Psychological Review
1990, Vol. 97, No. 2, 309-312

0033-295X/90/$00.75

THEORETICAL NOTE

A Note on the Stop-Signal Paradigm, or How to Observe the Unobservable

Hans Colonius
Department of Psychological Sciences
Purdue University

A new theoretical analysis of the stop-signal paradigm is proposed. With the concepts of crude and net hazard functions, the nonobservable control-latency distribution can be estimated from observable reaction times. This result allows a test of Logan & Cowan's (1984) model without any simplifying assumptions.

In studies using response time as the major dependent variable, the stop-signal procedure has been used by several investigators in an attempt to reveal the time course of hypothesized underlying sensory or cognitive processes or both (e.g., Lappin & Eriksen, 1966; Ollman, 1973). In this paradigm, subjects are given a primary reaction time task to perform, and a stop signal that tells subjects to withhold their responses is sometimes presented shortly after the primary stimulus. Logan and Cowan (1984) presented a theory of inhibition of thought and action that focuses on this paradigm. The purpose of this note is to present a new theoretical analysis of the stop-signal paradigm that allows a more stringent test of Logan and Cowan's model and, in fact, of all theories based on this paradigm.

Logan and Cowan's (1984) model accounts for response inhibition in terms of a "horse race" between two sets of processes, one that generates a response for the primary task and one that reacts to the stop signal:

> If the primary-task process finishes before the stop-signal process, the response is executed; if the stop-signal process finishes before the primary-task process, the response is inhibited. To model this situation, the finishing times of the primary-task and the stopping process are assumed to be independent random variables. (pp. 298-299)

Note that the primary-task processing time is observable both in the normal condition (i.e., the no stop-signal condition) and in the stop-signal condition given it "wins the race"; because of the nature of this paradigm, however, the stop-signal processing time can never be observed. A major goal of a formal analysis of this paradigm, then, is to derive an estimate of this unobservable control latency. Logan and Cowan achieved this by assuming the stop-signal processing time to be constant. Assuming, in addition, that the control-latency distribution peaks around its mean, they offered an argument that this constancy assumption yields a reasonable approximation to what would follow from a true horse-race mechanism. In a later article, Logan and Burkell (1986) introduced specific parametric assumptions for the control-latency distribution.

Rather than arguing about the soundness of this approximation, in what follows I demonstrate that under the assumptions of Logan and Cowan's (1984) race model, the entire distribution of control latencies can be estimated in a straightforward manner from directly observable reaction times (observing the unobservable, so to speak). Moreover, this analysis permits an overall test of the assumptions of the race model that is more direct than in previous analyses.

The Horse-Race Model

To state explicitly the assumptions that go into the race model, some terminology is needed. Let T_p and T_s denote random variables representing the primary-task processing time and the stop-signal processing time, respectively. With t_d denoting the stop-signal delay, that is, the time between the onset of the primary-task stimulus and the stop signal, the observable entities $\Pr[T_p > t \mid \text{no stop signal}]$ and $\Pr[T_p > t \mid T_p < T_s + t_d]$ are defined for all nonnegative real numbers t. The first entity is a survival distribution, representing, for each instant of time t, the probability that the primary-task processing takes longer than t ms in the experimental condition without a stop-signal presentation. The second entity is a conditional survival distribution, representing the probability that the primary-task processing takes longer than t ms given it wins the race against the processing of a stop signal presented t_d ms later than the primary-task stimulus. Obviously, both distributions can be estimated from empirically observed relative-frequency distributions in the normal and the stop-signal conditions. The probability of observing a response in the stop-signal condition, $\Pr[T_p < T_s + t_d]$, is abbreviated $q(t_d)$. Moreover, there are two unobservable entities, (a) the control-latency survival distribution, $\Pr[T_s > t \mid t_d]$ (here and later, $\Pr[\quad \mid t_d]$ refers to a condi-

This note is based on a talk presented at the Hoosier Mental Life Conference, Indiana University, Bloomington, May 1989. The work was partially supported by a grant from Stiftung Volkswagenwerk.

I thank James Townsend and Richard Schweickert for their helpful comments on an earlier version of this note.

Correspondence concerning this article should be addressed to Hans Colonius, Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907.

tional probability, with conditioning on the event {stop-signal presented t_d ms after primary stimulus}), and (b) the bivariate survival distribution for T_p and T_s given by $\Pr[T_p > t_p) \cap (T_s > t_s) | t_d]$.

The following specific assumptions are central to Logan and Cowan's (1984) model:

[A1] *Context independence.* The primary-task processing time distribution is the same whether or not a stop signal is presented; or, more formally, $\Pr[T_p > t | \text{no stop signal}] = \Pr[T_p > t | t_d]$ for all t, t_d.

[A2] *Stochastic independence.* Primary-task processing time and stop-signal processing time are stochastically independent; formally, $\Pr[(T_p > t_p) \cap (T_s > t_s) | t_d] = \Pr[T_p > t_p | t_d]\Pr[T_s > t_s | t_d]$ for all t_p and t_s.

[A3] *Stop-signal delay invariance.* The distribution of the stop-signal processing time is the same for all values of the stop-signal delay; formally, $\Pr[T_s > t | t_d] = \Pr[T_s > t | t'_d]$ for all t_d, t'_d, and t.

Because each of the above assumptions contains a statement about unobservable entities, none of them is testable if considered separately. However, as our subsequent analysis will demonstrate, Assumptions [A1] and [A2] taken together lead to a distribution- and parameter-free estimate of the control latency, thereby allowing a test of Assumption [A3].

Observing the Unobservable

The next step of this analysis involves defining two types of *hazard functions,* a concept that has proved to be a useful tool in reaction time analysis (see, e.g., Bloxom, 1984; Colonius, 1988; Luce, 1986; Ratcliff, 1988; Thomas, 1971; Townsend & Ashby, 1983). The *net hazard function* for T_p is defined by

$$g_p(t) = \lim_{\delta \to 0+} (1/\delta) \Pr[t < T_p \le t + \delta | T_p > t].^1$$

Intuitively, $g_p(t)$ is the instantaneous tendency of T_p taking on the value t given that the primary-task processing has not been terminated by time t or been given a stop signal. On the other hand, the *crude hazard function* for T_p is defined by

$$h_p(t; t_d) = \lim_{\delta \to 0+} (1/\delta) \Pr[t < T_p \le t + \delta | (T_p > t) \cap (T_s + t_d > t)].$$

The interpretation of $h_p(t; t_d)$ is analogous to that of $g_p(t)$, except that now conditioning occurs with respect to the event that both primary-task processing and stop-signal processing have not been terminated by time t. The definition makes obvious that the net hazard rate, $g_p(t)$, can be estimated from the density $f_p(t)$ and the survival distribution $\Pr[T_p > t]$, both of which are observable (or, more precisely, statistically estimable). On the other hand, the crude hazard rate, $h_p(t; t_d)$, is not estimable because it involves the unobservable bivariate survival distribution for T_p and T_s. The main result is a consequence of the following lemma about these hazard functions. (For proofs, see the appendix.)

Lemma. If Assumptions [A1] and [A2] hold, then the net and the crude hazard rate for T_p are identical; that is, $g_p(t) = h_p(t; t_d)$ for all t and any t_d.

Theorem. If Assumptions [A1] and [A2] hold, then $\Pr[T_s + t_d > t | t_d] = f^*_p(t; t_d)q(t_d)/f_p(t)$ for all t and t_d, where f_p denotes the density for T_p, $f^*_p(t; t_d)$ denotes the conditional density for T_p given the event $\{T_s + t_d > T_p\}$ and a stop signal presented with a delay of t_d ms, and $q(t_d)$ stands for the probability of observing a response with a stop signal of t_d ms delay.

For the interpretation of this result, two comments are in order. First, the unobservable survival distribution of the stop-signal processing time, $\Pr[T_s > t | t_d]$, is expressible entirely in terms of observable (or, more precisely, statistically estimable) entities. Second, the theorem provides a test of Assumption [A3]. In fact, according to this assumption, varying the stop-signal delay t_d should only result in shifting the survival distribution $\Pr[T_s > t | t_d]$ horizontally (in the direction of the time axis). A violation of this shift invariance property, however, could be due to a failure of any of Assumptions [A1], [A2], and [A3].

This last observation points to a more general principle in testing race models (see also Ashby & Townsend, 1986; Vorberg, Colonius, & Schmidt, in press). Assume an experimental situation in which each of the processes competing in the joint condition can also be observed in a separate condition under which no race takes place. The context-independence assumption would hold that the distribution functions in the separate conditions are identical to the corresponding marginal distributions in the joint condition. However, because in the joint condition only the processing time of the winner of the race is observable, the marginals are not observable. Thus, stochastic independence is not testable without the assumption of context independence and vice versa.[2] Interestingly, in the joint condition, even if in addition to the processing time of the winning process the winner's identity is observable as well, stochastic independence is not testable (this is known as the nonidentifiability result in competing risk theory; see, for example, David & Moeschburger, 1978; for an application in probabilistic choice theory, see Marley & Colonius, in press).

Discussion and Conclusion

Before some practical aspects of using the above result are discussed, its significance for theories about the psychological processes involved in the stop-signal paradigm should be assessed. First, given that Assumptions [A1] and [A2] hold, the theorem provides a means for estimating the unobservable control-latency distribution without any further simplifying assumptions about the underlying race mechanism, such as the control latency's being constant (Logan & Cowan, 1984). Second, the model's prediction that varying the stop-signal delay should result only in shifting the corresponding control-latency distributions accordingly can be tested directly by inspecting the estimates of the control-latency distribution. If this test fails, however, my analysis does not give a clue as to which of these assumptions has to be dropped. Consequently, all arguments concerning the plausibility of each of these assumptions (see,

[1] In accordance with most of the reaction time literature, existence of a density for all processing-time random variables is assumed here. Note that $\delta \to 0+$ stands for $\delta \to 0$ and $\delta > 0$. See the appendix for the equivalence of this definition of the hazard function with the usual one in terms of density and survival distribution.

[2] See also the discussion of Miller's (1982) race model in Luce (1986, p. 128).

for example, the discussion in Logan & Cowan, 1984) remain potentially relevant.

Another point is worth mentioning. It is sometimes suggested (e.g., Logan & Cowan, 1984) that the primary-task processing time (T_p) should be broken into two additive subprocesses: an early process that can be inhibited by the stop signal and a subsequent ballistic process that cannot be inhibited once it has started. If the ballistic component is assumed to be a constant, t_m (motor time) for example, the analysis would still hold with $t - t_m$ replacing t. If the ballistic component is assumed to be randomly varying, then matters become much more difficult. In certain empirical situations in which some information about the distribution functions of the components is available (e.g., an exponential ballistic component), decomposition techniques proposed by Luce (1986, chap. 3; see also Ashby & Townsend, 1980) may be applicable to estimating the density and the conditional density of the early process, which would then replace f_p and f_p^*, respectively, in the theorem. However, note that the need to introduce a ballistic component has been challenged recently in an interesting study by de Jong, Coles, Logan, and Gratton (1990) with electromyogram and continuous response measures as well as reaction time.

Next, estimating the control-latency distribution, $\Pr[T_s > t \mid t_d]$, according to the theorem involves computation of the ratio of two densities, $f_p^*(t)$ and $f_p(t)$. Two comments are in order here. First, replacing these densities by their corresponding empirical relative-frequency functions probably results in an unstable estimate for the control latency, particularly at regions of the time axis where the density in the denominator, $f_p(t)$, is near 0. Fortunately, there are several smooth, nonparametric density-estimation methods available (see, for example, Silverman, 1986) that should be applied to the raw data. Although this application will in general yield more stable estimates of the control-latency distributions, the feasibility of this estimation may still depend on the particular data set under investigation. Finally, a practical graphical test to check whether variation of the stop-signal delay (t_d) results in generating a shift family of control-latency distributions is provided by quantile–quantile plots. This technique, including weighted least squares estimators, has been explained fully by Nair (1984).

References

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

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Bloxom, B. (1984). Estimating response time hazard functions: An exposition and extension. *Journal of Mathematical Psychology, 28,* 401–420.

Colonius, H. (1988). Modeling the redundant signals effect by specifying the hazard function. *Perception & Psychophysics, 43,* 604–606.

David, H. A., & Moeschburger, M. L. (1978). *The theory of competing risks.* London: Charles Griffin & Co.

de Jong, R., Coles, M. G. H., Logan, G. D., & Gratton, G. (1990). In search of the point of no return: The control of response processes in speeded choice reaction performance. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 164–182.

Lappin, J. S., & Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *Journal of Experimental Psychology, 72,* 805–811.

Logan, G. D., & Burkell, J. (1986). Dependence and independence in responding to double stimulation: A comparison of stop, change, and dual-task paradigms. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 549–563.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91,* 295–327.

Luce, R. D. (1986). *Response times. Their role in inferring elementary mental organization.* New York: Oxford University Press.

Marley, A. A. J., & Colonius, H. (in press). The "horse-race" random utility model and its competing risks interpretation. *Journal of Mathematical Psychology.*

Miller, J. O. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279.

Nair, V. N. (1984). On the behavior of some estimators from probability plots. *Journal of the American Statistical Association, 79,* 823–831.

Ollman, R. T. (1973). Simple reactions with random countermanding of the "go" signal. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 571–581). New York: Academic Press.

Ratcliff, R. (1988). A note on mimicking additive reaction time models. *Journal of Mathematical Psychology, 32,* 192–204.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman & Hall.

Thomas, E. A. C. (1971). Sufficient conditions for monotone hazard rate and application to latency-probability curves. *Journal of Mathematical Psychology, 8,* 303–332.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* New York: Cambridge University Press.

Vorberg, D., Colonius, H., & Schmidt, R. (in press). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology.*

(*Appendix follows on next page*)

Appendix

Proofs

Proof of the Lemma

Consider the crude hazard function

$$h_p(t; t_d) = \lim_{\delta \to 0+} (1/\delta) \Pr[t < T_p \leq t + \delta \,|\, (T_p > t) \cap (T_s + t_d > t)].$$

In terms of partially unconditioned probability, this is

$$\frac{\lim_{\delta \to 0+} (1/\delta) \Pr[(t < T_p \leq t + \delta) \cap (T_s + t_d > t) \,|\, t_d]}{\Pr[(T_p > t) \cap (T_s + t_d > t) \,|\, t_d]}. \quad \text{(A1)}$$

By stochastic independence between T_p and T_s, $\Pr[T_s + t_d > t \,|\, t_d]$ can be factored out in the numerator and in the denominator and thus cancels. Then, invoking context independence (Assumption [A1] in main text) gives $h_p(t; t_d) = g_p(t)$ for all t, completing the proof.

Proof of the Theorem

The net hazard function can be rewritten for T_p as follows:

$$g_p(t) = \lim_{\delta \to 0+} (1/\delta) \Pr[t < T_p \leq t + \delta \,|\, T_p > t]$$

$$= \{\lim_{\delta \to 0+} (1/\delta) \Pr[t < T_p \leq t + \delta]\}/\Pr[T_p > t] \quad \text{(A2)}$$

$$= f_p(t)/\Pr[T_p > t],$$

where $f_p(t)$ denotes the density for T_p. For the crude hazard function $h_p(t; t_d)$, Equation A1 may be rewritten as

$$\frac{\lim_{\delta \to 0+} (1/\delta) \Pr[(t < T_p \leq t + \delta) \cap (T_s + t_d > T_p) \,|\, t_d]}{\Pr[(T_p > t) \cap (T_s + t_d > t) \,|\, t_d]}.$$

Introducing the conditional density $f_p^*(t; t_d)$ in the numerator, the previous expression becomes

$$h_p(t; t_d) = f_p^*(t; t_d) q(t_d)/\Pr[(T_p > t) \cap (T_s + t_d) > t \,|\, t_d]. \quad \text{(A3)}$$

By stochastic independence between T_p and T_s (Assumption [A2] in main text), the denominator factors into $\Pr[T_p > t \,|\, t_d]$ and $\Pr[T_s + t_d > t \,|\, t_d]$. By the lemma, Equations A2 and A3 can be set equal:

$$f_p(t)/\Pr[T_p > t] = f_p^*(t; t_d) q(t_d)/[\Pr[T_p > t \,|\, t_d] \Pr[T_s + t_d > t \,|\, t_d].$$

By context independence, the survival distributions for T_p are identical and thus cancel out. Solving the previous equation for $\Pr[T_s + t_d > t \,|\, t_d]$ yields the theorem.

Received June 7, 1989
Revision received September 18, 1989
Accepted October 3, 1989 ■