

# Feedback functions for reinforcement: A paradigmatic experiment

DRAZEN PRELEC and R. J. HERRNSTEIN  
*Harvard University, Cambridge, Massachusetts 02138*

A reinforcement schedule states a rule for obtaining reinforcement as a function of behavior actually emitted and perhaps as a function of additional variables. These functions are here called "feedback functions." Behavior actually emitted is, in turn, a function of obtained reinforcement. This reciprocal interdependency was quantified for an experiment in which pigeons chose among either three or two alternatives. Shifting from three to two alternatives, and vice versa, produced changes in the distribution of responding which were approximately accounted for by equations that combined the feedback functions with the matching law for reinforced responding. These equations predicted, among other things, a violation of the constant-ratio rule of formal choice theory and an absence of simultaneous contrast effects between response alternatives reinforced on variable-interval schedules. Both predictions were approximately confirmed.

Studies of reinforced behavior by animal and human subjects (de Villiers & Herrnstein, 1976) have suggested the following simple equation to describe response strength:

$$B_1 = \frac{kR_1}{R_1 + R_e} \quad (1)$$

$B_1$  is the rate of responding sustained by a given level of reinforcement,  $R_1$ ;  $k$  is the asymptotic rate of responding for  $B_1$ , approached as  $R_1$  dominates the denominator on the right, or the total amount of reinforced behavior when different forms of behavior are measured on a common scale (Herrnstein, 1974);  $R_e$  is the reinforcement which is not associated with  $B_1$ , so that the denominator expresses total reinforcement.  $k$  and  $R_e$  are typically handled as parameters estimated from data on  $B_1$  and  $R_1$ .

Given a choice between two forms of behavior,  $B_1$  and  $B_2$ , sustained by two sources of reinforcement,  $R_1$  and  $R_2$ , and also given the same values of  $k$  and  $R_e$  for both behaviors, Equation 1 yields the following prediction for choice:

$$\frac{B_1}{R_1} = \frac{B_2}{R_2} \quad (2)$$

This work was supported by Grant MH-15494 from the NIMH to Harvard University. We are grateful to Gene Heyman and Will Vaughan for helpful comments and criticisms. Reprints may be obtained from the second author, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

This relationship has also been widely substantiated, at least approximately (de Villiers, 1977; Herrnstein, 1970), for two or more choices. Because of the symmetry expressed by Equation 2, this account of response strength is commonly called the "matching law."

In Equations 1 and 2,  $R_1$  and  $R_2$  refer to reinforcements actually received by the subject, not just available reinforcements. Since reinforcements actually received customarily depend on the subject's behavior, the equations do not contain a true independent variable, a variable external to the subject's responding. In principle, the foregoing equations omit a class of "feedback functions" accounting for reinforcements actually delivered to a subject.

In general, a feedback function should take the form:

$$R_i = f(B_i, r_i, a_i) \quad (3)$$

That is to say, reinforcements actually received,  $R_i$ , will be some function of some or all of behavior actually emitted,  $B_i$ , of reinforcements available for presentation,  $r_i$ , and of possible other factors,  $a_i$ , such as the presence of various signaling stimuli. Whatever the feedback function for an environment is, it would be inserted in the relevant response strength equation and the resulting new equation would be solved for  $B_i$ . The result should then predict behavior as a function of the true independent variables.

The present study demonstrates the procedure for taking into account feedback functions. We have chosen a procedure in which the subjects are reinforced on a variable-ratio (VR) schedule and either one or two variable-interval (VI) schedules. This particular configuration was chosen because its feedback functions lead

to several surprising predictions. First, they imply that the preference for the VR schedule vis-a-vis a VI schedule should depend on the presence or absence of the second VI schedule. This violates the common assumption in operant research (Catania, 1966; de Villiers, 1977; Herrnstein, 1970) of relative stability of preference measures. It also violates the closely related "constant-ratio rule" or the "independence from irrelevant alternatives" of formal choice theories in psychology and economics (Luce, 1959), which, simply stated, is that "the ratio of the likelihood of choosing *a* to the likelihood of choosing *b* is a constant irrespective of the number or composition of other alternatives in the set presented for choice" (Atkinson, Bower, & Crothers, 1965, p. 140). The feedback functions also imply an absence of contrast effects between the two VI schedules, thus violating the ubiquitous finding of mutually inhibiting interactions between the responding on one VI schedule and the reinforcement from another (de Villiers, 1977; Herrnstein, 1970).

## METHOD

### Subjects

Four food-deprived pigeons at 80% of free-feeding weights served as subjects. The subjects had no prior experimental histories.

### Apparatus

A standard operant conditioning chamber for pigeons was used. For some sessions, the chamber contained three standard response keys, each requiring a force of about 15 g (.15 N) to operate. The three keys were side by side on one wall of the chamber. For other sessions, the center key was blocked out with a square piece of aluminum, leaving the two side keys. The feeder opening, a 2-in. square, was under the center key. Pecking any key operated a relay to produce an audible click. A white masking noise sounded continuously during sessions. Programming and recording were done by PDP-8 computer (Digital Equipment Corporation).

### Procedure

Sessions lasted 45 min or until 60 food reinforcements (3 sec each) were presented, whichever came first. Magazine training and keypeck training were accomplished in the standard manner in two sessions. Thereafter, pecking at the center key (when present) was always reinforced on a VI schedule (Fleshler & Hoffman, 1962). Pecking at the right key was reinforced on the same schedule, though it was programmed independently. Pecking at the left key was reinforced on a VR schedule. Reinforcement could never occur within 1 sec of a switch to one of the keys with a VI schedule (COD 1 sec). There was no COD for the variable-ratio key. This procedure ran for 52 sessions, either with all three keys available or with the center one covered and inaccessible to the subjects.

Values of the schedules were selected so as to sustain responding on all available keys. This required minor adjustments of schedule parameters during approximately the first 3 weeks of sessions. No changes were made during the sessions used for data analysis. For subjects 51 and 52, the final VI schedules averaged 60 sec and the final VR schedule averaged 50 responses. For subjects 53 and 54, the corresponding values were 50 sec and 67 responses. All schedules used the Fleshler-Hoffman progression.

## RESULTS AND DISCUSSION

### Derivations of Feedback Functions

The feedback function for VR schedules is the simplest of all (Herrnstein & Loveland, 1975). It stipulates only that a certain fraction of pecks operate the food magazine, as follows:

$$R_1 = \frac{B_1}{C_1} \quad (4)$$

The obtained rate of reinforcement,  $R_1$ , is  $1/C_1$  of the actual rate of responding,  $B_1$ .

For VI schedules, the feedback function is more complex. A series of intervals sets minimum times between successive reinforcements for pecking at a given key. After each interval, the next peck at the key (not falling within a COD period) produces reinforcement. If the intervals are exponentially distributed, as they were in this case, the feedback function (Baum, 1973) for the center key approximates:

$$R_2 = \frac{B_2 r_2}{B_2 + \alpha_2 r_2} \quad (5)$$

$R_2$  is the rate of reinforcement actually obtained when the rate of responding is  $B_2$ ;  $r_2$  is the maximum possible rate of reinforcement programmed by the schedule (in this case, either 60 or 72 reinforcements/h.) If behavior had been randomly distributed in time and if there had been no COD, then  $\alpha_2 = 1.0$  and Equation 5 would be an exact, not an approximate, feedback function.  $\alpha_2$  is most likely to rise about 1.0 because of the COD and because of nonrandom shifts in the average rate of responding. The feedback function for the right key ( $B_3$ ) is the same as Equation 5 with suitable changes in subscript.

With three keys present, there are four feedback functions to consider: one for each of the three response keys ( $B_1$ ,  $B_2$ , and  $B_3$ ) and one for the background reinforcement,  $R_e$ , associated with a background of unmeasured behavior,  $B_e$ . It has been shown (Herrnstein, 1974) that the existence of  $B_e$  is a corollary of the matching law itself, not an additional premise. We do not know the reinforcement schedules giving rise to  $R_e$ , since it is not under experimental control. However, from Equations 2 and 4, the feedback function nevertheless follows:

$$\frac{B_e}{R_e} = \frac{B_1}{R_1} = C_1$$

$$R_e = \frac{B_e}{C_1} \quad (6)$$

Equation 6 is, in fact, a feedback function for unprogrammed reinforcement ( $R_e$ , see Equation 1) expressed in terms of the ratio requirement for key 1 (i.e.,  $C_1$ ). It differs from the feedback functions given above in that it presupposes the matching law (i.e., Equation 2). In contrast, the feedback functions for the VR (Equation 4) and the VI (Equation 5) schedules are true irrespective of the underlying law of response strength.

A ratio schedule that is being responded to has the property of entering into the relationship between all concurrent activities and their respective reinforcements, given the matching law. Consider the center key,  $B_2$ . Assuming Equations 2, 4, and 5, we may write:

$$\frac{B_2}{R_2} = \frac{B_1}{R_1}$$

$$B_2 = C_1 \frac{B_2 r_2}{B_2 + \alpha_2 r_2} \quad (7)$$

Solved for  $B_2$ , this is a quadratic with solutions:

$$B_2 = r_2(C_1 - \alpha_2); 0 \quad (8)$$

Analogously,

$$B_3 = r_3(C_1 - \alpha_3); 0 \quad (9)$$

As a function of programmed reinforcement rate, variable-interval responding (i.e.,  $B_2$  and  $B_3$ ) either obeys the nondegenerate solution or ceases altogether, according to this analysis. A degenerate (i.e., 0) solution is to be expected in any plausible equation for responding as a function of programmed, rather than obtained, reinforcement, for when the subject fails to earn the programmed reinforcement, its unreinforced behavior should cease. The two solutions in Equations 8 and 9 state the two possibilities for behavior as a function of programmed reinforcement.

The matching law implies that the total amount of behavior,  $k$ , is invariant (Herrnstein, 1974); therefore, from Equations 6, 8, and 9, it follows that responding to the variable-ratio key obeys the equation:

$$B_1 = k - [B_2 + B_3 + B_e]$$

$$= k - [r_2(C_1 - \alpha_2) + r_3(C_1 - \alpha_3) + C_1 R_e] \quad (10)$$

but only when all three keys are concurrently available and being responded to at all. When the center key ( $B_2$ ) is eliminated, which is tantamount to  $r_2 = 0$ :

$$B_1 = k - [r_3(C_1 - \alpha_3) + C_1 R_e] \quad (11)$$

In summary, this analysis says that responding on the two VI schedules ( $B_2$  and  $B_3$ ) depends on their individual programmed reinforcement rates ( $r_2$  and  $r_3$ , respectively), on the size of the ratio requirement ( $C_1$ ), and on a parameter expressing non-randomness ( $\alpha_2$  and  $\alpha_3$ , respectively). More to the point, it says that variable-interval responding here is unaffected by reinforcements to other alternatives, hence will show no contrast effects. A ratio schedule that is being responded to thus decouples variable-interval behavior from the interactions and contrast effects usually seen, except insofar as they may show up in the distribution parameters,  $\alpha_2$ , and  $\alpha_3$ .

The ratio behavior,  $B_1$ , however, should show considerable sensitivity to the presence or absence of responding to the center key,  $B_2$ . In fact, if  $\alpha_3$  and  $R_e$  do not change, the behavior that disappears from  $B_2$  when the center key is blocked out should simply transfer<sup>1</sup> to  $B_1$ . Consequently, the proportion  $B_1 / (B_1 + B_3)$  should rise when  $B_2$  is eliminated and fall when it is restored.

**Data**

Average results for the four subjects are plotted in Figure 1. Session numbers, starting with Session 34, are along the x-axis and responses per minute for each key are along the y-axis. When the center key was present, pecking that key ( $B_2$ ) was reinforced on the same VI schedule as pecking the right key ( $B_3$ ). Responding on the left key ( $B_1$ ) was always reinforced on the VR schedule.

The center key was eliminated on three occasions, for two, four, and one session. Each time, there was a large increase in responding to the left, variable-ratio key and a much smaller increase in responding to the right, variable-interval key. These effects are qualitatively consistent with the derivations from the matching equations just presented. The approx-

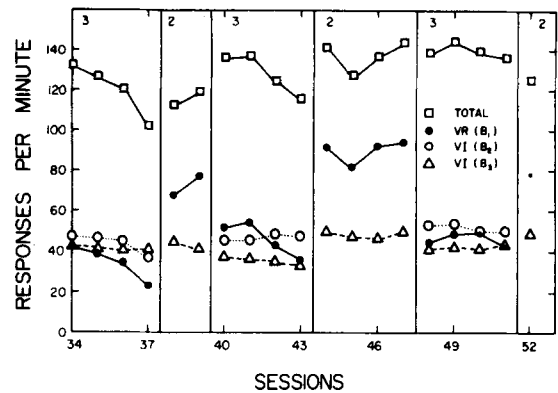


Figure 1. Rate of responding on each available key and in total from the 34th session on, averaged over subjects. The presence or absence of the center key determined whether there were three or two available keys, as indicated by the numeral in the top of each panel.

Table 1  
Summary of Data (Responses/Min, Standard Deviations, and Coefficients of Variation) for Sessions 34-52

		Subject 51			Subject 52			Subject 53			Subject 54		
		VR (B <sub>1</sub> )	VI (B <sub>2</sub> )	VI (B <sub>3</sub> )	VR (B <sub>1</sub> )	VI (B <sub>2</sub> )	VI (B <sub>3</sub> )	VR (B <sub>1</sub> )	VI (B <sub>2</sub> )	VI (B <sub>3</sub> )	VR (B <sub>1</sub> )	VI (B <sub>2</sub> )	VI (B <sub>3</sub> )
Responses	3 key	62	51	45	47	40	37	41	55	39	18	46	42
	2 key	123		51	77		45	88		39	40		54
$\sigma$	3 key	20	3.8	8.8	12.1	4.5	6.2	11.9	7.3	5.5	9.5	8.8	8.0
	2 key	13.8		7.1	12.7		7.8	11.7		7.9	23.3		15.0
$\sigma$ /Mean	3 key	.32	.07	.20	.26	.11	.17	.29	.13	.14	.53	.19	.19
	2 key	.11		.14	.16		.17	.13		.20	.58		.28

imate constancy of total responding with respect to the presence or absence of the center key is also consistent with the present analysis.<sup>2</sup>

Table 1 presents individual data for the four subjects. The simplest way to quantify the effects of inserting and removing the center key is to compare average rates of responding during three- and two-key conditions. Table 1 (upper portion) shows that the responding lost when the center key is blocked out mostly switches to the left key. Averaged over the four subjects, the left key gained 40 responses/min, the center key lost 48 responses/min, and the right key gained 7 responses/min.

Although Table 1 approximates the prediction that responding to the variable-ratio key should absorb the effects of deleting and restoring the opportunity to peck at the center key, there are two obvious departures. One is the change in responding at the right key (B<sub>3</sub>), whereas the analysis implies no change at all. According to Equation 9, responding to the right key depends on three quantities:  $r_3$ —the programmed rate of reinforcement on the VI schedule;  $C_1$ —the value of the ratio schedule on the left key; and  $\alpha_3$ —the time distribution parameter. Only  $\alpha_3$  could have changed here. This parameter equals 1.0 when responding is randomly distributed in time and there is no COD. The COD can increase  $\alpha$  in two ways: it adds a small interval to the inter-reinforcement time whenever the subject switches to the center or right key and it probably introduces some nonrandomness into the temporal pattern of responding. It is therefore likely that eliminating the center key would reduce the value of  $\alpha_3$  simply by eliminating some gaps in the temporal pattern of responding at the right key. Consequently, we would predict a small increase in responding to the right key when the center key is removed and an approximately equal decrease when it is restored. In addition, the analysis in Equations 8-12 concerns asymptotic responding and may therefore fail to account for transitory effects.

A change in the value of  $\alpha_3$  may also explain the other deviation from the analysis. Responding to the variable-ratio key gained only 40 responses/min (on

the average) when blocking out the center key eliminated 48 responses. However, a decrease in the value of  $\alpha_3$  predicts just such a difference, as is apparent from a comparison of Equations 10 and 11.

Responding on each of the interval schedules depends on its own value of  $\alpha$  and on fixed parameters of the procedure. Responding on the ratio schedules, in contrast, depends on *all* of the other forms of responding in the situation: B<sub>2</sub>, B<sub>3</sub>, and B<sub>e</sub> (see Equations 8-11) in the three-key procedure. It is to be expected, therefore, that ratio responding would show more variability than interval responding (and also more variability in the three-key procedure than in the two-key one, since there is an additional degree of freedom). Table 1 presents the relevant data. Shown are the standard deviations ( $\sigma$ ) and coefficients of variation ( $\sigma$ /mean), separated into two- and three-key procedures. For each subject, with either two or three keys present, the standard deviations for ratio responding were larger than for interval responding. The coefficients of variation attempt to compensate for variations in the overall level of responding. Except for subject 54, the highest coefficient of variation was for ratio responding in the three-key procedure and the lowest was for interval responding on the center key when it was present. For subject 54, the coefficient of variation was slightly (.05) higher for the VR schedule in the two-key procedure than in the three-key procedure.

### Test of Constant-Ratio Rule

Formal choice theories often assume the "constant-ratio rule," or the "independence from irrelevant alternatives" (Atkinson, Bower, & Crothers, 1965; Luce, 1959). As noted earlier, this rule says that the relative preference in a binary choice is unaffected by changes in the larger context in which it occurs. The stability of relative preference is also a common finding in operant research (Catania, 1966; de Villiers, 1977; Herrnstein, 1970). Insofar as the present findings obey the matching equations as used here, they violate this supposition. Figure 2, which compares preferences before and after each transition from three to two alternatives, and vice versa, shows that our subjects

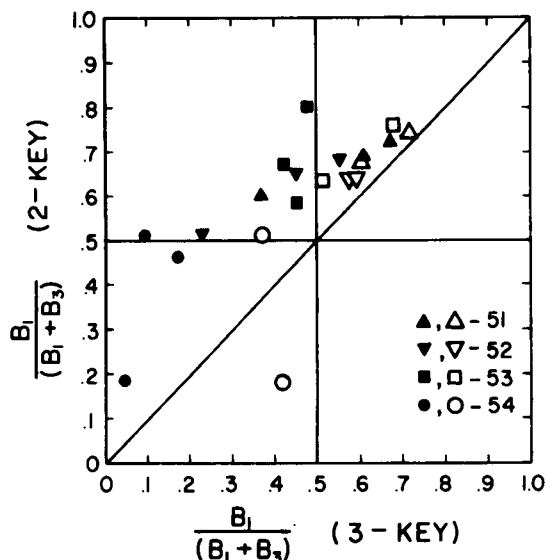


Figure 2. Preference for the VR schedule on the left key ( $B_1$ ) vis-à-vis the VI schedule on the right key ( $B_3$ ), measured by the proportion of the rates of responding,  $B_1/(B_1 + B_3)$ , for each subject. Each point plots the proportion for the session on either side of a transition from three available keys (x-axis) to two to three keys; for open data points, the transitions were from two to three keys; for closed points, transitions were from three to two keys. Each point therefore plots the value of the proportion for a single session against its value for the very next session, for individual subjects.

have consistently violated the constant-ratio rule.

The y- and x-axes express responding on the VR schedule ( $B_1$ ) as a proportion of that responding plus variable-interval responding on the key that was always present ( $B_3$ ), for just one session before and after each transition from three to two alternatives, and vice versa. The y-axis shows the relative strength of ratio responding compared to interval responding on the right key,  $B_1/(B_1 + B_3)$ , when the center key was blocked out; the x-axis shows the same proportion,  $B_1/(B_1 + B_3)$ , when the center key was accessible. There were three transitions from three to two alternatives and two going the other way. For four subjects, this yields 20 tests of the effect of context on preference.

The constant-ratio rule says that the points should approximate the diagonal. The present analysis says that blocking out the center key increases  $B_1$  but not  $B_3$ , so the points should rise above the diagonal.<sup>3</sup> The points in Figure 2 lie above the diagonal in every case but one. The open circles, which lie closer to the diagonal, are for transitions from two to three alternatives, suggesting that they were somewhat slower than the transitions from three to two (closed circles). The points in the upper left quadrant represent preference reversals: the ratio alternative was pecked less than .5 in the three-key procedure and more than .5 in the two-key procedure. Each subject contributed at least one example of preference reversal.

Formal choice theorists (Luce, 1977) have uncovered a class of violations of the constant-ratio rule based on similarity or discriminability among choices. Figure 2 appears to be a different sort of violation. Here, the constant-ratio rule is undermined by differences in the underlying feedback functions for ratio and interval alternatives, given the matching law. Deleting one alternative frees a certain amount of behavior to be allocated among the remaining alternatives. With the feedback functions for ratio and interval schedules, most of the behavior thus freed transfers to the ratio alternative, thereby violating the constant-ratio rule without regard to discriminability or to changes in motivational state.

REFERENCES

ATKINSON, R. C., BOWER, G. H., & CROTHERS, E. J. *An introduction to mathematical learning theory*. New York: Wiley, 1965.

BAUM, W. M. The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior*, 1973, 20, 137-153.

CATANIA, A. C. Concurrent operants. In W. K. Honig (Ed.), *Operant behavior*. New York: Appleton-Century-Crofts, 1966.

DE VILLIERS, P. A. Choice in concurrent schedules and a quantitative formulation of the law of effect. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

DE VILLIERS, P. A., & HERRNSTEIN, R. J. Toward a law of response strength. *Psychological Bulletin*, 1976, 83, 1131-1153.

FLESHLER, M., & HOFFMAN, H. S. A progression for generating variable interval schedules. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 529-530.

HERRNSTEIN, R. J. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 1970, 13, 243-266.

HERRNSTEIN, R. J. Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, 1974, 21, 159-164.

HERRNSTEIN, R. J., & LOVELAND, D. H. Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior*, 1975, 24, 107-116.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

LUCE, R. D. The choice axiom after twenty years. *Journal of Mathematical Psychology*, 1977, 15, 215-233.

NOTES

1. A different approach to ratio responding gives a closely related answer. Inserting the feedback function for ratio responding (Equation 4) directly into the matching law gives:

$$B_1 = \frac{k \frac{B_1}{C_1}}{\frac{B_1}{C_1} + R_2 + R_3 + R_e}$$

$R_2$ ,  $R_3$ , and  $R_e$  are obtained reinforcements associated with the behaviors other than  $B_1$ . The equation is a quadratic with solutions:

$$B_1 = k - C_1(R_2 + R_3 + R_e); 0.$$

However, Equations 6 and 7 imply that, if the ratio is responded to, then:

$$C_1(R_2 + R_3 + R_e) = B_2 + B_3 + B_e.$$

Consequently, the nondegenerate solution just above may be rewritten as:

$$B_1 = k - (B_2 + B_3 + B_e),$$

which is Equation 10. In short, the derivation in the text presupposes that the ratio is responded to. Here, in contrast, we make no such presupposition and the analysis yields the additional option that the ratio is not responded to at all, but is otherwise the same. It is well known, in fact, that ratio behavior may "strain," which is to say, cease. If it had done so in this experiment, then the constraints on variable-interval responding imposed by  $C_1$  in Equation 7 and, implicitly, in Equation 9 would have been lifted. If ratio responding had ceased here, we would doubtless have observed contrast effects between the two interval schedules. As noted earlier, however, we chose schedule parameters to maintain responding to accessible keys.

2. Earlier work (by Herrnstein and Loveland, summarized in de Villiers, 1977) has shown that pigeons approximate the matching equation on concurrent VI VR schedules. To match, the subject's variable-interval responding must produce the same ratio of responses to reinforcements on the interval schedule as is programmed by the variable ratio. This, it should be noted, is true whatever parameter values are set for VI and VR schedules, just so long as the subjects match. Matching in concurrent VI VR schedules is therefore a case in which the subject's responding must obey unusual constraints. The subjects here approximately matched, although they were not exposed to each concurrent schedule quite long enough to stabilize responding or to approximate matching as closely as has been observed before with comparable schedules.

3. The difference between the proportions of ratio responding for two- and three-key procedures is given in the following derivation:

$$\begin{aligned} & \frac{B_1'}{B_1' + B_3'} - \frac{B_1}{B_1 + B_3} \\ &= \frac{k - [r_3(C_1 - \alpha_3) + C_1 R_e]}{k - C_1 R_e} \\ & \quad - \frac{k - [r_2(C_1 - \alpha_2) + r_3(C_1 - \alpha_3) + C_1 R_e]}{k - [r_2(C_1 - \alpha_2) + C_1 R_e]} \\ &= \frac{r_2 r_3 (C_1 - \alpha_2)(C_1 - \alpha_3)}{(k - C_1 R_e)^2 - r_2(k - C_1 R_e)(C_1 - \alpha_2)}. \end{aligned}$$

The superscripted measures of responding are for the two-key procedure. Given the derivation of this equation, the quantity must be positive (i.e., the points should lie above the diagonal), and it is constant (i.e., the points should lie along a line parallel to the diagonal) if  $\alpha_3$  and  $R_e$  are not affected by the transitions between three and two keys.

(Received for publication September 12, 1977;  
revision accepted November 16, 1977.)