PSYCHOLOGICAL SCIENCE

Research Article

MENTAL CHRONOMETRY: Beyond Reaction Time

Richard A. Abrams and David A. Balota

Department of Psychology, Washington University

Abstract—*Details of response execution were examined in two classic human information processing paradigms: lexical decision and memory scanning. In the lexical decision experiment, word frequency influenced both the time of response onset and kinematic properties of the response. In the memory scanning experiment, when the probe was present in the memory set, the responses were both initiated sooner and were more forceful when the memory set consisted of two items than when it consisted of six items. When the probe was absent from the memory set, responses were initiated sooner but were less forceful when the memory set consisted of two items than when it consisted of six items. The results suggest that the amount of activation in support of a given response can modulate both the time taken to initiate the response and the force with which the response is executed. These findings bear on all models of human cognitive performance that have been developed within the mental chronometry tradition.*

Over 100 years ago, Donders (1868–1869/1969), a Dutch physiologist, developed a method to measure what he called "The Speed of Mental Processes." His procedure was quite elegant. Donders recorded the time people required to complete two simple tasks that were identical with the exception of a single additional mental operation inserted into one of the tasks. He argued that the difference in completion times for the two tasks could be used as an estimate of the duration of the target mental operation. In this classic work, Donders derived estimates for the durations of the mental processes involved in stimulus identification and response selection.

Since Donders' original work, there has been a considerable increase in sophistication regarding the inferences that are and are not possible from response latency data (see Meyer, Osman, Irwin, & Yantis, 1988, for a review). Despite the diversity of the mechanisms proposed to account for response latency data, all information processing models make an important assumption: Functionally early processes such as stimulus identification and response selection are assumed to be completed prior to the onset of an overt response. The response is then believed to be ballistically "triggered" when such processing is completed. Thus, the primary interest in studies of mental chronometry is the influence of factors on *when* a response is initiated, not *how* the response is executed.

Recently, however, there has been evidence accumulating that characteristics of the response itself can be influenced by the same factors that affect the earlier stimulus identification and response selection processes. For example, Balota, Boland, and Shields (1989) demonstrated that the *duration* of a pronunciation can be influenced by some of the same manipulations that affect the time to initiate the pronunciation. Also, Osman, Kornblum, and Meyer (1986) have shown that response latency can be affected by factors that have their impact after the point in processing when response selection has already occurred (i.e., after the "point of no return"). Finally, Coles et al. (1985) have shown that activation may grow continuously prior to the production of an overt response in a two-choice reaction time paradigm, sometimes resulting in covert muscular activity that favors the incorrect response. These results suggest that the parameters involved in the execution of a response may be affected by factors that previously were believed to influence only the earlier stimulus identification and response selection stages.

The present experiments extend these recent demonstrations in two ways. First, we investigated the extent to which kinematic features of a response may be related to response latency by using an *arbitrary* response (a limb movement) that is not inherently tied to the factors being manipulated. This approach eliminates some limitations of previous work. For example, in the Balota et al. (1989) study described above, a linguistic variable (associative context) was shown to influence the duration of a linguistic response (naming duration). But stress, a high correlate of duration, can provide useful cues to a listener in speech. Hence, the response used in the Balota et al. study was not arbitrary with respect to the manipulation, and the findings may not generalize beyond the domain of speech production.

Second, we studied a single graded response under stimulus conditions that unequivocally defined the response. In the Coles et al. (1985) study, described above, subjects responded to centrally presented target letters (*H* or *S*) by squeezing handles with either their right or left hands. Targets could be flanked by either compatible stimuli (*HHHHH*) or incompatible stimuli (*SSHSS*). They found that squeeze responses were stronger when the flanking stimuli were compatible with the target than when they were incompatible. However, as noted above, they also found that incompatible stimuli produced partial squeeze responses in the incompatible response channel (i.e., the opposing hand). Thus, as Coles et al. claimed, their results reflect *response competition* processes. In the present experiments, there were no characteristics in the stimuli (e.g., flanking incompatible letters) that simultaneously led to incompatible responses.

Address correspondence to either author at Department of Psychology, Washington University, St. Louis, MO 63130. e-mail: C39823RA@WUVMD.bitnet.

EXPERIMENT 1

In the first experiment, we examined the influence of word frequency on characteristics of the response in a lexical decision task, the premier reaction time task used to examine variables that influence word identification processes. On each trial, subjects judged the lexical status of a letter string. To indicate their decision, subjects moved a handle rapidly in one direction if the string was a word, and in the other direction if the string was a nonword.[1] When the handle passed a target location 5.3 cm to the left or right of the starting position, a tone sounded indicating the completion of the response.

The predictions in this paradigm are straightforward. If word frequency influences only those processes involved in word identification, as most models assume, then word frequency should affect only the time taken to initiate the "word" response. However, if there is also an impact of word frequency on the execution of the response, then the time taken to reach the target location after the response has been initiated should also be affected.[2]

Method

Subjects

Twenty naive right-handed undergraduate volunteers from Washington University participated in this experiment.

Materials

Thirty-six high- and 36 low-frequency target words were selected from the Kučera and Francis (1967) norms. The low-frequency words had counts less than seven per million while the high-frequency words had counts greater than 36 per million. The high- and low-frequency words along with 72 pronounceable nonwords were matched on number of syllables, phonemes, and letters. There were also 23 words and 23 pronounceable nonwords that served as practice and buffer items.

Equipment and procedure

Stimuli were presented on a CRT display, and responses were obtained from a handle that was mounted on a track directly in front of the CRT. The subject grasped the handle, which moved freely from side to side, in his or her right hand. The handle was connected to a transducer that produced a position-dependent voltage that was digitized and sampled at a rate of 1 kHz with a resolution of 0.01 cm. Movement onset was defined to be the first moment in time when the velocity of the handle exceeded 3.3 cm/s and remained above that value for 20 ms. The end of movement was defined to be the first moment at which the handle passed a point 5.3 cm from the starting location. Subjects were instructed to pass through that point as soon as possible to indicate their response. (Movement analysis was performed using the procedure described in Meyer et al., 1988.)

Results

The mean response latencies and kinematic features of the ensuing responses are displayed in Table 1. First, as has been shown previously (e.g., Balota & Chumbley, 1984) response latency was slowest for nonwords, faster for low-frequency words, and fastest for high-frequency words. These differences presumably reflect the time needed to make the central decision

Table 1. *Results from lexical decision experiment (Experiment 1)*

Dependent measure	Nonword	Word Low-Frequency	 High-Frequency
Reaction time (ms)***	626.6	613.1	499.2
Percent correct***	93.0	86.0	98.0
Movement duration (ms)**	136.3	131.8	128.0
Final velocity (cm/s)**	79.5	78.0	80.6
Peak acceleration (cm/s/s)**	151.6	152.9	159.8
20 ms after movement onset			
Position (cm)	.21	.20	.22
Velocity (cm/s)	15.6	15.7	16.3
Acceleration (cm/s/s)**	96.9	96.8	101.6
50 ms after movement onset			
Position (cm)*	1.17	1.17	1.23
Velocity (cm/s)**	47.4	48.4	50.3
Acceleration (cm/s/s)	106.0	108.5	111.0

*$p < .05$; **$p < .005$; ***$p < .001$. All comparisons are between high- and low-frequency words using a t test with 19 degrees of freedom and are in the predicted direction. Movement duration is the time interval from movement onset until the handle passed a target location 5.3 cm from the starting location; final velocity is the velocity of the handle at the end of that interval; acceleration is proportional to force.

---

[1] Half of the subjects had rightward responses assigned to nonwords, and half had rightward responses assigned to words. The counterbalancing of direction and response was also used in Experiment 2.

[2] Of course, word frequency may affect characteristics of the response other than the time to reach the target location. In theory, an infinite number of force–time functions could move the limb from the starting position to the target in a given time period. Nevertheless, any overall change in the force of the response would also affect the time taken to reach the target.

regarding the lexicality of the stimulus. More importantly, the effect of word frequency did not end at the initiation of the response. Kinematic features of the responses to the low- and high-frequency words differed. For example, among other differences, high-frequency word responses were accelerating more rapidly 20 ms after movement onset, they were moving faster and had already travelled farther 50 ms after movement onset. The total duration of the movement to the target location was shorter for high-frequency words as compared to low-frequency words. Thus, word frequency not only influenced the latency with which the lexical decision response was initiated, it also influenced the force with which the response was produced.[3]

EXPERIMENT 2

In order to insure that the results of Experiment 1 were not due to idiosyncratic characteristics of the lexical decision task, an attempt was made to extend the results to a second widely used reaction time task. Experiment 2 employed a memory scanning task, the premier task used to isolate variables that influence retrieval processes in short-term memory (e.g., Sternberg, 1966). On each trial, subjects studied a memory set consisting of either two or six digits. They were then shown a probe digit that either was or was not a member of the previously presented memory set. Subjects responded by moving the handle in one direction if the probe was present in the memory set, and in the other direction if the probe was absent from the set.

The predictions in this paradigm are also straightforward. If the size of the memory set influences only the time taken to retrieve the probe from short-term memory, as most models assume, then set size should affect only the time taken to initiate the response. However, if memory set size also influences processes involved in the execution of the response, then set size should also influence the time taken to reach the target location after the response has been initiated. In addition, because the set size manipulation is relevant for both trials in which the target is present in the memory set, and for trials in which the target is absent from the memory set, the memory scanning paradigm permits an investigation of the relation between response latency and response kinematics for both positive and negative responses. Such a comparison was not possible for the negative (nonword) responses in Experiment 1.

Method

Subjects

Twenty right-handed undergraduate volunteers at Washington University participated in this experiment. None of the subjects had participated in Experiment 1.

Materials

The digits 1 through 9 served as the stimuli for the memory sets and probes. Each subject received 45 target trials in each of the four conditions that were produced by crossing memory set (2 or 6) and presence in the memory set (present or absent). In addition, subjects received 36 practice trials.

Equipment and procedure

The equipment and procedure were the same as that used in Experiment 1, with the following exceptions: On each trial, subjects studied the memory set until they were ready for the test probe. They then pressed a button on the response handle that cleared the screen. The probe was presented 500 ms later. Subjects were required to rapidly move the handle in one direction if the probe was present in the memory set and in the other direction if the probe was absent from the set (see footnote 1).

Results

Mean response latencies and kinematic features of the ensuing responses are displayed in Table 2. As expected, responses on both "present" and "absent" trials were initiated sooner and were more accurate when the memory set consisted of two digits as compared to six digits. These differences presumably reflect the impact of set size on the processes involved in retrieval from short-term memory (Sternberg, 1966). More importantly, kinematic aspects of the responses again differed as a function of set size, and interestingly, the direction of the difference depended on whether the response was positive or negative. For the positive (i.e., "present") responses (as for the "word" responses in the lexical decision task), set-size-two responses were accelerating more rapidly, moving faster, and had already travelled 50 ms farther after movement onset relative to set-size-six responses. The overall movement duration was shorter for the set-size-two responses compared to the set-size-six responses. However, for the negative (i.e., "absent") responses, set-size-six responses were moving more quickly, accelerating at a greater rate, and had a greater final velocity than set-size-two responses. Thus, memory set size influenced the time needed to search the set in memory and initiate the appropriate response, as well as the force with which the response was executed. Moreover, the relation between onset latency and movement kinematics differed between positive and negative trials.[4]

---

[3] There are two further points to note here. First, consider the details of the distributions of the dependent measures in the various conditions. It is possible, in theory, that the observed differences in responses to high- and low-frequency words arose from different probability mixtures of relatively forceful and relatively less forceful responses in the two conditions. There were no reliable differences in the standard deviations of any of the force measures between the two conditions ($ts(19) < 1.4$, $ps > .15$). Second, it is important to note that the present data do not necessarily indicate that response production begins before "earlier" processes such as stimulus identification are completed. The primary finding here is that word frequency can influence some characteristics of the response other than its latency. Whether this effect occurs early or late in the information processing stream is an issue that cannot be fully addressed within the present experiments.

[4] As in Experiment 1, we examined the standard deviations of the various dependent measures for evidence that the differences in means might have arisen from different probability mixtures of more and less forceful responses in the different conditions. As in Experiment 1, there was little evidence of such a mixture: Of the 18 comparisons performed

Table 2. *Results from memory scanning experiment (Experiment 2)*

Dependent measure	Absent Set size 6	Absent Set size 2	Present Set size 6	Present Set size 2
Reaction time (ms)	572.8***	465.4	555.1***	415.1
Percent correct	89**	94	90**	95
Movement duration (ms)	114.2	117.1	114.8*	111.1
Final velocity (cm/s)	87.2*	84.0	80.3	80.5
Peak acceleration (cm/s/s)	171.6*	161.7	161.9	167.9
20 ms after movement onset				
Position (cm)	.24	.25	.18	.20
Velocity (cm/s)	17.5*	16.5	17.1	17.8
Acceleration (cm/s/s)	103.9*	96.9	104.6*	110.6
50 ms after movement onset				
Position (cm)	1.28**	1.23	1.21**	1.29
Velocity (cm/s)	50.7*	48.4	50.3*	52.7
Acceleration (cm/s/s)	122.1*	118.5	116.0*	122.4

*p < .05; **p < .01; ***p < .001. All comparisons are between set-size two and set-size six responses using a t test with 19 degrees of freedom.

CONCLUSIONS

Our results suggest that it is necessary to extend current chronometric models beyond stimulus recognition and response selection to processes involved in response execution.[5] Experiment 1 indicated that high-frequency words led to shorter response latencies and more forceful responses than low-frequency words in a lexical decision task. Experiment 2 replicated the relation between onset latency and force for "present" responses in a memory scanning task, but revealed the opposite relation for "absent" responses. Specifically, for absent responses the condition that yielded the fastest response latencies (set-size-two) actually led to the least forceful responses.

First, consider the results of the lexical decision experiment. As in most response latency paradigms, this task requires the subject to choose one of two alternative responses ("word" or "nonword"). One dimension considered as useful in discriminating words from nonwords is familiarity of the stimulus: Words are more familiar stimuli then nonwords (Balota & Chumbley, 1984; Besner & Swan, 1982; Seidenberg & McClelland, 1989). As a result, the degree of familiarity of the stimulus in a lexical decision task is directly related to the evidence in favor of the word response or the nonword response. For high-frequency words, there is considerable familiarity and strong support for the word response. However, for low-frequency words there is considerably less familiarity, and therefore evidence may exist in favor of both "word" and "nonword" responses (Balota & Chumbley, 1984). Because it takes time to resolve the conflict between word and nonword responses, onset latencies should be slower for low-frequency than for high-frequency words, which is what we observed.

There are at least two possible loci for the effect that word frequency had on the force of the responses in the lexical decision task. First, it is possible that subjects adjust the force of their response based on the strength of the evidence in favor of the response: Stimuli that have higher levels of familiarity would yield more forceful "word" responses because more evidence exists in favor of the "word" response. Thus, responses to high-frequency words would be more forceful than those to low-frequency words, as we observed. Second, it is possible that word-frequency also has an effect at a more peripheral, response–conflict locus. As discussed above, low-familiarity (low-frequency) words may produce evidence favoring both the "word" and "nonword" responses simultaneously. Because these two responses are incompatible, any such conflict would decrease the force of the "word" response for low-frequency words when the response is eventually emitted. This second possible locus of the force effect suggests that characteristics of the stimulus itself may simultaneously lead to conflicting responses, thereby influencing the force of the correct response. This is similar to what may have happened in the Coles et al. (1985) study discussed earlier, in which central and flanking stimuli led to incompatible responses.

Although it is not possible to identify the precise locus of the impact of word-frequency on the force of the responses in the lexical decision task, the memory scanning data are relatively

---

for the present and absent responses, only one attained significance (final velocities were more variable for present responses with a set-size of two than for present responses with a set-size of six, $t(19) = 2.34$, $p < .05$).

[5] Of course, one might argue that the effects observed on response kinematics are relatively small compared to the large effects on onset latency. Although this is true, we do not feel that this weakens our conclusions. The major thrust of the present work is to demonstrate that different factor levels in two widely used reaction time paradigms do not simply yield identical responses that are ballistically triggered with different latencies. Hence, the appropriate question is whether kinematic features differ at all, not whether the effects of factors on response kinematics are "large" compared to the effects of the same factors on response latency. The present results clearly indicate that there are reliable and systematic differences in the responses that are produced in the various conditions of two widely studied reaction time paradigms.

more informative regarding the effects of set size on response force. In the memory scanning paradigm, there is nothing inherent in the stimulus (a probe digit) that would lead to opposing responses for either the set-size-two or set-size-six memory sets. Instead, the results from the memory scanning paradigm reveal the effects of more central mechanisms involved in short-term memory retrieval processes.

Results from the memory scanning paradigm can be accounted for by a simple extension of Sternberg's (1966) serial exhaustive search model. The extension involves an additional assumption, proposed above for the lexical-decision data: Specifically, we assume that the force of a response is adjusted on the basis of the relative amount of evidence available in favor of the selected response as compared to that available supporting the alternative response. Consider the application of the model to the present trials. Following Sternberg, one would argue that subjects search through the complete memory set on each trial before making the central present/absent decision. Because fewer memory items need to be compared to the memory probe for smaller memory sets, set-size-two trials would be expected to have shorter onset latencies than set-size-six trials, as we observed. The difference in the force of the "present" responses for different memory set-sizes arise from the proposed modulation of response force on the basis of the amount of evidence in favor of the "present" response relative to the amount of evidence in favor of the "absent" response. Because there are more non-matches to the memory probe in a memory set containing six items as compared to two, more evidence would exist favoring the "absent" response from a set-size-six trial. If the subject uses the relative evidence in favor of each response to adjust the response force, then one would predict more forceful "present" responses for set-size-two trials compared to set-size-six trials, which is what we observed.

Now, consider the application of the same model to the "absent" responses. Because the probe needs to be compared to fewer items on trials with a memory set of two, onset latencies would be expected to be faster for set-size-two trials (compared to set-size-six trials), as was true for the "present" responses. This is the result we observed. However, the model predicts that the force of the "absent" responses should actually exhibit a pattern opposite to that for the "present" responses. Specifically, although the memory search through a set of six items takes longer to complete than a search through two items, more evidence for an "absent" response would be available after searching through six non-matching items as compared to two. Thus, one would expect that the "absent" responses would be more forceful on set-size-six trials than on set-size-two trials, precisely as the data indicated. Thus, our results are easily explained by a simple serial exhaustive search model in which response force is modulated by the amount of evidence in favor of the selected response relative to the amount of evidence in favor of the alternative response. The model predicts the intriguing dissociation between effects of memory set-size on response latency and response force for the "absent" responses, as well as the correspondence between measures of latency and force for the "present" responses.

These conclusions complement and extend current models of cognition by showing that factors that have previously been viewed as modulating only early processes such as stimulus recognition and response selection influence not only *when* a response is initiated, but also *how* the response is executed.

Acknowledgements—This work was supported by NIH grants RO1-A607406 to D. Balota and R29-MH45145 to R. Abrams. The present work represents a totally collaborative effort.

REFERENCES

Balota, D.A., Boland, J.E., & Shields, L.W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language, 28*, 14–36.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340–357.

Besner, D., & Swan, M. (1982). Models of lexical access in visual word recognition. *Quarterly Journal of Experimental Psychology, 34A*, 313–325.

Coles, M.G.H., Gratton, G., Bashore, T.R., Eriksen, C.W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 529–553.

Donders, F.C. (1868–69/1969). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Psyiologish Laboratorium der Utrechtsche Hoogeschool: 1868–1869. Translated in *Acta Psychologica* (1969) *30*, 412–431.

Kučera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Meyer, D.E., Abrams, R.A., Kornblum, S., Wright, C.E., & Smith, J.E.K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, 95, 340–370.

Meyer, D.E., Osman, A.M., Irwin, D.E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26, 3–67.

Osman, A.M., Kornblum, S., & Meyer, D.E. (1986). The point-of-no-return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 243–258.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

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

(Received 5/29/90; Revision accepted 1/29/91)