again, what we know here concerns identification paradigms, not search ones. And since most search studies attempt to keep the error rate very low, it is usually assumed that speed-accuracy tradeoffs are not at issue. In fact, this assumption is highly suspect, and especially so when we are dealing with distributions. The problem is that many speed-accuracy functions have a steep derivative when the error rate is low, so relatively small changes in that rate correspond to quite substantial shifts in the response time distribution. Thus, if during an experimental run the subject adjusts his or her location on the speed-accuracy curve so as to maintain an overall prescribed error rate, this will be reflected in increased apparent variability in the response time distribution.

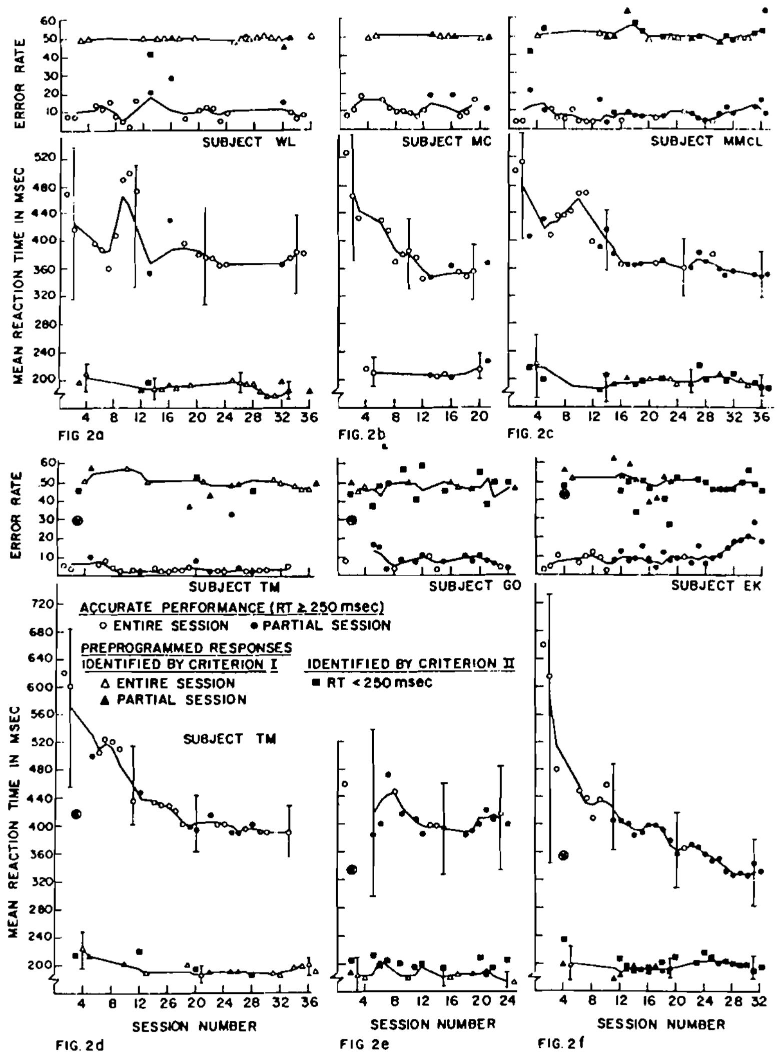

Figure 5

Data over trials from six subjects in an absolute identification experiment run under time pressure. The trials were partitioned, according to two different criteria, into responses that are believed to be preprogrammed (or fast guesses) and those in which the subject attempts to identify the signal. Note that the preprogrammed trials are approximately at chance performance and at a speed normally found for simple reaction times; the other type has an error rate of about 5 to 10 % and the mean time is about 200 msec slower. This is Figure 2 of Swensson (1972); reprinted by permission.

CONCLUSION

My caution is a simple one. Sequential effects and local speed-accuracy tradeoffs are ubiquitous and sizable in identification paradigms, and without evidence to the contrary we must assume they exist in memory search designs, even when overall accuracy rates are held below 5 %. If such effects do exist, and certainly this should be studied empirically, then estimates of overall response time distributions are bound to be misleading, and so attempts to evaluate models in terms of them must be treated with a degree of skepticism. I do not mean this as an argument against using distributions to evaluate models, which I believe is essential, but as a prood both to experimentalists and to theorists. The former should determine if such effects exist and, if so, try to devise procedures to eliminate them. The latter should assume these effects will be found and that experimentalists will be unable to eliminate them (possibly because they are deeply imbedded in the search process), and so theorists should attempt to incorporate them into the models. I am aware that this is not easy to do, in large part because the number of free parameters tends to profilerate, a problem that is always getting out of hand when we attempt to model even only moderately complex processes. But to act as if trial-by-trial changes are not there is, I am afraid, overly optimistic and potentially quite misleading.

REFERENCES

/1/ Baird, J.C., Green, D.M., and Luce, R.D., Variability and sequential effects in cross-modality matching of area and loudness, Journal of Experimental Psychology: Human Perception and Performance 6 (1980) 277-289.

/2/ Christie, L.S. and Luce, R.D., Decision structure and time relation in simple choice behavior, Bulletin of Mathematical Biophysics 18 (1956) 89-112.

/3/ Falmagne, J.-C., Cohen, S.P., and Dwivedi, A., Two-choice reactions as an ordered memory scanning process, in: Rabbitt, P.M.A. and Dornic,S. (eds.), Attention and Performance V (Academic Press, New York, 1975).

/4/ Falmagne, J.-C. and Theios, J., On attention and memory in reaction time experiments, Acta Psychologica 30 (1969) 316-323.

/5/ Kirby, N.H., Sequential effects in two-choice reaction time: Automatic faciliation or subjective expectancy?, Journal of Experimental Psychology: Human Perception and Performance 2 (1976) 567-577.

/6/ Kirby, N., Sequential effects in choice reaction time, in: Welford, A.T. (ed.), Reaction Times (Academic Press, London, 1980).

/7/ Laming, D.R.J., Information Theory of Choice Reaction-Times (Academic Press, London, 1968).

/8/ Laming, D.R.J., Choice reaction performance following an error, Acta Psychologica 43 (1979a) 199-224.

/9/ Laming, D.R.J., Autocorrelation of choice-reaction times, Acta Psychologica 43 (1979b) 381-412.

/10/ Link, S.W., The relative judgment theory of two choice response time, Journal of Mathematical Psychology 12 (1975) 114-135.

/11/ Link, S.W. and Heath, R.A., A sequential theory of psychological discrimination, Psychometrika 40 (1975) 77-105.

/12/ Lockhead, G.R., Sequential predictors of choice in psychophysical tasks, in: Kornblum, S. and Requin, J. (eds.), Preparatory States and Processes (Erlbaum, Hillsdale, 1984).

/13/ Luce, R.D., Response Time: and their role in inferring elementary mental organization (Oxford Press, New York, 1985).

/14/ Luce, R.D., Baird, J.C., Green, D.M., and Smith, A.F., Two classes of models for magnitude estimation, Journal of Mathematical Psychology 22 (1980) 121-148.

/15/ Marley, A.A.J. and Cook, V.T., A fixed rehearsal capacity interpretation of limits on absolute identification performance, British Journal of Mathematical and Statistical Psychology 37 (1984).

/16/ Pike, A.R. and Dalgleish, L., Latency-probability curves for sequential decision models: A comment on Weatherburn, Psychological Bulletin 91 (1982) 384-388.

/17/ Ratcliff, R, A theory of memory retrieval, Psychological Review 85 (1978) 59-108.

/18/ Remington, R.J., Analysis of sequential effects in choice reaction times, Journal of Experimental Psychology 82 (1969) 250-257.

/19/ Schneider, W. and Shiffrin, R.M., Controlled and automatic human information processing: I. Detection, search, and attention, Psychological Review 84 (1977) 1-66.

/20/ Sternberg, S., High speed scanning in human memory, Science 153 (1966) 652-654.

/21/ Sternberg, S., Evidence against self-terminating memory search from properties of RT distributions. Paper presented at Psychonomic Society, St. Louis (November,1973).

/22/ Swensson, R.G., The elusive tradeoff: Speed versus accuracy in visual discrimination tasks, Perception & Psychophysics 12 (1972) 16-32.

/23/ Swensson, R.G. and Edwards, W., Response strategies in a two-choice reaction task with a continuous cost for time, Journal of Experimental Psychology 88 (1971) 67-81.

/24/ Theios, J., Reaction time measurements in the study of memory processes: Theory and data, in: Vower, G.H. (ed.), The Psychology of Learning and Motivation 7 (Academic Press, New York, 1973).

/25/ Theios, J. and Smith, P.G., Can a two-state model account for two-choice reaction time data?, Psychological Review 79 (1972) 172-177.

/26/ Theios, J., Smith, P.G., Haviland, S.E., Traupmann, J., and Moy, M.C., Memory scanning as a serial self-terminating process, Journal of Experimental Psychology 97 (1973) 323-336.

/27/ Townsend, J.T. and Ashby, F.G., Stochastic Modeling of Elementary Psychological Processing (Cambridge University Press, Cambridge, 1983).

/28/ Treisman, M., The magical number seven and some other features of absolute judgment: An application of criterion-setting theory. Journal of Mathematical Psychology 29 (1985).

/29/ Treisman, M. and Williams, T.C., A theory of criterion setting. I. Sequential dependencies, Psychological Review 91 (1984) 68-111.

/30/ Wald, A., Sequential Analysis (Wiley, New York, 1947).

/31/ Weatherburn, D., Latency-probability functions as bases for evaluating competing accounts of the sensory decision process, Psychological Bulletin 85 (1978) 1344-1347.

/32/ Weatherburn, D. and Grayson, D., Latency-probability functions: A reply to Pike and Dalgleish, Psychological Bulletin 91 (1982) 389-392.