

Low-variance stimulus-response latencies: Deterministic internal delays?

ALFRED B. KRISTOFFERSON
McMaster University, Hamilton, Ontario, Canada

When special procedures are used to minimize S-R latency variance, all responses fall within a distribution which has a standard deviation near 10 msec. This minimum SD is the same whether the mean latency is at the simple RT limit or as much as 400 msec longer than that limit. Over this range, the latency distribution is everywhere the same, symmetrical and highly-peaked and not typical of RT. Above a mean latency of 550 msec, variance increases as the mean increases in the way that would be expected if SD/M were constant for the delay in excess of 550. By way of interpretation, it is proposed that there are internal time delays which can be inserted into the S-R chain. These delays can be adjusted to any value between 0 and about 400 msec, but once set, they can be deterministic. Other considerations are discussed which suggest that the deterministic time delays are in the afferent part of the S-R chain.

The simple reaction time distribution can easily be shifted right or left along the time axis. When it is shifted some distance toward longer latencies, the responses are often called time estimations rather than simple reactions even though no basis for a qualitative distinction has been established.

Shifting of latency distributions has been accomplished in several different ways. Snodgrass, Luce, and Galanter (1967) used narrow payoff bands centered at various times closely following the action stimulus. Ollman and Billington (1972) told their subjects to try to produce a certain target latency and informed them of their latency after each trial. Naatanen, Muranen, and Merisalo (1974) presented a second stimulus a fixed time after the action stimulus and asked that the response be made in synchrony with the second stimulus. The latter method, response-stimulus synchronization, is the one used in the experiments reported in this paper. With it, the second stimulus, which is brief, serves as the payoff band served in the Snodgrass et al., experiments.

Snodgrass, Luce, and Galanter (1967) found that the latency distribution will follow the payoff band along the time axis in a quite precise way. They also found that the variance of the distribution is minimal for some band position within the usual reaction time range; shifting the distribution away from this optimal position in either direction increases the variance. They explained their results by postulating

This research was supported by Grant A7919 from the National Research Council of Canada. The paper was written while the author was on leave at the MRC-Applied Psychology Unit, Cambridge, England. I thank the members of the Unit and its director, Alan Baddeley, for their generous hospitality, and Alan Wing also for many valuable discussions related to this paper. Subject G.H. is Gordon Hopkins, who is still cheerful despite this ordeal; his contribution was substantial.

that the distributions consist of mixtures of three kinds of stimulus-response events: (1) true RTs triggered by the action stimulus, (2) time estimation responses triggered by the warning signal, and (3) time estimation responses triggered by the action stimulus. By positioning the distribution in the location which minimizes its variance, they hoped to maximize the proportion of true RTs in the mixture and reveal the true reaction time distribution. It should be possible to do this, they argued, because time estimates are less accurate than true RTs; in particular, time estimates have a larger variance and the variance increases with the mean so that the SD is about 10% of the mean. That this is so had been shown by Woodrow (1930, 1933) and was confirmed by Snodgrass et al. for a range of time estimation means from about 600 to 5,000 msec.

Reaction time instructions call for a rapid response while time estimates require that the response be delayed for some additional time. It may be that they differ only with respect to the insertion of an additional delay into the stimulus-response chain. The inference that the SD of such internal delays increases as the mean increases is supported by many kinds of experiments. The ones mentioned above, by Woodrow, Snodgrass et al. and Naatanen et al., are examples. Other examples are Treisman (1963), for time interval production and reproduction, and Bartlett and Bartlett (1959), for synchronizing a response with one of a train of brief stimuli. Even duration discrimination can be interpreted as involving a stimulus-triggered internal time delay (Kristofferson & Allan, 1973), and Getty (1975) has recently shown that the ratio of the SD of the psychometric function for duration discrimination to its mean is approximately constant at 0.06 over the range from 200 to 2,000 msec. As a final example, Wing and Kristofferson (1973a) have isolated the variance

in interresponse times which is due to a hypothetical central clock which controls the average times between successive, regularly spaced responses. The result of their Experiment A, averaged over subjects and plotted here as SD vs. the mean, is shown in Figure 1. Again, the conclusion that the SD is a fixed percentage of the mean, here about 2.3% is not unreasonable.

These experiments all show that longer internal delays are more variable. Some of them support the stronger postulation of a principle resembling Weber's law, as Getty (1975) has described his model. The variety of experimental operations suggests that the principle may have considerable generality.

Snodgrass et al., quite properly did not conclude that they had experimentally isolated the RT distribution. In fact, while they did not always report obtained SDs, the values that they do report are no lower than those observed in many other studies, even when the payoff band is situated so as to minimize the variance. Their smallest SDs were between 20 and 25 msec. Such values are typical in carefully controlled experiments regardless of whether the action stimulus is auditory or visual. Smaller values have sometimes been reported, but in such cases some latencies have been excluded from the analysis. Saslow (1974), also using the payoff band technique along with very extensive training procedures, obtained the smallest SDs that this author has encountered in the literature, the smallest single value being 16 msec. However, most of Saslow's values also exceeded 20 msec. Experiments which use variable foreperiods yield even larger SDs, as for example the extensive experiments by Green and Luce (1971).

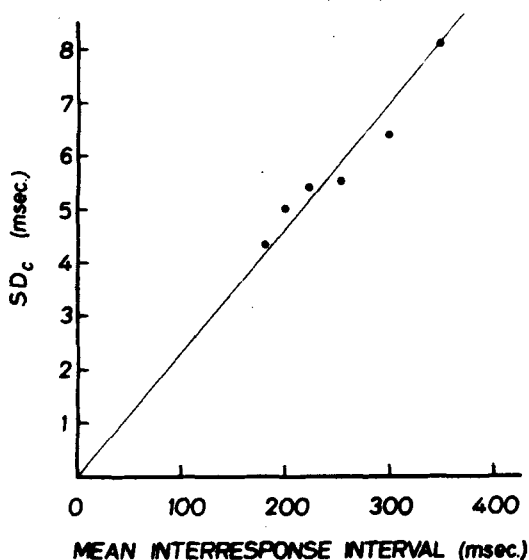


Figure 1. Standard deviation vs. the mean for intervals generated by theoretical timekeeper. The model is from Wing and Kristofferson (1973a), and the data are from their Experiment A. Average parameters for four subjects. The line represents $SD/M = 0.023$.

Ollman and Billington (1972) and Kornblum (1973) have stated models of simple reaction time which have much in common with the theory of Snodgrass et al. (1967). Their models are intended to permit them to extract true RT distributions from their data. Kornblum's one reported subject yielded a true distribution with a SD of 27 msec. Ollman does not state the necessary figures, but his graphs indicate true SDs no smaller than 25 to 40 for his three subjects. These values are no lower than the raw values reported by others.

Thus, the smallest simple reaction time SD seems to be near 20 msec. Conjectures about the form of the RT distribution and about elementary timing mechanisms are based upon such data and might need to be reconsidered if a procedure is found which significantly reduces the variability.

Snodgrass, Luce, and Galanter (1967) considered the form of the distribution in great detail and arrived at no certain conclusion. No one general form was found to be wholly satisfactory. There is still no agreement on a single general form but there is some agreement about certain general features of the distribution as Snodgrass et al. described them. The distributions are asymmetrical, usually positively skewed, and they have sharply peaked modes and high tails. Distributions of time estimates, on the other hand, are usually symmetrical, a finding which might encourage a qualitative distinction between time estimates and RTs.

There is, however, a recent exception to the generalization given above about the form of the RT distribution. Saslow (1974), who, as mentioned earlier, reports the smallest variability, also found that the latency distribution was symmetrical, even when the payoff band was within the normal RT region.

The present paper is concerned with three questions arising from the discussion above: (1) The relationship between the mean and the variability of internally timed delays; (2) the minimum variability of S-R latencies; and (3) the form of the distribution when variability is reduced below previously observed levels. The response-stimulus synchronization method is used, as Naatanen et al. (1974) have done and as Bartlett and Bartlett (1959) employed in a somewhat more complex form.

METHOD

A trial is diagrammed in Figure 2. It begins with a 100-msec ready signal followed by P_1 , a 10-msec auditory pulse, 1,000 msec later. After another delay, the P_1P_2 interval, P_2 is presented. P_2 is identical to P_1 , and the interval between them is fixed in duration throughout a session. The subject is asked to depress the response button in synchrony with P_2 . Thus, he must trigger his response some time before P_2 if he is to hit P_2 , and it is intended that the response will be a delayed response triggered by P_1 when the P_1P_2 interval is greater than the simple RT limit.

The moment that the response strikes the button is indicated by the arrow in Figure 2, and the response latency, the time between

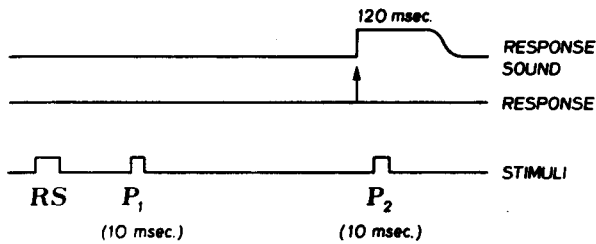


Figure 2. Schematic of one synchronization trial. P_1P_2 interval is onset to onset. Response latency is measured from the onset of P_1 .

the onset of P_1 and the electrical contact of the button, is measured to the nearest millisecond and stored by the computer.

Striking the button produces a *response sound* which the subject always hears. The loudness of the response sound is about the same as that of P_2 . It has a rise time of approximately 1 msec, a constant amplitude for 90-100 msec, and a total duration of about 120 msec. It is caused by the finger striking the button and the properties of the metal box on which the button is mounted. The subject receives immediate *perceptual feedback* from the relationship between the response sound and P_2 . They are qualitatively different and specifically identifiable.

P_1 and P_2 are 1,000-Hz sinusoids gated at zero-crossing with a rise-decay time of 2.5 msec. Their loudness is 68 dB re 0.0002 dynes/cm² when on continuously. They are delivered over earphones, while the RS originates from a speaker in the room. The RS is lower in pitch and is noisy.

The response button has a diameter of 1.1 cm and protrudes 1 cm out of the top of the response box, which lies flat on a table top and has the dimensions 43 × 25 × 5 cm. The subject sits with his elbow on the table and his index finger on the response button. The button must be depressed 0.5 mm to make contact and a static weight of 80 g is required to do that.

The time between P_2 and the RS for the next trial is fixed so that trial duration depends upon the P_1P_2 interval. A block is 100 consecutive trials and four blocks comprise a session, with a rest of 1.5 min between blocks. A session lasts 45 min or less, hence the average trial duration is about 6 sec. But within a session, the P_1P_2 interval is fixed and the trials occur at perfectly regular intervals. Time of day is controlled and subjects ran every day, except weekends. Subject AK participated in two sessions per day.

The first block of 100 trials is discarded from every session, and each session therefore yields 300 trials. The subject is permitted to skip a trial if he is distracted for any reason, but that rarely happens and 300 latencies are usually obtained.

Feedback

No payoff scheme is used. The major feedback is provided by the response sound in relation to P_2 . In addition, the subject is informed of his means and variances after each session and he is urged to minimize his variance and his synchronization error, which is defined as the time between his mean latency and the onset of P_2 , with a negative sign denoting early responding.

The perceptual feedback is complex, but it is remarkably effective. When the response is late by more than some small amount, a gap can be heard between P_2 and the response sound. When the response is early, its sound overlaps P_2 , unless it comes very much too early. when the response occurs during P_2 , and only then, P_2 is immediately terminated. Consequently, when the response hits during the first few milliseconds of P_2 , the loudness of P_2 is reduced so that it cannot be heard, and such a successful response is immediately signaled by the absence of P_2 . Therefore, the perceptual feedback consists of a set of different events which occur on different proportions of trials and which must depend upon, for example, the duration of P_2 and the characteristics of the response latency distribution. The research necessary to specify these events more exactly has not yet been done.

The Response

The button is depressed by a downward movement of the index finger with the hand cupped and the index finger supported by the thumb and middle finger. The movement is small in amplitude, since the finger remains on the button, but the elbow is the pivot and some wrist flexion is involved. No attempt has yet been made to measure parameters of the movement, but it is complex and it depends to some extent upon the P_1P_2 interval. When the interval is long, muscular involvement is reported by the subject to begin shortly after P_1 . The finger moves very little, since it remains on the button. The response which strikes the button seems to be a rapid impulse superimposed on the background of muscular involvement or tension. As the P_1P_2 interval is shortened, a point is reached beyond which the initial response to P_1 cannot be made and the entire response must be only the final impulsive one. This change in the form of the total movement occurs at a P_1P_2 interval of about 300 msec.

EXPERIMENT I

The first experiment was undertaken by the author as subject for the purpose of determining roughly how much practice would be required with a single P_1P_2 interval to achieve stable performance and some hope of stationarity. The P_1P_2 interval was set at 940 msec and kept fixed for 60 sessions.

Figure 3 shows the effect of practice upon the mean and the variance of the latencies with the data grouped into five-session blocks. Each point in the figure summarizes 1,500 responses. The mean shows little change, being quite stable after 20 sessions. The modal response occurs about 10 msec before P_2 , a result which is typically found with other subjects as well.

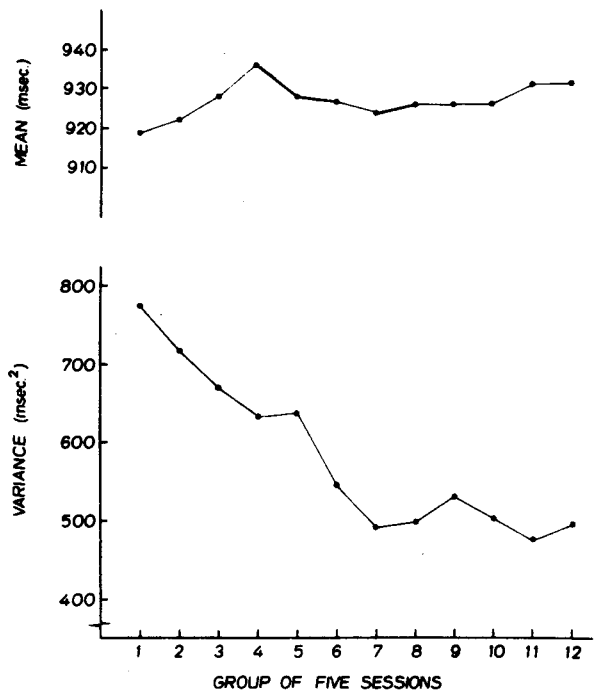


Figure 3. Mean (top) and within-block variance (bottom) for the 60 consecutive sessions of Experiment I. P_2 occurred 940 msec after P_1 .

The latency variance diminishes progressively and slowly for 30 sessions. A week's break was taken after Session 20, and the point for Group 5 is slightly elevated. For the final 30 sessions, the variance is stable. The quantity plotted here is the mean within-block variance. It averaged 497 msec² during the final 30 sessions, corresponding to a SD of 22 msec, very close to the minimum known for RT.

It is unlikely that additional practice would reduce the variance any further. But there are other factors which do reduce it further and do so almost by a factor of 2, as the next experiment will show.

The latency distribution obtained during the final seven sessions is shown in Figure 4. All 2,100 latencies fall within a 150-msec window centered on the mean of 930, and 97.5% of them fall within a 100-msec window. The total variance of this distribution is 505 msec². Since the within-block variance is 497, as noted above, it is apparent that changes between blocks and between sessions are negligible. The SD is 2.4% of the mean, and the error of synchronization is 1.1%.

The time at which P_2 occurs is given by the black rectangle. Most responses occur before P_2 and no responses occur more than 60 msec after P_2 . The shape and placement of the distribution give no hint of a direct effect of P_2 on any of the responses and encourage the view that the responses are triggered only by P_1 .

The conclusion that time estimation distributions are symmetrical is confirmed by Figure 4. This one is symmetrical and it has a sharp peak. It also has straight sides and narrow skirts. To say that it is the convolution of an isosceles triangle (Kristofferson & Allan, 1973), having a base of 100 msec, with a normal distribution having a SD of 9 msec would be a sufficient description.

EXPERIMENT II

Despite the excessive practice required to obtain the distribution of Figure 4, the distribution is not rigidly fixed. This experiment shows that the distribution can be moved without losing the variance reduction gained in Experiment I, providing it is moved by changing the P_1P_2 interval in small steps. The purpose is to study the parameters of the distribution as its mean is varied down to the RT limit. An alternative is to repeat Experiment I at each of many P_1P_2 intervals. Such would be necessary because a large change in the mean would surely lose the practice gain, or most of it.

Accordingly, for Session 61, the P_1P_2 interval was reduced by 10 msec to 930 msec, and it was further reduced by 10 for each subsequent session until the RT limit was reached at 160 msec (the descending series). Then the interval was increased in steps back to the starting point (the ascending series). For the ascending series, the step sizes were larger and of

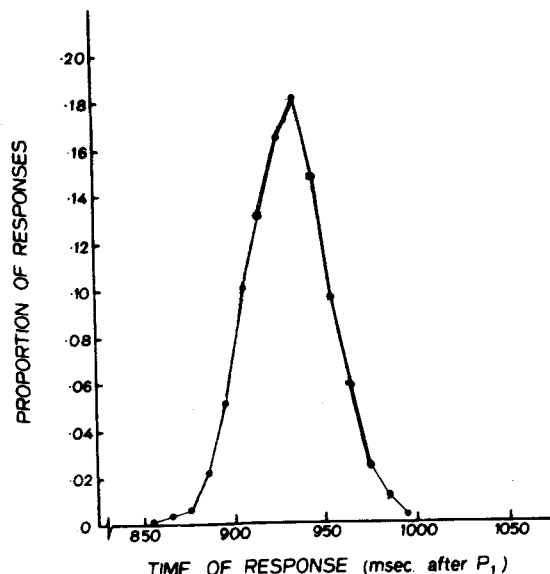


Figure 4. Frequency distribution of the 2,100 response latencies from the final seven sessions of Figure 3. Bin size = 10 msec. Rectangle marks time of occurrence of P_2 . Mean 930. SD = 22.

various sizes, but from 3 to 10 sessions were conducted at each before moving to the next higher step.

The mean latency can be shifted rapidly and with impressive precision. A change in the P_1P_2 interval of 10 msec induces a change in the mean latency of almost the same amount, and it does so within the first block of the session, even when the overall interval is over 900 msec.

Shifting the mean downward did not lose the practice gain. In fact, the variance *decreased* rapidly and by a large amount below the already low level reached in Experiment I. This can be seen in Figure 5, in which the solid line represents the descending series. The decrease in variance continued until the mean of the distribution reached 550 msec, the upper bound of Region II. Through Region II, the variance remained *constant*. When the mean crossed the 300-msec bound, the variance rapidly increased, and at 160 msec, it reached the level typical of simple RT.

The dashed line in Figure 5 gives the result for the ascending series for all sessions at each step except the first session. Overall, the variances are lower than they were for the descending series, which probably means that the practice gain was slightly disrupted by the shifts in the descending series. The general form of the function for the ascending series is similar to that for the descending series except at the upper end. Returning to Region IV after the experience with shorter intervals, the subject's variance is only about half what it was at the end of Experiment I.

Region II is of special interest because within it latency variance is minimal and latency variance is independent of mean latency. Figure 6 amplifies Region II, replotting the data from Figure 5 and

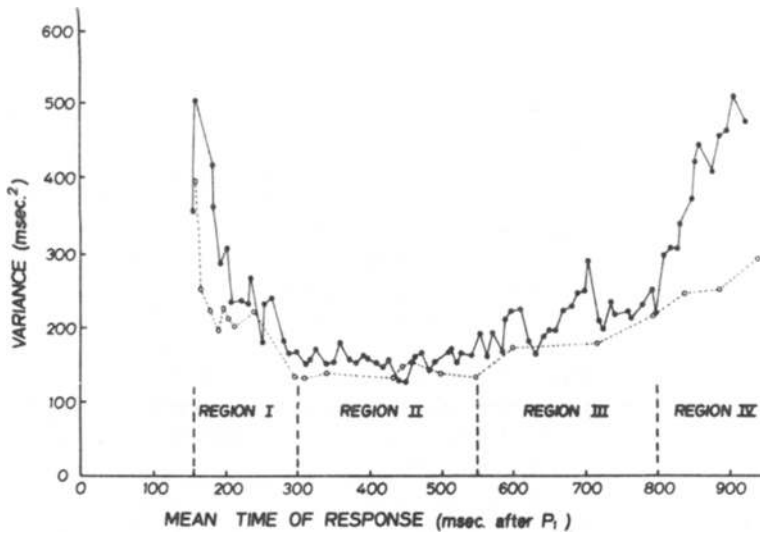


Figure 5. Mean within-block variance as a function of mean latency in Experiment II. Filled circles are the descending series with 300 latencies per point. Open circles are the ascending series sessions after Sessions 1 at each point, with 900 or more latencies per point. The bounds of the regions were set from these data for discussion purposes and do not have a priori significance.

adding the data from the first sessions at each step for the ascending series. The average slope for the three lines is $-0.003 \text{ msec}^2/\text{msec}$.

The latency distributions within Region II do not differ from each other in any discernible way. Distributions from the two ends of the region, 310 and 550, are approximately superimposed in Figure 7 by equating their modal bins. They are very similar, and the total of 1,500 responses all fall within a 70-msec window. They are also very similar in shape to the distribution of Figure 4, and the same general description applies here too. Figure 7 differs from Figure 4 in that the implied isosceles triangle has a base of 50 rather than 100 msec.

The error of synchrony depends upon the P_1P_2 interval only to a slight extent. Figure 8 shows that the error is generally negative and less than 10 msec. It is positive near the RT limit and nearly zero in the middle of the range.

EXPERIMENT III

Why does the variance increase as the RT limit is approached in Region I? As noted above, in Method, crossing the upper bound of Region I is accompanied by a change in the form of the response movement. Perhaps the training received in the upper regions is increasingly inappropriate as the RT limit is approached. The hypothesis that training which is specific to Region I might flatten the function so that Regions I and II are both horizontal is suggested by the flatter slope seen for the ascending series in Region I in Figure 5.

On the other hand, it is possible that simple RT must have *more* variability than time estimation responses because of the close temporal proximity of R and S. Having just received an input might somehow interfere with triggering a response. The variance near the RT limit in Figure 5, after all,

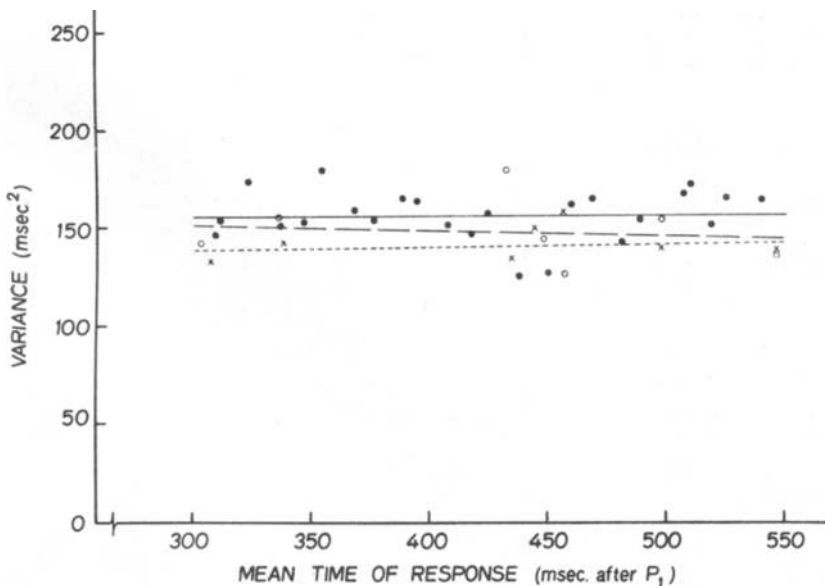


Figure 6. Independence of variance and mean within Region II. Filled and open circles are from Figure 5. Crosses are the first session at each point during the ascending series. The three least-squares lines have slopes of 0.004, -0.029 , and $0.017 \text{ msec}^2/\text{msec}$. Mean obtained variances are 157, 149, and 142.

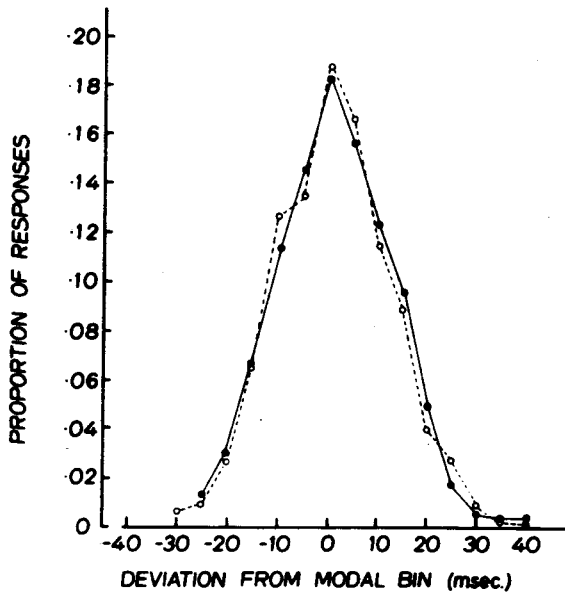


Figure 7. Superimposed frequency distributions from the two ends of Region II. Solid line: $M = 547$, $SD = 12$, $N = 600$. Dashed line: $M = 307$, $SD = 12$, $N = 900$. Subject A.K.

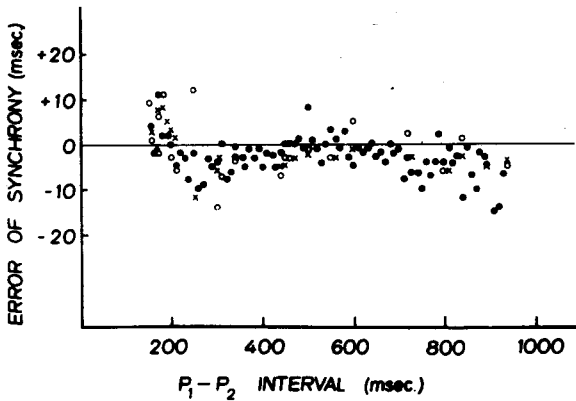


Figure 8. A negative error of synchrony is the mean time by which the response precedes the onset of P_2 . The error is shown here with the same symbols as before for all the data of Experiment II.

within the range of variability for well-practiced simple RT.

Experiment III was done with a fresh subject for the purpose of exploring Region I in greater detail. The subject, G.H., was a young male graduate student who had detailed knowledge of this field of research.

Fifteen sessions of simple RT were done first with the procedure described above, in Method, except only that P_2 was omitted. The results are displayed in Figure 9.

The RT variance decreased progressively for about nine sessions (3,600 trials) and fluctuated around a level of 350 msec^2 during the final five sessions ($SD = 18.7$). This result is very similar to those reported by Saslow (1974) in that the SD is unusually

small and the distribution is symmetrical. Of the 1,500 responses from the final five sessions, 6 were between 60 and 100 msec and 6 were between 220 and 250 msec. The remainder were between 100 and 200 msec.

Beginning with Session 16, P_2 was put in at a P_1P_2 interval of 170 msec so that it would occur a few milliseconds following the modal RT response, and synchronization trials began and continued with this interval for 29 sessions. The upper curve in Figure 10 shows how the latency variance changed during the 29 sessions. For the first few sessions it was greater than it had been, but it rapidly decreased to a level of 126 msec^2 during the final nine sessions.

Next a scan out and back over Region I was conducted, changing the P_1P_2 interval by 10 msec per session out to 310, running eight sessions at 310, and returning by steps of 10 to 170. Figure 11 gives the result. The variance increased slightly toward the upper bound of Region I instead of decreasing as it had for subject A.K.

The small increase in variance in Figure 11 might be due to moving further from the locus of initial training, so the next step was to set the P_1P_2 interval at 310 msec and do further practice at that point.

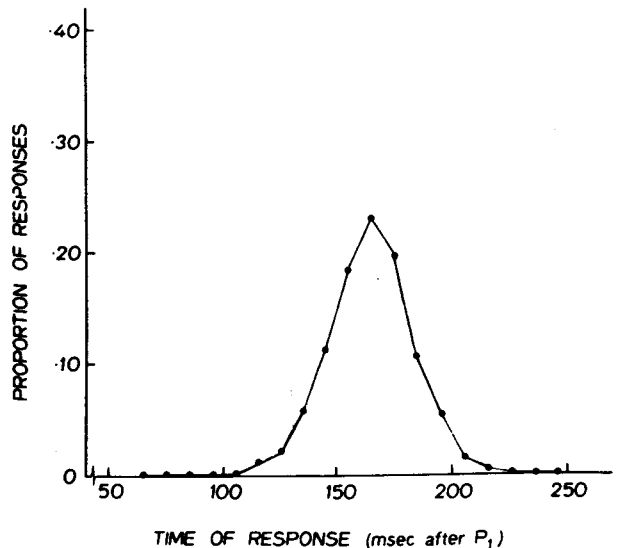
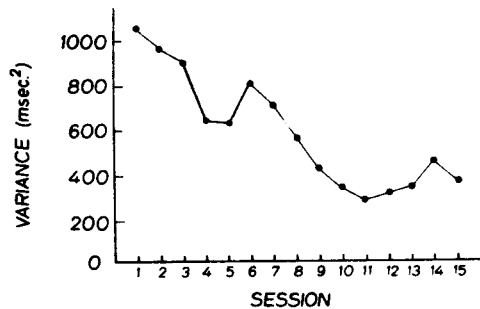


Figure 9. Change in simple RT variance over 15 sessions for subject G.H. (top). Below is the frequency distribution of the 1,500 response of the final five sessions.

The curve labeled 310 in Figure 10 shows again a slow decrease in variance. During the final five sessions, the mean within block variance was 107 msec² and the total variance was 112 msec². Nine sessions were then run at 360, followed by 11 sessions at 410. These are also drawn in Figure 10. During the last five sessions of each, the variance at 360 was 103 and at 410 it was 125 msec².

Latency distributions for G.H. are graphed as Figure 12. They have SDs of 11, 10, 10, and 11 msec from left to right. All responses occur within narrow windows, the window size being 95 msec for the left-most distribution and 75 msec for each of the other three. In shape, the distributions conform to the description given earlier for subject A.K.

It seems probable that in the limit latency variance is independent of mean latency over both Region I and Region II.

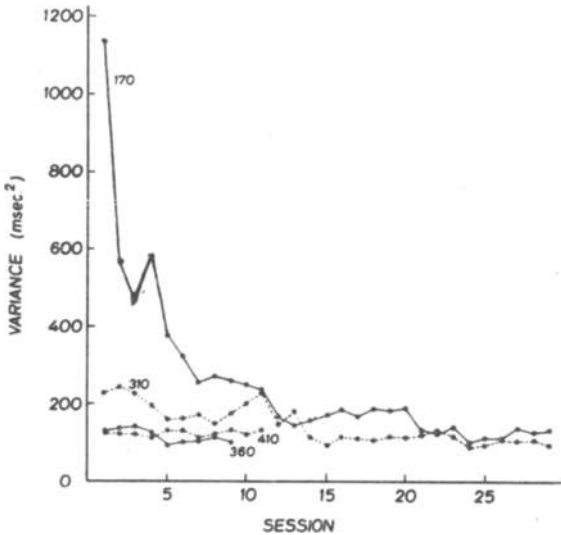


Figure 10. Changes in variances with practice at each of four P₁P₂ intervals run in the order 170, 310, 360, 410. Subject G.H. Session 1 at 170 followed immediately after the RT series of Figure 9. N = 300 per point.

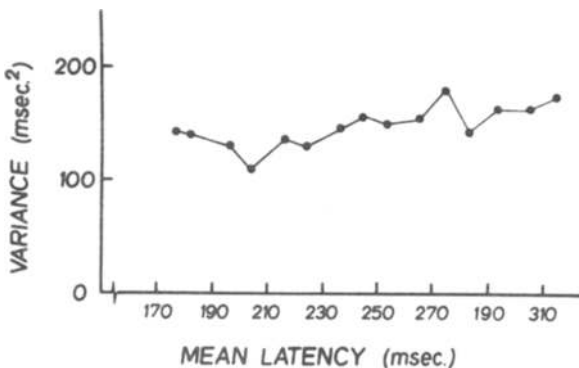


Figure 11. Variance vs. mean latency over Region 1 for subject G.H. The region was scanned beginning and ending at 170, and the data are combined ascending and descending series.

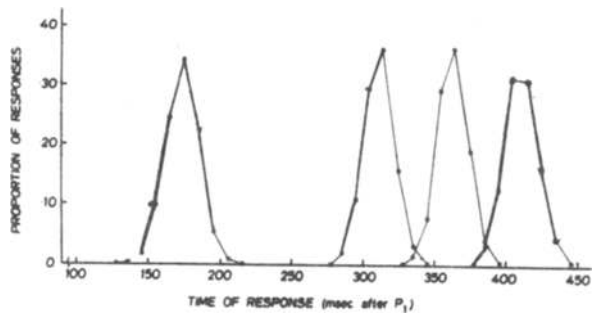


Figure 12. Latency distributions for G.H. after specific training at each of four P₁P₂ intervals. The two distributions at the left are at the ends of Region I. The two on the right are within Region II. SDs are 11, 10, 10, and 11 msec.

EXPERIMENT IV

The conclusion that the SD of time estimates is a constant proportion of the mean, discussed above in the introduction, does not conform to the results obtained under the conditions of these experiments. Since the SD is unchanging as the mean increases from 165 to about 550 msec, the ratio of the SD to the mean has been seen above to change from 0.067 to 0.020 over that range. Increasing the mean above 550 does produce an increase in the SD, and the purpose of this experiment is to study the function over a range wider than that of Experiment II.

After finishing Experiment I, subject A.K. continued the ascending series. Four sessions were done at each step between 990 and 2,190 msec. The step size was 50 msec.

Means and SDs for the entire ascending series are given in Table 2. The ratio of the SD to the mean is plotted as a function of the mean in Figure 13. While the SD does increase above 550, the initial increase is so slow that the ratio continues to decrease slightly to a low of 0.018. The ratio then gradually increases to about 0.027 at 2,190 msec.

The most adequate simple description of these data yields the theoretical curve drawn in Figure 13. It states that for mean latencies less than some value, L, the variance is constant (here the variance is 142 msec² and L is 550 msec). When the mean exceeds L, the variance is increased by an increment which is equal to K²(M - L)². That is to say, the ratio of the SD to the mean is a constant K for increments that are added in excess of the limit L. It also assumes additive variances. The value of K for Figure 13 is 0.034.

The model is a reasonable general description of the data. The points at the left end are in Region I and they do not fall on the function as would be expected from the earlier discussion of Region I. But there may be a systematic failure of the model in the fairly wide region between 700 and 1,200 msec. The obtained ratio is virtually constant there. A more complex model may be necessary, but additional experiments are necessary to establish whether

Table 1
Number of Responses in Each 5-msec Bin for the Latency Distributions of Figure 12

X	Lower Bound of Modal Bin			405
	175	310	360	
-50	3			
-45	5			
-40	7			
-35	16	1	2	
-30	32	6	5	1
-25	86	24	17	13
-20	176	51	31	34
-15	288	114	87	65
-10	373	191	186	121
- 5	445	259	258	213
0	473	292	291	262
5	371	259	263	240
10	228	150	183	229
15	109	90	106	154
20	40	34	42	89
25	18	13	21	46
30	8	1	5	22
35	1	1	3	5
40	1			2

Note—X is the distance of the lower bound of a bin from the lower bound of the modal bin.

that is the case. Also, conclusions from this experiment must be tentative because only the ascending series was done.

SEQUENTIAL DEPENDENCE

As already noted, when performance has been stabilized, differences between sessions and between blocks are negligible. But it remains possible that there are short-term changes within blocks which contribute to the total variance. If so, the true variance may be even less than the very small values that are obtained.

To find out, lag 1 sequential effects were analyzed for several groups of consecutive sessions of stable performance for the two subjects. Figures 14 and 15 summarize the results for A.K.

In both cases, there is evidence of temporal autocorrelation, the mean latency on trial (n + 1) being a positive function of the latency on trial n. If this autocorrelation were due only to a wandering of the mean of an otherwise fixed true distribution, then the SD on trial (n + 1) should be an inverted-U function of the latency on trial n. The value of the SD at the extremes would then be a better estimate of the true SD. The results are equivocal in this regard, as the figures show, perhaps because even with the samples of 1,500 each, the numbers of responses at the extremes are quite small.

There is some degree of autocorrelation for A.K., and it is probable that his obtained minimum SD of 12 msec in Region II is somewhat inflated.

The smaller SD of 10-11 msec obtained for G.H. within Regions I and II is not subject to the same qualification. Figure 16 shows that there is no lag 1

autocorrelation for G.H. and no evidence of instability from trial to trial.

ADDITIONAL SUBJECTS

Four adult female subjects have also been studied but in no case was it possible to continue long enough to obtain convincingly stable data. Three of these subjects were trained with a P₁P₂ interval of 460 msec, G.R. and M.K. for 20 sessions each and B.B. for 50 sessions. G.R. and M.K. appeared to level off after 10 sessions at SDs of 21 and 19 msec. B.B. showed a very slowly diminishing variance, her SD reaching 24 msec after 45 sessions. M.K. contributed additional sessions with an interval of 310 msec, and her SD reached 14 msec.

The fourth additional subject, C.H., began with 31 sessions at a P₁P₂ interval of 310. After 25 sessions, her SD averaged 12.5 msec. She was then used to scan Region I as G.H. had done to produce the data of Figure 11 except that she began and ended at the upper rather than the lower edge of the region. Her variance increased slightly as she approached the RT limit but not as markedly as did A.K.'s in Experiment II. Her SD was 15 msec when her mean was 170 msec at the lower boundary..

These subjects all demonstrate that time estimation latencies can have as little variability as simple RTs. Two of the four subjects gave lowest SDs close to those reported in the main experiments, and one of

Table 2
Means and SDs for Subject A. K., Ascending Series of Experiments II and IV

P ₁ P ₂ Interval	Mean	SD	P ₁ P ₂ Interval	Mean	SD
155	157	19.9	1090	1081	19.5
165	164	16.0	1140	1132	20.5
170	177	15.0	1190	1180	21.7
180	188	14.1	1240	1237	24.2
190	195	15.1	1290	1287	27.5
200	203	14.7	1340	1336	28.8
210	212	14.3	1390	1389	29.8
250	238	15.0	1440	1430	31.7
300	294	11.7	1490	1482	33.5
310	307	11.5	1540	1536	36.2
340	339	11.9	1590	1578	41.1
440	435	11.6	1640	1623	40.8
450	445	12.2	1690	1673	44.0
460	457	12.6	1740	1721	41.2
500	498	11.8	1790	1761	46.1
550	547	11.7	1840	1820	42.8
600	599	13.3	1890	1869	48.8
720	717	13.4	1940	1929	45.5
800	794	14.8	1990	1971	49.8
840	837	15.8	2040	2013	55.5
890	885	16.1	2090	2059	57.3
940	936	17.2	2140	2112	58.6
990	981	17.6	2190	2165	59.1
1040	1034	18.7			

Note—X is the distance of the lower bound of a bin from the lower bound of the modal bin.

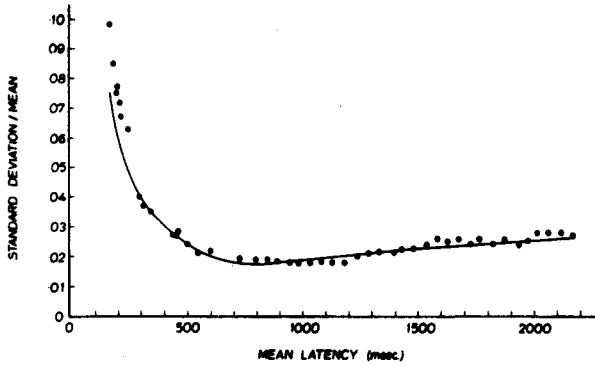


Figure 13. Ratio of SD to the mean for the ascending series by subject A.K. from Experiments II and IV. $N = 900$ or more per point. The model that generates the theoretical curve is explained in the text.

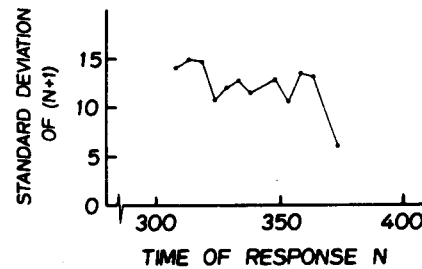
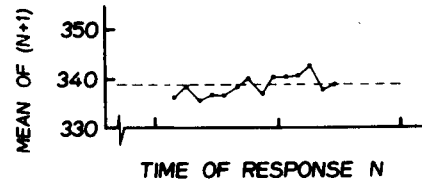
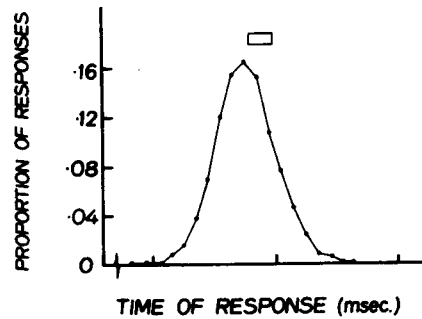


Figure 15. Same autocorrelation analysis as Figure 14. $N = 1,500$ with P_1P_2 interval at 340. $M = 338$. Total variance = 150. Mean within block variance = 142.

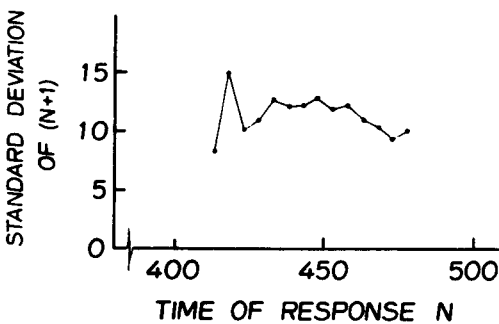
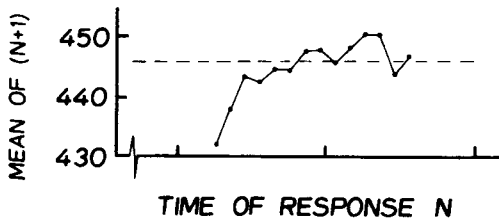
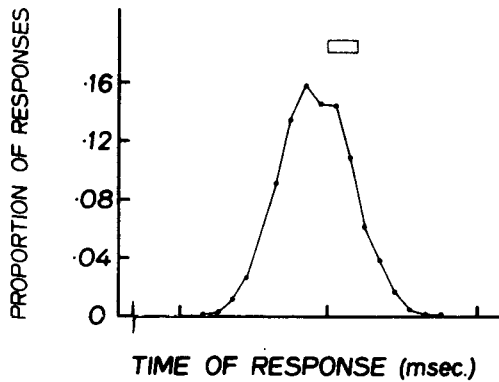


Figure 14. Lag 1 autocorrelation analysis of 1,500 latencies with P_1P_2 interval of 450 for subject A.K. Top figure shows the total distribution ($M = 446$, total variance = 150, mean within block variance = 145.) Middle and bottom figures show the mean and SD on trial $(n + 1)$, given that the response on trial n was of the latency on the abscissa.

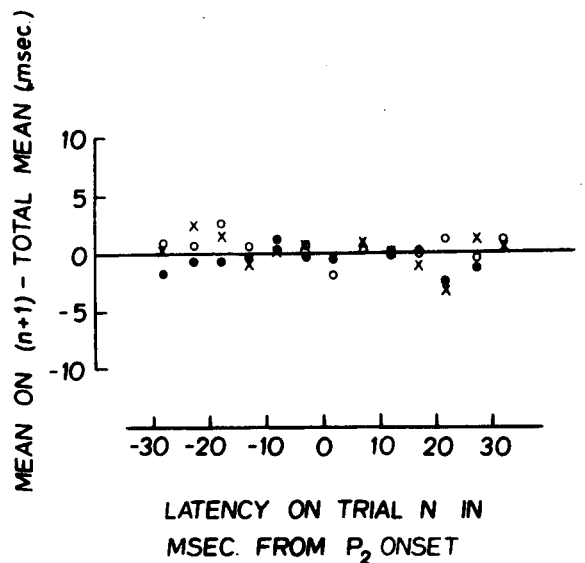


Figure 16. Three autocorrelation analyses for subject G.H. each with 1,500 latencies. The data are the three rightmost distributions of Figure 12 for P_1P_2 intervals of 310, 350, and 410. The mean latency on a trial does not depend upon the latency of the preceding trial.

them, C.H., confirmed the finding that variability does not increase as the mean is increased within Region I. They also demonstrate that there are large individual differences in the amount of training required to produce low variability performance.

DISCUSSION

These experiments find that, in the limit, stimulus-response latencies have several significant properties which have not previously been revealed. Under some conditions, responses are almost certain to be elicited by a stimulus and the temporal certainty associated with the response is also high. The latter is reflected by the inverse of the latency variability. Latency standard deviations as small as 10 msec are reported above. This reduces the upper bound on the theoretical minimum latency variance by a factor of about 4 over that provided by previous research. To state the point differently, the results give maximum justification for the simple statement that a stimulus will (certainly) elicit a response and the response will occur at a particular time plus or minus an amount which is less than 50 msec.

Such a statement declares that the stimulus and the response are both instantaneous, time-point events. The stimulus is the rapid onset of the 10 msec auditory pulse. The response is electrical contact of the response switch, and it is, obviously, a consequence of a movement and not the movement itself. The response occurs after the initiation of the movement. Since the movement must be somewhat variable, it is difficult to escape the conclusion that the temporal certainty of the initiation of the movement is even greater than that indexed by the minimum latency variance of 100 msec².

The experiments also demonstrate that the degree of temporal certainty associated with the response does not depend upon the mean latency until the mean latency is quite long. The observable minimum latency variance is the same when the response is a time estimation response as when it is a simple RT. The only estimate of the upper bound on this principle is the 550-msec upper bound on Region II found in Experiment II. The validity of the principle, however, is shown also in Experiment III and elsewhere.

The most immediate inference is one that goes against the grain of current thinking about timing mechanisms. It is that a time delay can be inserted into an S-R chain without increasing the variability of the chain. The added delay is a time interval which can be set at any value and, once set, is deterministic. Such an added delay can have any value between zero (simple RT) and about 400 msec (the upper bound of Region II minus simple RT).

The third major result has to do with the form of the latency distribution. When the variance is

minimized, the distribution has a form which is not typical of reaction time distributions. Instead as described earlier in more detail, it is symmetrical and sharply peaked. Further, its form is everywhere the same within Regions I and II and the deterministic added delays appear to affect only its mean.

Therefore, the minimum variance of an S-R chain is 100 msec² or less. The sharp peaks on the distributions and their simple form suggest that this variance is not divided among a large number of different sources. The work by Wing and Kristofferson (1973a) on interresponse timing also isolated a component of variance which they identified as efferent delay variance in addition to the central component shown in Figure 1. The efferent delay variance depends upon the form of movement, of course, and their responses were repeated at regular intervals rather than made only once. For a response similar to the one used here, they obtained estimates of efferent latency variance which ranged from 10 to 50 msec². Such amounts are small relative to usual RT variance but they loom larger in the present context.

If the usual tripartite parsing of the S-R chain into afferent, central, and efferent components is considered and the efferent component is subtracted from the minimum total variance, the remainder would be approximately that of a triangular distribution with a base of about 50 msec. That such a triangle represents a unit of central temporal variability is a highly speculative idea at the present time, but it is implied by Kristofferson's time quantum theory (Kristofferson, 1967). Within this speculative context, one is here left with the conclusion that the afferent latency component is deterministic. This latter conclusion is consistent with the assumptions of the attention-switching model of successiveness discrimination (Allan & Kristofferson, 1974; Kristofferson, 1967), with the recent work on temporal order discrimination by Allan (1975), and with certain conclusions by Ollman and Billington (1972).

The minimum obtained variance is constant over Regions I and II, but the conclusion that this is due to an adjustable, deterministic internal delay is not inescapable. One could proffer a multiple factor theory to account for it, for example, and argue that the constancy is merely an almost exact balance between two, or more, opposing factors. An unknown factor, X, might operate increasingly to inflate the variance as the RT limit is approached, or it might operate increasingly to suppress the variance as the interval increases. A possible X might be found in the perceptual feedback. As the P₁P₂ interval becomes small, P₁ may interfere with the extraction of information from the relation between the response sound and P₂. Thus, as the interval increases, the feedback becomes more effective and the variance diminishes. This could be the reason

for the increase in variance actually seen as the RT limit was approached in Experiment II.

But the point is that the SD is only 10 msec when the mean is 550 msec. However X is thought to operate, postulating X implies that the true SD near the RT limit is less than 10 msec and the conclusion that the entire S-R chain is deterministic is rapidly approached. There certainly are highly stable, long-duration elements somewhere in the system.

If there are deterministic delays, the important next questions ask after their nature and their locus. This author knows of no plausible physiological explanation. Meijers and Eijkman (1974) approach the problem in their discussion of the motor system, but they do not provide all of the answer that is needed here.

As for locus, only a little can be said. From the form of the response movement, the interval between the initiation of the movement and the time point of the response can be excluded as a sufficient locus although fine adjustments in mean latency might be made there. Afferent latency is a possible locus, especially because of the evidence that it can be almost free of variability. If so, it would be significant, because there would then be many opportunities for the delays to participate in single acts of information processing. If, on the other hand, the delays are located within autonomous motor programs, one would be unlikely to find them contributing to, say, duration discrimination.

In time quantum theory, the quanta have been assumed to have a fixed duration and that duration has been shown to be related to the frequency of the alpha rhythm (Kristofferson, 1967). Deterministic delays could be due to accurate counting of quanta with an upper limit on perfectly accurate counting and much practice to achieve error-free counting.

For mean latencies longer than the upper bound of Region II, variance increases with the mean. The theoretical curve in Figure 13 assumes that the maximum deterministic delay remains in the S-R chain and that further delay is achieved through the insertion of a timer like the central timekeeper of Wing and Kristofferson (1973a). The ratio of SD to mean obtained in Experiment IV agrees well with that obtained by them and shown here in Figure 1. This implies that their central timekeeper is not merely a response timer; it can be triggered by inputs arriving over the S-R chain. These conclusions are subject to the qualifications stated above concerning Experiment IV.

Wing and Kristofferson (1973b) rejected the hypothesis that a deterministic timekeeper controls interresponse intervals in their experiments. This conclusion also supports the notion that the deterministic delays are located in the afferent end of the S-R chain. It may be that deterministic delays cannot control interresponse intervals because the deterministic delays must be triggered by a stimulus, or it

may be that they do not control interresponse intervals because their output must traverse the more variable central part of the chain before triggering a response.

There is a subjective report which should be recorded here. It is a pronounced subjective effect which occurs in these experiments and also in duration discrimination experiments which employ the same P_1P_2 stimulus patterns. When the interval between P_1 and P_2 is long, the improvement in performance with practice is paralleled by a decrease in the perceived duration of the stimulus interval. When the interval is long, say 1,000 msec, the effect is very large and unmistakable, even though a direct comparison cannot be made. Part way through practice, a more direct comparison can be made and the subject can experience the perceptual shortening occurring within a block, often over just a few trials. After performance has stabilized, the shortened perceived duration is experienced on the first trial of a session and throughout the session. This observation bears on the hypothesis that the deterministic delay is in the afferent latency. It supports that idea, but not in a necessary way because it is unlikely that the afferent latency of the percept is the same as the afferent latency of the S-R chain. However, the two must share a common path at least in part.

Why do these experiments provide answers to the questions raised in the introduction which are so different from previous answers? In the main, they are different answers, although Saslow (1974) approached minimum variance and also found symmetrical distributions within the RT range. The differences are due to several matters of experimental control which are by no means isolated and clearly identified by the experiments. The perceptual feedback feature is one of these, and it is that, in combination with certain training procedures, which is important. Exhaustive training is required, and the training must be with conditions fixed over many sessions. Changing even the P_1P_2 interval from session to session by large amounts would produce very different results. Also, it is sometimes the case that performance with a particular interval improves *more* as a result of practice with other intervals than it does as a result of direct practice with it, although the reverse also happens and a particular practice gain may be lost as a result of experiencing other intervals. Extensive experience is necessary, both particular and general, followed by evidence of stable performance.

REFERENCES

- ALLAN, L. G. The relationship between judgements of successiveness and judgements of order. *Perception & Psychophysics*, 1975, 18, 29-36.
- ALLAN, L. G., & KRISTOFFERSON, A. B. Successiveness discrimination: Two models. *Perception & Psychophysics*, 1974, 15, 37-46.

- BARTLETT, N. R., & BARTLETT, S. C. Synchronization of a motor response with an anticipated sensory event. *The Psychological Review*, 1959, **66**, 203-218.
- GETTY, D. J. Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics*, 1975, **18**, 1-8.
- GREEN, D. M., & LUCE, R. D. Detection of auditory signals presented at random times: III. *Perception & Psychophysics*, 1971, **9**, 257-268.
- KORNBLUM, S. Simple reaction time as a race between signal detection and time estimation: A paradigm and model. *Perception & Psychophysics*, 1973, **13**, 108-112.
- KRISTOFFERSON, A. B. Attention and psychophysical time. *Acta Psychologica*, 1967, **27**, 93-100.
- KRISTOFFERSON, A. B., & ALLAN, L. G. Successiveness and duration discrimination. In S. Kornblum (Ed.), *Attention and performance IV*. New York and London: Academic Press, 1973.
- MEIJERS, T. M. M., & EIJKMAN, E. G. J. The motor system in simple reaction time experiments. *Acta Psychologica*, 1974, **38**, 367-377.
- NAATANEN, R., MURANEN, V., & MERISALO, A. Timing of the expectancy peak in simple reaction time situation. *Acta Psychologica*, 1974, **38**, 461-470.
- OLLMAN, R. T., & BILLINGTON, M. J. The deadline model for simple reaction times. *Cognitive Psychology*, 1972, **3**, 311-336.
- SASLOW, C. A. Dual payoff band control of reaction time. *Perception & Psychophysics*, 1974, **15**, 108-114.
- SNODGRASS, J. G., LUCE, R. D., & GALANTER, E. Some experiments on simple and choice reaction time. *Journal of Experimental Psychology*, 1967, **75**, 1-17.
- TREISMAN, M. Temporal discrimination and the indifference interval: Implication for a model of the "internal clock". *Psychological Monographs*, 1963, **77**, No. 13.
- WING, A. M., & KRISTOFFERSON, A. B. Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, 1973, **14**, 5-12. (a)
- WING, A. M., & KRISTOFFERSON, A. B. The timing of inter-response intervals. *Perception & Psychophysics*, 1973, **13**, 455-460. (b)
- WOODROW, H. The reproduction of temporal intervals. *Journal of Experimental Psychology*, 1930, **13**, 473-499.
- WOODROW, H. Individual differences in the reproduction of temporal intervals. *American Journal of Psychology*, 1933, **45**, 271-281.

(Received for publication April 2, 1976;
revision received June 4, 1976.)