PSYCHOMETRIKA—VOL. 58, NO. 2, 281–314
JUNE 1993

GRICE-REPRESENTABILITY OF RESPONSE TIME DISTRIBUTION FAMILIES

EHTIBAR N. DZHAFAROV

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
AND
THE BECKMAN INSTITUTE FOR
ADVANCED SCIENCE AND TECHNOLOGY

Any family of simple response time distributions that correspond to different values of stimulation variables can be modeled by a deterministic stimulation-dependent process that terminates when it crosses a randomly preset criterion. The criterion distribution function is stimulation-independent and can be chosen arbitrarily, provided it is continuous and strictly increasing. Any family of N-alternative choice response time distributions can be modeled by N such process-criterion pairs, with response choice and response time being determined by the process that reaches its criterion first. The joint distribution of the N criteria can be chosen arbitrarily, provided it satisfies certain unrestrictive conditions. In particular, the criteria can be chosen to be stochastically independent. This modeling scheme, therefore, is a descriptive theoretical language rather than an empirically falsifiable model. The only role of the criteria in this theoretical language is to numerically calibrate the ordinal-scale axes for the deterministic response processes.

Key words: response time, choice, criterion, deterministic process, stochastic process.

Introduction

Response time (RT) distributions are commonly modeled in terms of unobservable processes that develop in time and result in overt responses when the processes meet certain termination conditions. RT is a random variable due to random fluctuations postulated to occur in the termination conditions or in the processes themselves. In a simple version of this modeling scheme the processes are scalar (real-valued) functions of time, and the termination condition is satisfied when a process crosses some critical level for the first time. This critical level is traditionally referred to as the "criterion". I will use the neutral term "response process" to refer to a scalar process that has to cross a criterion to initiate an observable response.

Grice (1968, 1972; Grice, Cahnam, & Boroughs, 1984; Grice, Nullmeyer, & Spiker, 1982) proposed simplifying further this modeling scheme by assuming that for any given combination of relevant external factors (target stimulus characteristics, speed-accuracy requirements, etc.): (A) the corresponding response process is a deterministic function of time, monotonically increasing from an initial (zero) level; (B) the criterion is randomly preset at a nonnegative value that remains constant throughout the entire duration of the response process. It is further assumed that: (C) the criterion distribution may only depend on a small subset of the external factors, which normally, after an asymptotic performance level has been achieved, would not include factors that may vary from trial to trial (such as target stimulus characteristics, or foreperiod duration). These assumptions constitute the essence of the Grice modeling scheme for simple RT.

The author is indebted to Richard Schweickert, James Townsend, and Duncan Luce for valuable comments and criticism.

Requests for reprints should be sent to Ehtibar N. Dzhafarov, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61820.

0033-3123/93/0600-91410$00.75/0

For choice RT, when one of N possible responses is to be chosen in every trial, Grice proposes considering N deterministic response processes and N corresponding criteria, each process-criterion pair satisfying assumptions A, B, and C above. It is assumed that: (D) the criteria corresponding to different responses are stochastically independent; (E) response choice and response time are determined by the response process that reaches its criterion first.

To fit empirical RT distributions, Grice also makes specific assumptions (as it turns out, not quite mutually compatible, see section 2.2) concerning the shapes of criterion distributions (approximate normality) and response processes (e.g., a negative exponential function for simple RT). Such auxiliary assumptions are of no importance for the present discussion, as it focuses on the principal idea of modeling RT through stimulation-dependent deterministic response processes and stimulation-independent random criteria. Intuitively, this idea seems quite restrictive as it is, aside from auxiliary assumptions, simply because there are many obvious ways to present it as a particular or limit case of much more general schemes.

One direction of generalization is to allow for a stimulation-dependent stochasticity in response processes themselves, in addition to that in preset criteria. This opens additional possibilities, such as stochastic interdependence between criteria and processes, or (in the choice RT paradigm) stochastic interdependence between processes leading to different responses. Another direction of generalization is to view a criterion as a stochastic process (perhaps a nonstationary one) within a trial. Termination conditions for response processes can also be made more complex than a simple crossing-a-criterion rule. One could postulate, for example, a race between a deterministic process and its rate: a response is produced when the process exceeds its criterion, or when the time derivative of the process exceeds its own criterion, whichever is first. More generally, a termination condition may be a set of equations-inequalities involving high-order derivatives of response processes. A different line of generalization is to postulate that criteria may change within trials depending on response process characteristics: for example, if the response process rate at a preset criterion level is below a certain constant, the criterion resets at a higher level. These and many other directions of generalization can be combined to form still more sophisticated schemes that would include the Grice modeling scheme as their particular or marginal case. One gets the impression, therefore, that the Grice modeling scheme should be in principle empirically falsifiable. That is, one can generate families of hypothetical RT distributions (in accordance with one of the "sophisticated" models) that the Grice modeling scheme would not account for.

This intuition seems supported by the following observation. McGill (1963) proposed modeling RT in terms of stochastic response processes reaching a fixed criterion level, an idea that has been employed in several paradigmatic models (Green & Luce, 1974; Laming, 1968; Link, 1975; Pike, 1973; Ratcliff, 1978; see Luce, 1986, for a review). It is easy to show (see sections 1.3 and 2.3) that the Grice modeling scheme is formally equivalent to the following special case of the McGill modeling scheme: Any response process can be decomposed (additively or multiplicatively, depending on an arbitrary choice of scale) into a stimulation-dependent deterministic function and a stimulation-independent stationary noise. Stochastic processes of this structure have indeed been used for modeling RT distributions: one simple example is a Wiener diffusion process (Pacut, 1977; Ratcliff, 1978). However, one can easily construct stochastic processes that would not be decomposable in this fashion. "Noise" in the processes can always be made nonstationary, stimulation-dependent, or such that no monotonic transformation can make it additively or multiplicatively combined with deterministic components of the processes. Again, this observation seems to suggest

that one can construct families of hypothetical RT distributions that cannot be accounted for by the Grice modeling scheme.

This intuition is wrong, however. It is shown in this paper that the Grice modeling scheme is a descriptive theoretical language rather than an empirically falsifiable model. *Any* family of hypothetical RT distributions can be modeled according to assumptions A through E above. Assumption C is in fact unnecessarily weak: Variability in RT can always be attributed to a randomly preset criterion (or several such criteria in the choice RT paradigm) whose distribution does not depend on *any* external factors. Consequently, in the absence of additional constraints derived from considerations other than RT analysis, one never faces the necessity to generalize the Grice modeling scheme, or equivalently, to go beyond a stationary stimulus-independent noise superimposed on deterministic functions in the McGill modelling scheme.

Moreover, the choice of the criterion distributions turns out to be completely arbitrary, provided they satisfy certain (rather mild) restrictions. In particular, the mutual stochastic independence of the criteria in the choice RT paradigm can be replaced by any "well-behaved" stimulation-independent joint distribution function. Whatever the choice of the criteria, one can always find stimulation-dependent deterministic response processes that would generate any given family of RT distribution function.

The mathematical justification of these statements is very simple, and it is surprising that they have not been commonly acknowledged in the RT area. By undertaking a systematic investigation of the Grice modeling scheme, this paper, in spite of its mathematical simplicity (or maybe due to it), helps to achieve a useful purpose. It clarifies the meaning and proper usage of some fundamental notions of RT modeling, such as termination criterion, deterministic versus stochastic components of response, in/ter/dependence of processes versus stochastic in/ter/dependence of their termination times, etcetera. The analysis presented in this paper can be used to determine what in a certain RT model constitutes a set of empirically testable assumptions, and what is merely an arbitrary choice of mathematical language.

1. Grice-Representability of Time Variables

Here, the basic concepts associated with the Grice modeling scheme for simple RT will be presented, and a formal analysis made of Grice-representability of time variables stochastically depending on a set of external factors. The following notation rules are adopted throughout this paper. Boldface capital letters, $\mathbf{T}$, $\mathbf{C}$, etcetera, denote random variables. Their realizations are denoted by the corresponding lowercase italics, t, c, etcetera, whereas their distribution functions are denoted by the corresponding script letters, $\mathcal{T}$, $\mathcal{C}$, etcetera. For example, $\mathcal{C}(c)$ is the distribution function of a random variable $\mathbf{C}$ at $\mathbf{C} = c$. Script letters are also used to denote multidimensional distribution functions and some related functions (such as conditional probabilities). The proofs of all statements formulated in this paper as lemmas or theorems are given in the Appendix, even when very short or obvious.

1.1. Time-Dimensioned Random Variables.

In the simple RT paradigm the object of modelling is a family of *time-dimensioned random variables* $\mathbf{T}(\Xi)$. Here Ξ denotes all external factors on which the distribution of $\mathbf{T}$ is known to depend, such as target signal intensity, temporal separation between target and warning signals, speed-accuracy emphasis, etcetera. Some of these factors may vary from trial to trial, or even within trials, others are fixed within an experiment. For our purposes the set of all possible values of Ξ can be viewed as an abstract

indexing set for T, generally not countable. If, for example, Ξ consists of two factors, signal intensity $A \in (0, A_{sup})$, and foreperiod duration $F \in (F_{inf}, F_{sup})$, then different pairs of real numbers from these intervals will define different random variables $\mathbf{T} = \mathbf{T}(F, A)$. For simplicity, the family $\mathbf{T}(\Xi)$ will be referred to as a (single) random variable, stochastically depending on Ξ. The terms "value of Ξ" or "a given Ξ" will refer, of course, to the set of values of all components of Ξ.

The time axis origin ($t = 0$) will be assumed to coincide with the onset of a target signal (determined by appropriate conventions if the target is not a step-function in time). RTs represented by the time variables $\mathbf{T}(\Xi)$, however, may attain any real values, including very small positive values and negative values, which means that the consideration includes "false alarms", or "anticipatory responses", as well as "true" responses. Although it is reasonable to assume that RT distributions are sufficiently smooth (see section 2.1), this assumption will not be needed as far as the simple RT paradigm is concerned.

Throughout this paper a RT model is understood as an algorithm generating $\mathbf{T}(\Xi)$ (or other random variables, defined later) with precisely known theoretical distributions for every possible value of Ξ. The modeling of RT, therefore, is expressly a nonstatistical notion here: for any given Ξ, the distribution of $\mathbf{T}(\Xi)$ is assumed to be known precisely, and these distributions should be generated (predicted) by a RT model for all possible values of Ξ, a set whose cardinality may be arbitrarily large. Obviously, if $\mathbf{T}(\Xi)$ is successfully modeled in this sense, then the model is also "true" in a statistical meaning, with respect to finite-size samples from $\mathbf{T}(\Xi)$ taken for a finite number of Ξ-values.

1.2. Grice-Representability.

Let $C_\Xi(t)$ be a *response process*, a scalar (real-valued) deterministic function of time depending on Ξ: for a given value of Ξ and a given moment t, the value of $C_\Xi(t)$ either is defined uniquely or is not defined at all. Let $\mathbf{C}$, a *criterion*, be a random variable whose values are defined on the same scale as those of $C_\Xi(t)$, but they do not depend on t or on Ξ. (The correspondence in the notation, $\mathbf{C}$, c, $C_\Xi(t)$, emphasizes the common dimensionality of response processes and criteria.)

The time variable $\mathbf{T}(\Xi)$ will be said to be *generated* by the pair $C_\Xi(t)$ and $\mathbf{C}$ if the following statement is true for any value of Ξ:

$$\mathbf{T}(\Xi) = \inf\{t\colon C_\Xi(t) \geq \mathbf{C}\}. \tag{1}$$

Here $\{t\colon \ldots\}$ should be read as "all values of t such that ...". The time variable $\mathbf{T}(\Xi)$ is the first instance when $C_\Xi(t)$ reaches or exceeds the random criterion $\mathbf{C}$. Note that (1) defines $\mathbf{T}(\Xi)$ as a deterministic function of two arguments: Ξ, and a Ξ-independent random variable $\mathbf{C}$. Simple RT will be called *Grice-representable* if it can be generated according to (1), by an appropriate choice of a Ξ-independent random criterion, $\mathbf{C}$, and a Ξ-dependent deterministic process, $C_\Xi(t)$. The possibility mentioned in the Introduction (see Assumption C), that some experimental manipulations (components of Ξ) may affect the distribution of $\mathbf{C}$, will be discussed in the Conclusion.

1.3. McGill-Representability.

Let us say that $\mathbf{T}(\Xi)$ is *McGill-representable* if one can find a stochastic scalar process $\mathbf{R}_\Xi(t)$, depending on Ξ, and a constant criterion c_0, such that:

$$\mathbf{T}(\Xi) = \inf\{t\colon \mathbf{R}_\Xi(t) \geq c_0\}. \tag{2}$$

The following two lemmas establish the equivalence mentioned in the Introduction, between the Grice-representability and a particular version of the McGill-representability.

Lemma 1.3.1. Grice-representability, (1), of $\mathbf{T}(\Xi)$ is equivalent to its McGill-representability, (2), with $\mathbf{R}_\Xi(t) = C_\Xi(t) + \mathbf{N}(t)$: a deterministic response process, $C_\Xi(t)$, additively superimposed with a stationary Ξ-independent noise, $\mathbf{N}(t)$.

Lemma 1.3.2. Grice-representability, (1), of $\mathbf{T}(\Xi)$ is equivalent to its McGill-representability, (2), with $\mathbf{R}_\Xi(t) = C_\Xi(t)\mathbf{N}(t)$: a nonnegative deterministic response process, $C_\Xi(t)$, multiplicatively superimposed with a nonnegative stationary Ξ-independent noise, $\mathbf{N}(t)$.

It follows from the proofs of these lemmas that certain inherent characteristics of stochastic processes $\mathbf{R}_\Xi(t)$, such as their autocorrelation functions, cannot in principle be reconstructed from RT distributions modeled by these processes. The only stochastic property of $\mathbf{R}_\Xi(t)$ that matters is the distribution of the stationary noise $\mathbf{N}(t)$, by definition assumed to be the same for all time moments.

1.4. Generation of Time Variables.

Now consider more closely the mechanism by which time-dimensioned random variables, $\mathbf{T}(\Xi)$, are generated by deterministic response processes, $C_\Xi(t)$, coupled with randomly preset criteria, $\mathbf{C}$. The following concept greatly simplifies analysis. The *highest-reached-level (HRL) form* of a process $C_\Xi(t)$ is defined as

$$\bar{C}_\Xi(t) = \sup\{C_\Xi(t'): t' \leq t\}. \tag{3}$$

At any moment t, the HRL-form of $C_\Xi(t)$ shows the highest value reached by the process up to that moment. Obviously, $\bar{C}_\Xi(t)$ itself can be considered a deterministic response process. This process is nondecreasing, and it is its own HRL-form: $\bar{\bar{C}}_\Xi(t) = \bar{C}_\Xi(t)$. Any nondecreasing process $C_\Xi(t)$ coincides with its own HRL-form.

Let $\mathcal{T}_\Xi(t)$ and $\mathcal{C}(c)$ be the distribution functions for $\mathbf{T}(\Xi)$ and $\mathbf{C}$, respectively: $\mathcal{T}_\Xi(t) = \text{Prob}\{\mathbf{T}(\Xi) \leq t\}$, $\mathcal{C}(c) = \text{Prob}\{\mathbf{C} \leq c\}$. The following lemma, which immediately follows from a contemplation of definitions (1) and (3), specifies the role of $\bar{C}_\Xi(t)$ in generating $\mathbf{T}(\Xi)$.

Lemma 1.4. If $\mathbf{T}(\Xi)$ is generated according to (1) by $C_\Xi(t)$ and $\mathbf{C}$, then

$$\mathcal{T}_\Xi(t) = \mathcal{C}(\bar{C}_\Xi(t)). \tag{4}$$

It follows that if two response processes have identical HRL-forms, they will generate the same $\mathbf{T}(\Xi)$ when coupled with a given criterion $\mathbf{C}$. The reverse is not generally true: processes with different HRL-forms may generate the same $\mathbf{T}(\Xi)$. Indeed, consider $\mathcal{C}(c)$ that has a plateau within an interval (c_*, c^*), that is, $\text{Prob}\{c_* < \mathbf{C} \leq c^*\} = 0$. Consider a response process whose HRL-form assumes the values c_* and c^* at two different moments, $t_*(\Xi)$ and $t^*(\Xi)$, respectively. Obviously, the course of the HRL-form between the coordinates $(t_*(\Xi), c_*)$ and $(t^*(\Xi), c^*)$ is irrelevant for the generated variable $\mathbf{T}(\Xi)$: due to (4), $\text{Prob}\{t_*(\Xi) < \mathbf{T}(\Xi) \leq t^*(\Xi)\} = 0$ in all possible cases.

This ambiguity can be removed by only considering criteria with strictly increasing distribution functions: $\text{Prob}\{c_* < c \leq c^*\} > 0$ whenever $\inf\mathbf{C} < c_* < c^* < \sup\mathbf{C}$,

where infC and supC are defined as inf{c: $\mathscr{C}(c) > 0$} and sup{c: $\mathscr{C}(c) < 1$}, respectively. It immediately follows from Lemma 1.4 that

Corollary 1.4. Two response processes, $C_{1,\Xi}(t)$ and $C_{2,\Xi}(t)$, coupled with a common criterion **C** with a strictly increasing distribution function, generate one and the same $\mathbf{T}(\Xi)$ if and only if their HRL-forms coincide between infC and supC: $\bar{C}_{1,\Xi}(t) = \bar{C}_{2,\Xi}(t)$ whenever infC $< \bar{C}_{i,\Xi}(t) <$ supC, $i = 1$ or 2.

Note that the course of an HRL-form after it exceeds supC or before it reaches infC (if this may happen) is irrelevant for the generated $\mathbf{T}(\Xi)$.

The strict monotonicity restriction does not lead to a loss of generality with respect to the generation of time variables. It can easily be shown that, without affecting the generated $\mathbf{T}(\Xi)$, one can always "collapse" the intervals of $\mathscr{C}(c)$-constancy, $(c_*, c^*]$, into their lower boundaries, c_*, thus removing all gaps in the common axis for **C** and $C_{\Xi}(t)$. However, Theorem 1.5 below makes such a demonstration unnecessary.

1.5. *Grice-Representability of Simple RT.*

Theorem 1.5 shows that any $\mathbf{T}(\Xi)$ can be generated according to (1), whatever the set of factors Ξ, and whatever the distributions of **T** for each particular value of Ξ. Moreover, the criterion **C** can always be chosen so that its distribution function, $\mathscr{C}(c)$, is *strictly increasing and continuous* between infC and supC. (The rationale for the continuity restriction will be given later.) In all other respects the choice of **C** is completely arbitrary: its only function is to numerically calibrate the scale for the response processes $C_{\Xi}(t)$, that is, to induce a metric on an otherwise purely ordinal-scale structure (scale properties for response processes will be discussed in section 1.9).

Theorem 1.5. Let **C** be an arbitrary random variable that does not depend on Ξ or on t, and whose distribution function $\mathscr{C}(c)$ is strictly increasing and continuous between infC and supC. Then for any time-dimensioned random variable $\mathbf{T}(\Xi)$, with a distribution function $\mathscr{T}_{\Xi}(t)$, there exists a deterministic response process $C_{\Xi}(t)$ such that $\mathbf{T}(\Xi)$ is generated by $C_{\Xi}(t)$ and **C** according to (1). This process is defined uniquely up to its HRL-form at all t for which $0 < \mathscr{T}_{\Xi}(t) < 1$:

$$\bar{C}_{\Xi}(t) = \mathscr{C}^{-1}(\mathscr{T}_{\Xi}(t)). \tag{5}$$

As far as simple RT is concerned, this theorem justifies the statement made in the Introduction: however complex the dependence of RT on external parameters may be, the Grice modeling scheme is a descriptive language, not an empirically testable model. The theorem is constructive: once $\mathbf{T}(\Xi)$ is given and **C** is chosen, $C_{\Xi}(t)$ can be constructed uniquely up to its HRL-form between infC and supC. To remove any remaining ambiguity, let us complement (5) with the following agreement: $\bar{C}_{\Xi}(t)$ is set equal to supC (or infC) whenever $\mathscr{T}_{\Xi}(t) = 1$ (respectively, 0).

The reason for assuming that $\mathscr{C}(c)$ is strictly increasing has already been discussed (section 1.4): this restriction is convenient but not critical. The continuity restriction is, however, essential for the Grice-representability.

Lemma 1.5. For a discontinuous $\mathscr{C}(c)$ one can always find a time-dimensioned variable **T** that cannot be generated by any deterministic process $C(t)$ coupled with **C**.

This result is especially apparent for discrete criteria: obviously, such criteria cannot be used to generate continuous time-dimensioned variables. It can even be

shown that for any discrete criterion one can find a discrete time-dimensioned variable that cannot be generated. Due to Theorem 1.5, however, any (discrete, continuous, or mixed) time-dimensioned variable can be generated by means of an arbitrarily chosen continuous criterion.

1.6. Choice of Criteria.

In the rest of this paper the term "criterion" will always include the two restrictions imposed in the formulation of Theorem 1.5: $\mathscr{C}(c)$ will always be considered strictly increasing and continuous between infC and supC. Since in all other respects $\mathscr{C}(c)$ can be chosen arbitrarily, one can subject it to further restrictions without restricting the class of Grice-representable time variables. Thus one can follow Grice (1968, 1972) in assuming that $\mathscr{C}(c)$ is a normal distribution, or (to exclude negative values) a truncated normal distribution. An invocation of the central limit theorem to substantiate this choice would be meaningless, because the common axis for C and $C_{\Xi}(t)$ is not calibrated unless $\mathscr{C}(c)$ has been specified. Any strictly increasing continuous $\mathscr{C}(c)$ can be made normal by a monotonic transformation, and vice versa. The same argument applies, of course, to other limit distributions, such as Weibull. Mathematical convenience is the only basis for a particular choice of $\mathscr{C}(c)$. It may be convenient, for example, to consider only nonnegative criteria, and thereby only nonnegative HRL-forms of response processes. Furthermore, one can always replace $C_{\Xi}(t)$ with its HRL-form, $\tilde{C}_{\Xi}(t)$, thus leaving only nonnegative and nondecreasing functions, as in Grice's actual modeling. Due to this possibility, the HRL-forms $\tilde{C}_{\Xi}(t)$ in the following text will sometimes be referred to as response processes. (Note that dealing with $\tilde{C}_{\Xi}(t)$ directly is not necessarily convenient: for example, to preserve smoothness of the response process derivatives, a nonmonotonic (say, $\cap$-shaped) $C_{\Xi}(t)$ might be preferable to $\tilde{C}_{\Xi}(t)$ in some applications.)

Considering "natural" or "canonical" choices for $\mathscr{C}(c)$, given $\mathbf{C} \geq 0$, two distributions seem to stand out. A standard uniform distribution is convenient because $\mathscr{C}^{-1}(p) = p$ in this case, and according to (5), the response processes $\tilde{C}_{\Xi}(t)$ then coincide with the generated distribution functions $\mathcal{T}_{\Xi}(t)$ (except, perhaps, for arbitrarily assigned units of measurement, because $\mathcal{T}$ is, by definition, dimensionless). Another convenient choice for $\mathscr{C}(c)$ is a unit exponential distribution, because the processes $\tilde{C}_{\Xi}(t)$ then numerically coincide with the log-survival functions, $-\log(1 - \mathcal{T}_{\Xi}(t))$. The conditional probability $\mathcal{T}_{\Xi}(t \mid \mathbf{T} > t_0)$ is determined in this case simply by the difference $\tilde{C}_{\Xi}(t) - \tilde{C}_{\Xi}(t_0)$:

$$\tilde{C}_{\Xi}(t) - \tilde{C}_{\Xi}(t_0) = -\log(1 - \mathcal{T}_{\Xi}(t|\mathbf{T} > t_0)).$$

This is a desirable property, for instance when considering distributions of "true" responses under the condition that no false alarms have occurred (discussed in section 2.7).

Strictly speaking, $\mathscr{C}(c)$ calibrates the response process scale only between infC and supC. Beyond these limits (if finite) the scale can be continued arbitrarily. It was agreed in section 1.5, however, that $\text{infC} \leq \tilde{C}_{\Xi}(t) \leq \text{supC}$, which means that the response process scale can always be considered truncated at these limits.

1.7. An Example.

Let $\Xi = (F, A)$, where A is intensity of a target signal and F is the foreperiod, the interval between warning and target signals. As usual, the onset of the target signal is taken to be the time axis origin. Let the theoretical distribution function for RT in this situation be as follows.

$\mathcal{T}_{F,A}(t) = 0$ if $t \leq -F$ (no response before the warning signal);

$\mathcal{T}_{F,A}(t) = 1 - \exp\{-k(t + F)\}$ if $-F < t \leq r$ (A-independent exponential distribution between the warning signal and some nonnegative constant r, an "irreducible minimum");

$\mathcal{T}_{F,A}(t) = 1 - \exp\{-k(r + F) - k(t - r)^{1+\phi(A)}\}$ if $t > r$, where $\phi(A)$ is some positive monotonic transformation of intensity into a dimensionless number, $\phi(0) = 0$.

Let us choose a unit-exponentially distributed **C**: $\mathcal{C}(c) = 1 - \exp(-c)$; $\mathcal{C}^{-1}(p) = -\log(1 - p)$. Then (5) yields the following expressions for $\bar{C}_{F,A}(t)$:

$$\bar{C}_{F,A}(t) = \begin{cases} 0 & \text{for } t \leq -F; \\ k(t + F) & \text{for } -F < t \leq r; \\ k(r + F) + k(t - r)^{1+\phi(A)} & \text{for } t > r. \end{cases}$$

After $t = -F$ this function is strictly increasing for any (F, A), so $C_{F,A}(t) = \bar{C}_{F,A}(t)$ is the only deterministic process that generates $\mathcal{T}_{F,A}(t)$.

1.8. Causal Consistency.

The set of external factors Ξ is generally a parametrization of a (multidimensional) *input process*: at least some components of Ξ may themselves be functions of time, or represent global descriptors of such functions. In the example above, intensity A is a global descriptor of a certain intensity function, for example, a step-function $A(t)$ defined as follows:

$$A(t) = \begin{cases} 0 & \text{for } t < 0; \\ A & \text{for } t \geq 0. \end{cases}$$

The foreperiod F in this example is a global descriptor for another function: for instance, an indicator function assuming a value of 1 if a warning signal has been presented, and 0 before that:

$$W(t) = \begin{cases} 0 & \text{for } t < -F; \\ 1 & \text{for } t \geq -F. \end{cases}$$

Therefore, $\Xi = (F, A)$ represents in this example the input function $[W(t), A(t)]$. Since this cannot cause confusion, the same symbol, Ξ, will be used for both the input function and its parametrization. In the example:

$$\Xi(t) = [W(t), A(t)].$$

The causal consistency of a time variable $\mathbf{T}(\Xi) = \mathbf{T}(\Xi(t))$ means simply that the distribution of **T** up to a moment t cannot depend on values of the input function after this moment:

$$\mathcal{T}_{\Xi}(t) = \mathcal{T}(\{\Xi(u): u \leq t\}). \tag{6}$$

Note that $\{\Xi(u): u \leq t\}$ should be understood as an ordered set (indexed by u), where $\Xi(t)$ indicates not only the value of the input function but also the time moment, t, when this value occurs.

In the generation of $\mathbf{T}(\Xi)$, the criterion **C** can be chosen arbitrarily, but the response process $C_{\Xi}(t)$ has an HRL-form $\bar{C}_{\Xi}(t)$ uniquely related to the generated $\mathbf{T}(\Xi)$.

To be physically realizable, the response process should itself be causally consistent with respect to the input functions $\Xi(t)$:

$$C_{\Xi}(t) = C(\{\Xi(u): u \leq t\}). \tag{7}$$

The following lemma, whose proof is trivially derived from (5), shows that (6) and (7) are "almost" equivalent, in the sense that $\bar{C}_{\Xi}(t)$ cannot be anticipatory, and thereby (1) always describes a physically realizable generation process.

Lemma 1.8. For any choice of the criterion **C**, the HRL-form of the response process $C_{\Xi}(t)$ is causally consistent, (7), if and only if the generated time variable $\mathbf{T}(\Xi)$ is causally consistent, (6).

The causality, (7), can only be violated within intervals where the HRL-form is constant. If $\bar{C}_{\Xi}(t) = c_0$ between t_1 and t_2, then $C_{\Xi}(t)$ can have arbitrary values within this interval, provided they do not exceed c_0. Therefore, as a formal game, one can make the process $C_{\Xi}(t)$ anticipatory between t_1 and t_2. This will have no effect on $\mathbf{T}(\Xi)$, because $\text{Prob}\{t_1 < \mathbf{T}(\Xi) \leq t_2\} = 0$. Moreover, Lemma 1.8 guarantees that it can never be necessary to consider anticipatory processes, even within the gaps in RT distributions: there is at least one response process, namely $\bar{C}_{\Xi}(t)$, which is causally consistent throughout the entire domain of $\mathbf{T}(\Xi)$.

In the example of section 1.7, $C_{F,A}(t) = \bar{C}_{F,A}(t)$ starts at the moment $-F$ and develops as a linear function of time, $k(t + F)$, until the onset of the target stimulus (at $t = 0$), after which it changes its course depending on the value of signal intensity A. In agreement with Lemma 1.8, the course of $\bar{C}_{F,A}(t)$ within the foreperiod, $-F < t \leq 0$, only depends on the difference $t - (-F)$, not on the value of $t = t - 0$ itself (nor does it depend on A). If this were not the case, $C_{F,A}(t)$ would anticipate the onset (or intensity) of the target stimulus.

1.9. Averaging Of Response Processes.

In section 1.1 Ξ has been introduced as a set of "all" external factors upon which theoretical RT distributions are known to depend. This does not imply that there exists an "exhaustive" set of factors, such that RT distributions cannot be properly identified unless all of these factors are known. Any set of factors Ξ defines a certain family of RT distributions, $\mathbf{T}(\Xi)$, and vice versa: a family of random variables $\mathbf{T}(\Xi)$ is not defined unless it has been related to (indexed by) some set of factors (indices), Ξ. By reducing this set to one of its proper subsets, $\Xi^* \subset \Xi$, one simply defines a new family of random variables, $\mathbf{T}(\Xi^*)$. This might be viewed as a serious complication for the Grice modeling scheme, for the following reasons.

Consider a time variable $\mathbf{T}(\Xi)$ generated by a response process $C_{\Xi}(t)$ coupled with a criterion **C**. Consider a factor $B \in \Xi$, and denote Ξ without B by Ξ^*. (B can also be viewed as a proper subset of Ξ; in the present context, any such subset can be redefined as an element of Ξ, and Ξ itself can be redefined as a two-element set consisting of B and Ξ^*.) Let a certain distribution function $\mathcal{B}(b)$ be imposed on B, so that it can be considered a random variable **B**, independent of Ξ^*. Then for any given value of Ξ^*, the conditional distributions of $\mathbf{T}(\Xi^*|\mathbf{B} = b)$ will generally depend on b, and so will $C_{\Xi^*,b}(t)$. Different values b of **B** will correspond to different deterministic response processes $C_{\Xi^*,b}(t)$, from which it follows that for any fixed Ξ^*, $C_{\Xi^*,\mathbf{B}}(t)$ is a *stochastic* process whose realizations vary according to the distribution function $\mathcal{B}(b)$.

On the other hand, one can always choose to ignore **B** altogether, and consider the marginal random variable $\mathbf{T}(\Xi^*)$ whose distribution function is

$$\mathcal{T}_{\Xi^*}(t) = \int_b \mathcal{T}_{\Xi^*}(t|\mathbf{B} = b)\ \mathrm{d}\mathcal{B}(b). \tag{8}$$

The integration here should be understood in the Lebesgue-Stieltjes sense. Obviously, this distribution function is uniquely determined by Ξ^*. Then, according to Theorem 1.5, one can always find a *deterministic* response process, $C_{\Xi^*}(t)$, such that being coupled with the same criterion **C**, it would generate $\mathbf{T}(\Xi^*)$:

$$\tilde{C}_{\Xi^*}(t) = \mathcal{C}^{-1}(\mathcal{T}_{\Xi^*}(t)).$$

In the example of section 1.7, let the foreperiod be a random variable **F** exponentially distributed: $\mathcal{F}(f) = 1 - \exp(-mf)$. Isolating F from $\Xi = (F, A)$, and considering RT a function of $\Xi^* = A$ conditioned on values of **F**, one obviously gets different distribution functions $\mathcal{T}_A(t|\mathbf{F} = f)$ for a given value of A. Correspondingly, the response process $C_{\mathbf{F},A}(t)$ is a stochastic process whose realizations vary in accordance with the exponential distribution function $\mathcal{F}(f)$. Suppose, however, that one chooses to ignore foreperiod altogether, and considers RT a function of amplitude A alone. The marginal distribution of $\mathbf{T}(\Xi^*) = \mathbf{T}(A)$ can easily be derived from section 1.7:

$$\mathcal{T}_A(t) = \begin{cases} \dfrac{k \exp(mt)}{m+k} & \text{for } t \leq 0; \\ 1 - \dfrac{m \exp(-kt)}{m+k} & \text{for } 0 < t \leq r; \\ 1 - \dfrac{m \exp\{-kr - k(t-r)^{1+\phi(A)}}{m+k} & \text{for } r < t. \end{cases}$$

Using the same criterion **C**, with a unit exponential distribution, the random variable $\mathbf{T}(A)$ can be generated by the following deterministic response process:

$$C_A(t) = \begin{cases} -\log\left[1 - \dfrac{k\exp(mt)}{m+k}\right] & \text{for } t \leq 0; \\ kt - \log\left[\dfrac{m}{m+k}\right] & \text{for } 0 < t \leq r; \\ kr + k(t-r)^{1+\phi(A)} - \log\left[\dfrac{m}{m+k}\right] & \text{for } r < t. \end{cases}$$

This situation might be viewed as an internal inconsistency in the Grice modeling scheme. On the one hand, a given value of a set of factors Ξ^* (returning to the general case) "evokes" a single deterministic response process $C_{\Xi^*}(t)$. On the other hand, Ξ^* "evokes" different response processes $C_{\Xi}(t) = C_{\Xi^*,\mathbf{B}}(t)$ depending on the (random) value of an additional factor, **B**, that has not been previously taken into account. The two seemingly contradictory situations take place within the same response process scale, defined by the common criterion **C**.

A formal answer is, of course, that there is no contradiction because the two sets of factors, Ξ^* and $\Xi = \{\Xi^*, B\}$, defined two different time-dimensioned (families of) random variables, $\mathbf{T}(\Xi^*)$ and $\mathbf{T}(\Xi)$, respectively. This is not, however, an intellectually satisfactory answer. $\mathbf{T}(\Xi^*)$ is not simply different from $\mathbf{T}(\Xi)$: intuitively, $\mathbf{T}(\Xi^*)$ is a

"cruder version" of $\mathbf{T}(\Xi)$, a result of an averaging of the latter. Formally, the distribution function $\mathcal{T}_{\Xi^*}(t)$ in (8) is indeed the mean value of the distribution functions $\mathcal{T}_{\Xi^*}(t|\mathbf{B} = b)$ with respect to the probabilistic measure induced by $\mathcal{B}(b)$. One might expect, therefore, that the response process $C_{\Xi^*}(t)$ should also be a "cruder version", an average, of the response processes $C_{\Xi}(t) = C_{\Xi^*,\mathbf{B}}(t)$, computed across different values of b.

To see that this is indeed the case, one needs the concept of the Q-mean for a strictly increasing continuous function Q. For any function $X(\mathbf{B})$ of a random variable $\mathbf{B}$ with a distribution function $\mathcal{B}(b)$, the Q-mean is defined as a value $\overline{X}$ such that:

$$Q(\overline{X}) = \int_b Q(X(b))\ \mathrm{d}\mathcal{B}(b).$$

For a linear Q the Q-mean is the conventional expected value, for $Q(x) = \log(x)$ or x^{-1}, it is the geometric (respectively, harmonic) mean, etcetera. If $Q(x)$ is a probability distribution function, the Q-mean can also be called the quantile-rank mean, because then the quantile rank of $\overline{X}$ within the distribution defined by $Q(x)$ is the mean of the quantile ranks of $X(b)$-values within the same distribution. Combining this definition with (8) and (5), one comes to the following simple statement.

Lemma 1.9.1. Let $\mathbf{B}$ be a random variable whose distribution function $\mathcal{B}(b)$ does not depend on a set of factors Ξ^*. Let a time variable $\mathbf{T}(\Xi) = \mathbf{T}(\Xi^*, B)$ be generated by a response process $C_{\Xi}(t)$ coupled with a criterion $\mathbf{C}$, whose distribution function is $\mathcal{C}(c)$. Then a response process $C_{\Xi^*}(t)$ generates the marginal time variable $\mathbf{T}(\Xi^*)$ when coupled with the same criterion $\mathbf{C}$, if and only if $\tilde{C}_{\Xi^*}(t)$ is the $\mathcal{C}$-mean of $\tilde{C}_{\Xi}(t)$ computed across $\mathbf{B}$:

$$\mathcal{C}(\tilde{C}_{\Xi^*}(t)) = \int_b \mathcal{C}(\tilde{C}_{\Xi}(t))\ \mathrm{d}\mathcal{B}(b) \tag{9}$$

(in the Lebesgue-Stieltjes sense).

In other words, at any moment t, the quantile rank of $\tilde{C}_{\Xi^*}(t)$ within the criterion distribution is the mean (computed across all values b of $\mathbf{B}$) of the quantile ranks of processes $\tilde{C}_{\Xi^*,b}(t)$ within the same distribution. The next statement shows that $\mathcal{C}$-averaging can also be applied to the processes $C_{\Xi}(t)$ directly, rather than to their HRL-forms.

Lemma 1.9.2. Under the assumptions of Lemma 1.9.1, if a response process $C_{\Xi^*}(t)$ is the $\mathcal{C}$-mean of $C_{\Xi}(t)$ computed across $\mathbf{B}$,

$$\mathcal{C}(C_{\Xi^*}(t)) = \int_b \mathcal{C}(C_{\Xi}(t))\ \mathrm{d}\mathcal{B}(b), \tag{10}$$

then $C_{\Xi^*}(t)$ generates the marginal time variable $\mathbf{T}(\Xi^*)$ when coupled with the same criterion $\mathbf{C}$.

In a summary, one can legitimately focus on only a small subset of "all" factors that influence RT distributions. RT dependence on this subset can still be modeled by deterministic response processes. One only has to realize that any such process is an average, the quantile-rank mean, of a set of different deterministic processes modeling

RT distributions related to a larger subset of factors. This will obviously be true for any given set of factors, however comprehensive (unless they predict RTs deterministically). For example, one can always add stimulation factors used in preceding trials to the list of actually used ones, in which case response processes will exhibit sequential interdependence; but one can legitimately average RT distributions across all trials with a given value of stimulation factors, thereby averaging the corresponding response processes, and getting rid of sequential effects.

The concept of $\mathscr{C}$-averaging is consistent with the ordinal structure of the response process scale. In fact, it is the only form of averaging that can be legitimately applied to ordinal scales, because $\mathscr{C}$-averaging is done in terms of quantile ranks of the scale values being averaged, rather than the numerical values themselves. This concept will now be considered more closely. As discussed in section 1.6, for any time variable $\mathbf{T}(\Xi)$ the generating criterion $\mathbf{C}$ can be chosen arbitrarily, provided $\mathscr{C}(c)$ is strictly increasing and continuous. This means that the class of admissible transformations of the response process scale coincides with that of all strictly increasing and continuous transformations. Having applied such a transformation, say R, to the response process scale, one redefines ("recalibrates") the response processes,

$$\text{new } C_{\Xi}(t) = R[\text{old } C_{\Xi}(t)], \tag{11}$$

and changes the criterion distribution,

$$\text{new } \mathscr{C}(\text{new } c) = \text{old } \mathscr{C}(\text{old } c) = \text{old } \mathscr{C}[R^{-1}(\text{new } c)]. \tag{12}$$

It follows that the new $\mathscr{C}$-mean equals [old $\mathscr{C}$ R^{-1}]-mean, with respect to any random variable $\mathbf{B}$. Combining (11) and (12) one gets

$$\text{new } \mathscr{C}\text{-mean of new } C_{\Xi}(t) = \text{old } \mathscr{C}\text{-mean of old } C_{\Xi}(t), \tag{13}$$

which means that both (9) and (10) are invariant under all admissible transformations of the response process scale.

2. Grice-Representability of Label-And-Time Variables

Here, the Grice-representability of N-alternative choice RT will be defined and analyzed, including generalized disjunctive RT (Donders's reaction type c, when a subject is allowed to give no response in some trials). It will also be considered how the Grice modeling scheme applies to hypothetical decompositions of RT (mixture models), and how it deals with the dichotomy of target-initiated versus target-unrelated responses.

2.1. Label-And-Time Variables.

In the N-alternative choice RT paradigm, the formal object of modeling is a family of two-component random variables, $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$, where: $\mathbf{I}$ is the response identity label ("which of the N responses is chosen"), a random variable with values $i = 1, \ldots, N$; $\mathbf{T}$ is a time-dimensioned random variable representing RT; and Ξ are external factors indexing the theoretical distributions of $[\mathbf{I}, \mathbf{T}]$. By analogy with $\mathbf{T}(\Xi)$ in the preceding sections, the family of random vectors $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ will be referred to as a (single) *label-and-time* random variable, stochastically depending on Ξ. In sections 2.7, 2.8, and the Conclusion, the meaning of the label-and-time variable will be generalized to incorporate hypothetical (unobservable) components of RT.

The causal consistency restriction for label-and-time variables is defined in an obvious way. Let $\mathcal{T}_{\Xi}(i, t)$ denote the (joint) distribution function for $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$:

$$\mathcal{T}_{\Xi}(i,\ t) = \text{Prob}\{\mathbf{I}(\Xi) = i \text{ and } \mathbf{T}(\Xi) \leq t\}.$$

$\mathcal{T}_{\Xi}(i,\ t)$ is defined here as a (cumulative) distribution function with respect to t, but a mass function with respect to i. Let the input function $\Xi(t)$ be defined as in section 1.8. Then $\mathcal{T}_{\Xi}(i,\ t)$ must be uniquely determined by the input function taken up to the moment t:

$$\mathcal{T}_{\Xi}(i,\ t) = \mathcal{T}_{i}(\{\Xi(u){:}u \leq t\}). \tag{14}$$

The causal consistency, in the form of (6), was the only restriction imposed on the time variables representing simple RT. In the case of the label-and-time variables, however, to avoid cumbersome technicalities, one has to impose additional constraints on $\mathcal{T}_{\Xi}(i,\ t)$. Namely, it will be assumed in the rest of this paper that: (a) for any Ξ and $i = 1, \ldots, N$, $\mathcal{T}_{\Xi}(i,\ t)$ is continuous; (b) for any Ξ, there exists a strictly increasing continuous function of time, $\psi_{\Xi}(t)$, such that for any $i = 1, \ldots, N$, $\mathrm{d}\mathcal{T}_{\Xi}(i,\ t)/\mathrm{d}\psi_{\Xi}(t)$ is continuous at all $\psi_{\Xi}(t)$, with the possible exception of a finite number of isolated points, $\psi_{\Xi}(t_1) < \cdots < \psi_{\Xi}(t_m)$. At these points the discontinuities in $\mathrm{d}\mathcal{T}_{\Xi}(i,\ t)/\mathrm{d}\psi_{\Xi}(t)$ may be either of the first kind (finite jumps) or the second kind (infinite limits at either or both sides of a discontinuity point). As far as observable RT distributions are concerned, these assumptions might appear unnecessarily general. It seems that experimental RTs can always be approximated by piecewise differentiable distribution functions, in which case one could put $\psi_{\Xi}(t) = t$ in the definition above, and allow $\mathrm{d}\mathcal{T}_{\Xi}(i,\ t)/\mathrm{d}t$ to have discontinuities of the first kind only. It is desirable, however, to keep the definition in the less restrictive form, due to the potential applicability of the analysis to hypothetical RT components, whose distributions are not observable.

2.2. Incomplete time variables.

In the classical disjunctive RT paradigm (Donders's reaction type c) the subject is instructed to give a designated response if a certain statement about stimulation is true, and to withhold a response if the statement is false. Giving no response in certain trials can formally be considered a special response, for which RT is not defined, or is unknown. Disjunctive RT can, therefore, be represented by label-and-time variables $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ with $i = 1, 2$: $\mathbf{T}(\Xi)$ is defined when $\mathbf{I}(\Xi) = 1$ (a designated response), but is undefined or unknown when $\mathbf{I}(\Xi) = 2$ (response is withheld). The two interpretations of $\mathbf{T}(\Xi)$ in the case when $\mathbf{I}(\Xi)$ is no-response, "$\mathbf{T}(\Xi)$ is undefined" and "$\mathbf{T}(\Xi)$ is unknown", correspond to two different approaches to modeling (and possibly, to two different decision strategies adopted by observers, as discussed in section 2.5). Here, only the case of undefined $\mathbf{T}(\Xi)$ will be analyzed. Intuitively, this means that no-response is a result of waiting indefinitely long (for a perceptual signal to initiate a response), rather than a deliberate decision to give no response.

Consider, for example, detection of weak signals. Let signal intensity A be the only factor related to RT, $\mathbf{T}(\Xi) = \mathbf{T}(A)$. According to the Grice modeling scheme, one can choose an arbitrary criterion $\mathbf{C}$, and relate every possible value of A to a deterministic nondecreasing function $\bar{C}_A(t)$. It is an empirical fact that, for a broad class of signals, $\mathbf{T}(A)$ stochastically decreases as A increases, in the sense that the quantiles of $\mathbf{T}(A)$ are decreasing functions of A (Dzhafarov, 1992). (In fact, this rule only holds for the quantiles that are sufficiently far from the distribution margins, but this will be ignored here for simplicity.) This means that for any moment t, $\bar{C}_A(t)$ is a decreasing function of A: weaker signals evoke slower processes.

To be specific, let $\mathbf{C}$ be unit-exponentially distributed: in this case $\text{Prob}\{\mathbf{C} > c\} > 0$ for any c. Then if $\bar{C}_A(t)$ for a given A increases without bound, we have $\text{Prob}\{\bar{C}_A(t)$

$< \mathbf{C}$ for all $t\} = 0$, and the generated $\mathbf{T}(A)$ will be a *complete* random variable, meaning that $\mathcal{T}_A(t) \to 1$ as $t \to \infty$. This property is conventionally assumed to hold for any random variable by definition. Nothing in the Grice modeling scheme, however, precludes the possibility that for a sufficiently small value of A, the process $\tilde{C}_A(t)$ increases up to only a finite value, $\sup\tilde{C}_A(t) < \infty$. For example, a response process $C_A(t)$ evoked by a weak signal may increase up to a certain value and then decrease back to zero. If $\sup\tilde{C}_A(t) < \infty$, then

$$\text{Prob}\{\tilde{C}_A(t) < \mathbf{C} \text{ for all } t\} = \text{Prob}\{\mathbf{C} > \sup\tilde{C}_A(t)\} > 0,$$

and for some $p < 1$, $\mathcal{T}_A(t) \to p$ as $t \to \infty$. In such cases the generated time variable $\mathbf{T}(A)$ will be said to be *incomplete*. Feller (1968, p. 309) calls such random variables "defective". The term "incomplete" seems more suitable in the present context, as it is more suggestive of an underlying process, and less of an exceptional status.

Indeed, incomplete time variables arise in the Grice modeling scheme quite naturally. A time variable $\mathbf{T}(\Xi)$ generated according to (1) will be incomplete, if and only if $\sup\tilde{C}_\Xi(t) < \sup\mathbf{C}$:

$$\mathcal{T}_\Xi(t) \to p < 1 \text{ as } t \to \infty \quad \text{iff} \quad \text{Prob}\{\mathbf{C} > \sup\tilde{C}_\Xi(t)\} = 1 - p > 0.$$

In the case of simple RT, an additional mechanism (discussed in section 2.8) is needed to generate a response in every trial when the criterion is set at a value exceeding $\sup\tilde{C}_\Xi(t)$. Conversely, if no such mechanism is postulated, and simple RT is modeled as a complete time variable $\mathbf{T}(\Xi)$, then $\sup\mathbf{C}$ must not exceed $\sup\tilde{C}_\Xi(t)$ for any value of Ξ. This simple rule is important for practical construction of models. For instance, Grice's own model violates this rule, because it assumes a normally distributed criterion (i.e., $\sup\mathbf{C} = \infty$) coupled with negative-exponentially developing response processes,

$$\tilde{C}_\Xi(t) = A_\Xi - B_\Xi \exp\{-\lambda_\Xi(t - t_\Xi)\},$$

with $\sup\tilde{C}_\Xi(t) = A_\Xi < \infty$. In the case of choice RT, Grice considers Gompertz (double-exponential) functions coupled with normally distributed criteria, with the same unfortunate consequence. (Interestingly, Grice's first version of his model, 1968, is free from this problem, as it couples normal distributions with linear response processes.) Ashby (1982) pointed out the same "global incompleteness" defect in the cascade model (McClelland, 1979). This is not a coincidence, because McClelland's model is equivalent to a Grice-scheme model with

$$\tilde{C}_\Xi(t) = 1 - \sum_k \exp\{-\lambda_{k,\Xi}(t - t_\Xi)\}, \ \mathbf{C} = \frac{c - \mathbf{N}_{0,1}}{\mathbf{N}_{a,\sigma^2}},$$

where $\mathbf{N}_{0,1}$ and $\mathbf{N}_{a,\sigma^2}$ are independent normally distributed variables with the parameters indicated by the subscripts, whereas c is a constant (the fixed criterion in McClelland's presentation). One then has $\sup\tilde{C}_\Xi(t) = 1 < \sup\mathbf{C} = \infty$.

At the same time, the notion of an incomplete time variable can be directly utilized in modeling disjunctive RT.

Theorem 2.2. Let $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$, $i = 1, 2$, be a disjunctive label-and-time variable:

$$\text{Prob}\{\mathbf{I}(\Xi) = 1 \text{ and } \mathbf{T}(\Xi) \le t\} = \mathcal{T}_\Xi(1, t) \to p(\Xi) \text{ as } t \to \infty;$$

$$\text{Prob}\{\mathbf{I}(\Xi) = 2 \text{ and } \mathbf{T}(\Xi) \le t\} \text{ is undefined.}$$

Then for any criterion $\mathbf{C}$ with a distribution function $\mathscr{C}(c)$, there exists a response process $C_{\Xi}(t)$ generating $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ in the following sense:

$$[\mathbf{I}(\Xi), \mathbf{T}(\Xi)] = \begin{cases} [1, \inf\{t\colon C_{\Xi}(t) \geq \mathbf{C}\}] & \text{if } \mathbf{C} \leq \sup C_{\Xi}(t); \\ [2, \text{undefined}] & \text{if } \mathbf{C} > \sup C_{\Xi}(t). \end{cases}$$

This process is defined uniquely up to its HRL-form:

$$\bar{C}_{\Xi}(t) = \mathscr{C}^{-1}(\mathscr{T}_{\Xi}(1, t)). \tag{15}$$

This theorem may be considered a generalization of Theorem 1.5. A way to further generalize it to incorporate situations with "$N-1$ real responses plus 1 no-response" ($N > 2$) will be discussed in section 2.5.

2.3. Grice-Representability of Label-And-Time Variables.

Consider now the N-alternative choice RT paradigm: here $\mathbf{T}(\Xi)$ in a label-and-time variable $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ is defined for any $i = 1, \ldots, N$. The label-and-time variable will be said to be Grice-representable if there are N deterministic response processes $C_{i,\Xi}(t)$, and N *mutually independent* criteria $\mathbf{C}_i$, such that

$$[\mathbf{I}(\Xi) = i, \mathbf{T}(\Xi) = t]$$

$$\text{iff} \tag{16}$$

$$\min_k\{\inf\{t\colon \underline{C_{k,\Xi}(t)} \geq \mathbf{C}_k\}\} = \inf\{t\colon \underline{C_{i,\Xi}(t)} \geq \mathbf{C}_i\} = t.$$

In other words, the i-th response is given at moment t if and only if the i-th response process crosses its criterion for the first time at moment t, whereas all other response processes have been below their criterion levels up to this moment. This should be considered a termination rule for *all* N response processes competing for their respective criteria, rather than only for the i-th response process, the winner of the competition.

Another way to define the Grice-representability of $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ is to consider N time-dimensioned random variables $\mathbf{T}_i(\Xi)$, generated by N pairs $\{C_{i,\Xi}(t), \mathbf{C}_i\}$ according to (1), such that

$$[\mathbf{I}(\Xi) = i, \mathbf{T}(\Xi) = t] \quad \text{iff} \quad \min\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\} = \mathbf{T}_i(\Xi) = t. \tag{17}$$

Again, one should keep in mind that only one of these time variables, the minimum one, attains a numerical value within any given "working cycle" of the system (whose duration equals this value). Numerical values for the rest of the time variables should be viewed as potential, or "virtual", termination times.

That the criteria are mutually independent was originally proposed by Grice for reasons of mathematical simplicity. This does not elucidate the theoretical significance of this restriction. Nor can one know what consequences it would have for Grice modeling to allow for a stochastic interdependence between the criteria. It seems important to answer these questions to more clearly understand the role of the criteria in the Grice modeling scheme. Therefore, it is useful to introduce the notion of the *generalized* Grice-representability of the label-and-time variables: $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ will be said to be Grice-representable in the generalized sense if (16) holds for N criteria $\mathbf{C}_i$ with some joint distribution function

$$\mathscr{C}(c_1, \ldots, c_N) = \text{Prob}\{\mathbf{C}_i \leq c_i,\ i = 1, \ldots, N\}.$$

Theorem 2.3.1 below shows that such a representation is possible for any label-and-time variable. Moreover, as one might expect from the analysis of simple RT, where the only role of the criteria is to numerically calibrate ordinal-scale axes for response processes, the joint distribution function $\mathscr{C}(c_1, \ldots, C_N)$ can be chosen arbitrarily, provided it is sufficiently well-behaved. In particular, the criteria can always be chosen to be mutually independent, which means that any label-and-time variable is also Grice-representable in the original, restricted sense.

The "well-behavedness" of the joint distribution of the criteria is specified by the following condition which will be referred to as the *admissibility* of a set of N criteria, $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$. This term is chosen to emphasize that the criteria serve to define a numerical N-dimensional frame of reference for the response processes, and the restrictions imposed on the choice of the criteria specify the admissible transformations of this frame of reference (which is essentially an "N-dimensional ordinal-scale" structure). A set of N criteria, $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$, is admissible if

1. the marginal distribution functions (marginals, for short) $\mathscr{C}_i(c_i) = \text{Prob}\{\mathbf{C}_i \leq c_i\}$ are strictly increasing and continuous on their domains, $(\text{inf}\mathbf{C}_i, \text{sup}\mathbf{C}_i)$, $i = 1, \ldots, N$;
2. for all $i = 1, \ldots, N$, the first-order conditional probabilities (conditionals, for short),

$$\mathscr{C}_i^-(c_1, \ldots, c_N) = \text{Prob}\{\mathbf{C}_j > c_j \text{ for all } j \neq i | \mathbf{C}_i = c_i\},$$

are nonzero (i.e., positive) on their domain $(\text{inf}\mathbf{C}_1, \text{sup}\mathbf{C}_1) \times \ldots \times (\text{inf}\mathbf{C}_N, \text{sup}\mathbf{C}_N)$ and satisfy the local Lipschitz condition with respect to all marginals $\mathscr{C}_k(c_k)$, $k = 1, \ldots, N$. (The superscript in $\mathscr{C}_i^-$ is a reminder that the conditionals are defined in terms of "greater than", rather than "less than or equal".)

The meaning of the local Lipschitz condition with respect to the marginals is as follows. For any *closed* subregion of $(\text{inf}\mathbf{C}_1, \text{sup}\mathbf{C}_1) \times \ldots \times (\text{inf}\mathbf{C}_N, \text{sup}\mathbf{C}_N)$ one can find a positive constant L, such that

$$|\mathscr{C}_i^-(c_1, \ldots, c_N) - \mathscr{C}_i^-(c_1^*, \ldots, c_N^*)| \leq L \sum_1^N |\mathscr{C}_k(c_k) - \mathscr{C}_k(c_k^*)|,$$

for any i and any two points $(c_1, \ldots, c_N)$ and $(c_1^*, \ldots, c_N^*)$ within the subregion. The Lipschitz condition implies (is stronger than) the continuity of $\mathscr{C}_i^-(c_1, \ldots, c_N)$, and it is implied by (is weaker than) the piecewise differentiability of $\mathscr{C}_i^-(c_1, \ldots, c_N)$ with respect to the marginals. As a result, the admissibility constraint imposed on $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ is a rather weak smoothness requirement, combined with the requirement that the distribution does not contain gaps of a certain shape (all $\mathscr{C}_i^-$ are nonzero). In particular, any set $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ whose density function is defined and positive on its domain is admissible. (Additional clarifications of the admissibility and Lipschitz condition can be found in the proof of Theorem 2.3.1.)

Theorem 2.3.1. Any label-and-time random variable $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ is Grice-representable in the generalized sense with respect to any admissible set of criteria $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$. The corresponding response processes $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$ are determined uniquely up to their HRL-forms.

Strictly speaking, the theorem is not constructive, but the standard theory of nonlinear differential equations provides a variety of techniques that can be used to find approximations (piecewise or integral) for the sought processes $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$.

The following result is contained in the proof of Theorem 2.3.1, and deserves to be stated separately as it clarifies further the natural occurrence of the incomplete time variables in the Grice modeling scheme. Recall that in (17) the time variables $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$ are assumed to be generated according to (1) by admissible (generally interdependent) criteria $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ coupled componentwise with response processes $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$.

Corollary 2.3.1. For any Ξ, at least one of the time variables $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$ in (17) is complete, but as many as $N-1$ remaining time variables may be incomplete.

A special importance has been attached in the literature (in the context of parallel versus serial processing) to the fact that the competing termination times are not necessarily all complete (Luce, 1986; Townsend, 1976; Townsend & Ashby, 1983). The issue is indeed important when, for example, a set of parallel processes is being considered, each related to a separate element or aspect of stimulation, and each having its own termination time, irrespective of its ordinal position with respect to other termination times. In the framework of the Grice modeling scheme, however, the completeness issue does not have a special significance: incomplete time variables are physically realizable and meaningful in exactly the same sense as complete ones. Indeed, the criteria $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ are all complete: in every trial each $\mathbf{C}_i$ assumes a certain finite value. The deterministic response processes $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$ are physically realizable since they are causally consistent with their input functions (sections 1.8 and 2.4). Therefore, the generation of the (potential) termination times $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$ can always be simulated by a physical system in real time. It is simply the case that some processes never exceed certain levels, and consequently they cannot win the who-is-first race if their termination criteria are set above these levels. At the same time, since at least one of the variables $\mathbf{T}_i(\Xi)$ is complete, the actual processing time, $\min\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$, is necessarily a complete variable. Observe that this minimum equals the marginal time variable $\mathbf{T}(\Xi)$, the RT aggregated across all N responses: it must be complete if (and only if) the response may not be withheld. As a result, a "global" incompleteness problem of the kind discussed in Section 2.2 does not exist here: predicted $\text{Prob}\{\mathbf{T}(\Xi) < \infty\}$ cannot be less than 1.

Returning now to the restricted formulation of the Grice-representability, the following theorem is a straightforward specification of Theorem 2.3.1. It is based on the simple observation that any set $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ of stochastically independent criteria is admissible, provided the marginals $\mathscr{C}_i(c_i) = \text{Prob}\{\mathbf{C}_i \leq c_i\}$ are strictly increasing and continuous on their domains.

Theorem 2.3.2. Any label-and-time random variable $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ is Grice-representable in the restricted sense, with respect to any N mutually independent criteria $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ whose distribution functions are strictly increasing and continuous on their domains. The corresponding response processes $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$ are determined uniquely up to their HRL-forms. Namely, for $i = 1, \ldots, N$,

$$\widetilde{C}_{i,\Xi}(t) = \mathscr{C}_i^{-1}\left(1 - \exp\left\{-\int_{-\infty}^{t} \frac{d\mathcal{T}_{\Xi}(i, u)}{1 - \mathcal{T}_{\Xi}(u)}\right\}\right), \tag{18}$$

where $\mathcal{T}_{\Xi}(t)$ is the marginal distribution function of $\mathbf{T}(\Xi)$ in $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$.

Theorems 2.3.1 and 2.3.2 establish the fact mentioned in the Introduction and demonstrated earlier for simple RT and disjunctive RT: the criteria in the Grice modeling scheme serve to induce a metrical structure (calibration) on the otherwise ordinal scales for response process values. This is their sole function, and as any other calibration problem, it is not subject to empirical testing. A simple parallel can be drawn between the choice of the criteria and the choice of a scaling factor in the calibration of a ratio scale. The exclusion of nonadmissible sets of criteria, for example, can be related to the exclusion of negative or zero scaling factors in the case of ratio scale transformations.

Applying Lemma 1.3.1 (or 1.3.2) to the termination times $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$, one can see that the equivalence between the McGill and Grice modeling schemes trivially extends to label-and-time variables, yielding the following picture: N stochastic processes of the type $\mathbf{R}_{i,\Xi}(t) = C_{i,\Xi}(t) + \mathbf{N}_i(t)$ competing for fixed (e.g., unity-level) criteria. This is a variant of the simple accumulator models (LaBerge, 1962; Pike, 1973; Vickers, 1970), and Theorems 2.3.1. and 2.3.2 tell us that such a model can always be used to describe choice RT distributions. Moreover, the distributions of $\mathbf{N}_i(t)$ (the Ξ-independent stationary noise) can be chosen arbitrarily, provided they are admissible in the above sense. In particular, it is a matter of arbitrary choice whether $\mathbf{N}_i(t)$ are or are not mutually independent for different values of i.

Theorem 2.3.2 could also be proved without reference to the general result provided by Theorem 2.3.1. Instead, one could first show that for any $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$, there exist N unique time variables $\mathbf{T}_i(\Xi)$ satisfying (17), and then one could apply Theorem 1.5 to each of these time variables. The applicability of (17) to any label-and-time variable has been proved by Marley and Colonius (1992; see also Marley, 1992), except that the restrictions they impose on $\mathcal{T}_{\Xi}(i, t)$ are unnecessarily stringent. Essentially the same result, in a somewhat different context, was earlier proved by Townsend (1976). There seems to be no conceptually transparent way, however, to generalize the Townsend-Marley-Colonius theorem to a set of interdependent time variables $\mathbf{T}_i(\Xi)$ satisfying (17), unless their joint distribution is defined through that of the Ξ-independent criteria, as it was done in Theorem 2.3.1. Therefore, this alternative approach to Theorem 2.3.2 would be less elucidating with respect to the role played by the criteria in the Grice modelling scheme.

2.4. Physical Realizability And Grouping of Responses.

It has been established in sections 1.8 and 1.9 that response processes modeling time variables $\mathbf{T}(\Xi)$ exhibit desirable consistency properties with respect to causality and averaging across external factors. Both these results trivially generalize to label-and-time variables. It is unnecessary to formulate these generalizations as separate formal statements, because Lemmas 1.8, 1.9.1, and 1.9.2 can be applied to label-and-time variables almost verbatim. One only has to treat $\mathbf{C}$, $C_{\Xi}(t)$, and $\tilde{C}_{\Xi}(t)$ as N-dimensional vectors, whereas the references to Theorem 1.5 in the proofs should be replaced by references to differential equations (A6, Appendix) in the general case, or (18) in the case of mutually independent criteria. To demonstrate causal consistency, for example, one observes that solutions of differential equations (A6) at some moment t cannot be affected by the course of $\mathcal{T}_{\Xi}(i, t)$ after that moment. In relation to averaging across external factors, one observes that analogues of (8) and (9) apply to, respectively, right-hand sides and left-hand sides of (A6).

Here, we will consider yet another consistency property of the Grice modeling scheme for label-and-time variables, this time a property that has no analogues in the

situations involving no choice. As an example, consider a 3-alternative label-and-time variable $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$, $i = 1, 2, 3$. Suppose that one decides to group the responses with labels 2 and 3 into a single response category, labeled 2*. The remaining response, originally labeled 1, then can be renamed into 1*, so that the choice is now between 1* and 2*, rather than among 1, 2, and 3. Formally, the grouping of the two responses defines a new label-and-time variable, $[\mathbf{I}^*(\Xi), \mathbf{T}^*(\Xi)]$, $i = 1^*, 2^*$. Suppose further that the original variable, $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$, is Grice-represented by three response processes $\{\tilde{C}_{1,\Xi}(t), \tilde{C}_{2,\Xi}(t), \tilde{C}_{3,\Xi}(t)\}$ coupled componentwise with three *mutually independent* criteria $\{\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3\}$; analogously, let $[\mathbf{I}^*(\Xi), \mathbf{T}^*(\Xi)]$ be Grice-represented by $\{\tilde{C}_{1^*,\Xi}(t), \tilde{C}_{2^*,\Xi}(t)\}$ coupled with two mutually independent criteria $\{\mathbf{C}_1^*, \mathbf{C}_2^*\}$. One should expect in this situation that, since the criteria are mutually independent, the following should be true.

1. If $\mathbf{C}_1^* = \mathbf{C}_1$ then $\tilde{C}_{1^*,\Xi}(t) = \tilde{C}_{1,\Xi}(t)$. Intuitively, if the metric induced on the axis for $\tilde{C}_{1,\Xi}(t)$ by $\mathbf{C}_1$ does not depend on the calibration of the two remaining axes (which is the essence of the mutual independence of the criteria), then the numerical values of the process should remain the same irrespective of what one does with the other two axes.

2. $\tilde{C}_{2^*,\Xi}(t)$ is related to $\{\tilde{C}_{2,\Xi}(t), \tilde{C}_{3,\Xi}(t)\}$ by a composition rule that only depends on $\mathbf{C}_2^*$, $\mathbf{C}_2$, and $\mathbf{C}_3$. Indeed, the distribution function of $[\mathbf{I}^*(\Xi) = 2^*, \mathbf{T}^*(\Xi)]$ is simply a sum of those for $[\mathbf{I}(\Xi) = 2, \mathbf{T}(\Xi)]$ and $[\mathbf{I}(\Xi) = 3, \mathbf{T}(\Xi)]$. Intuitively, $\tilde{C}_{2^*,\Xi}(t)$ should also be a "sum" of some kind of the processes $\tilde{C}_{2,\Xi}(t)$ and $\tilde{C}_{3,\Xi}(t)$, with a "summation rule" that depends on the calibration of the three scales, but not on the summands themselves.

The following lemma shows that these statements are indeed true. This is a simple but important characteristic of the Grice modeling scheme for choice RT. If it did not hold, a physical simulation of the response processes $\tilde{C}_{i,\Xi}(t)$ would depend on what and how many different versions or variants of a given response are being distinguished by an observer. The functions $\tilde{C}_{i,\Xi}(t)$ then, though physically realizable in the sense of causal consistency, could hardly be considered "real" physical processes, rather than mere mathematical constructs.

Lemma 2.4. Let a label-and-time variable $[\mathbf{I}^*(\Xi), \mathbf{T}^*(\Xi)]$ be a "grouped version" of another label-and-time variable, $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$. This means that for every $i^* = 1, \ldots, G$, there is a nonempty subset L_{i^*} of the set of labels $i = 1, \ldots, N$, such that

$$[\mathbf{I}^*(\Xi) = i^*, \mathbf{T}^*(\Xi) = t] \text{ iff } [\mathbf{I}(\Xi) \in L_{i^*}, \mathbf{T}(\Xi) = t].$$

Let $\{\mathbf{C}_i\}$ and $\{\mathbf{C}_{i^*}\}$ be two sets of mutually independent criteria, generating $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ and $[\mathbf{I}^*(\Xi), \mathbf{T}^*(\Xi)]$, respectively, when coupled with the corresponding sets of response processes, $\{\tilde{C}_{i,\Xi}(t)\}$ and $\{\tilde{C}_{i^*,\Xi}(t)\}$, where $i = 1, \ldots, N$, $i^* = 1, \ldots, G$. Then $\tilde{C}_{i^*,\Xi}(t)$ is a function of $\{\tilde{C}_{i,\Xi}(t)\}_{i \in L_{i^*}}$ such that:

$$1 - \mathscr{C}_{i^*}(\tilde{C}_{i^*,\Xi}(t)) = \prod_{i \in L_{i^*}} [1 - \mathscr{C}_i(\tilde{C}_{i,\Xi}(t))]. \tag{19}$$

It can be seen that the composition rule indeed depends on the distribution functions $\{\mathscr{C}_i\}_{i \in L_{i^*}}$ and $\mathscr{C}_{i^*}$ only. This directly relates to Statement 2 above, but Statement 1 follows as well, as a particular case. It is easy to see that if L_{i^*} consists of one element only, say i, and if the functions $\mathscr{C}_i$ and $\mathscr{C}_{i^*}$ coincide, then (19) yields $\tilde{C}_{i^*,\Xi}(t) = \tilde{C}_{i,\Xi}(t)$.

In the general case of stochastically interdependent criteria, $\tilde{C}_{i^*,\Xi}(t)$ is a compos-

ite of $\{\bar{C}_{i,\Xi}(t)\}_{i\in L_i^*}$ which, again, depends only on the distribution functions for the criteria. This time, however, all N criteria $\{\mathbf{C}_i\}$ and all G criteria $\{\mathbf{C}_{i^*}\}$ have to be involved in each particular composition, and the formulations lose the conceptual transparency of Lemma 2.4. This is not a principal issue, however, because the composite response processes can always be constructed by the following algorithm: first, the N interdependent criteria $\{\mathbf{C}_i\}$ are replaced with a set of N independent criteria $\{\overline{\mathbf{C}}_i\}$, and the response processes are recalibrated accordingly; second, the responses are grouped into G groups and composite response processes are computed according to (19) with respect to some set of G independent criteria $\{\overline{\mathbf{C}}_{i^*}\}$; finally, $\{\overline{\mathbf{C}}_{i^*}\}$ are replaced with the G interdependent criteria $\{\mathbf{C}_{i^*}\}$, and the composites sought are obtained by the corresponding recalibration.

2.5. Generalized c-Reaction.

Analysis of the "$N-1$ real responses plus 1 no-response" paradigm, $N \geq 2$, involves no new concepts. The object of modeling here is a label-and-time variable $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ with a distribution function $\mathcal{T}_\Xi(i, t)$ defined for all t if $i = 1, \ldots, N-1$, and undefined or unknown for $i = N$. If $\mathcal{T}_\Xi(N, t)$ is considered undefined, one can view $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ as an incomplete label-and-time variable:

$$\sum_{k=1}^{N-1} \mathcal{T}_\Xi(k, t) \to p(\Xi) \text{ as } t \to \infty. \tag{20}$$

Theorems 2.3.1 and 2.3.2 are still applicable, except that the generated termination times $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_{N-1}(\Xi)\}$ in (17) will now all necessarily be incomplete. After having chosen $\{\mathbf{C}_1, \ldots, \mathbf{C}_{N-1}\}$ and having computed $\{\bar{C}_{1,\Xi}(t), \ldots, \bar{C}_{N-1,\Xi}(t)\}$, one gets a straightforward generalization of Theorem 2.2: $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ equals $[N, \text{undefined}]$ if $\mathbf{C}_k > \sup C_{k,\Xi}(t)$ for all $k = 1, \ldots, N-1$; and $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ is defined by (16), with $i = 1, \ldots, N-1$, if $\mathbf{C}_k \leq \sup C_{k,\Xi}(t)$ for at least one k. The response processes are defined uniquely up to their HRL-forms.

This approach seems plausible in situations when "no-response" is assumed to be a result of indefinite waiting, as discussed in section 2.2. To generalize the example given in that section, assume that A is the magnitude of a change in intensity that may occur in two directions, positive or negative: the observer is instructed to respond "1" if $A > 0$, "2" if $A < 0$, and give no response (formally, "3") if $A = 0$. If the foreperiod F in this experiment may assume arbitrarily large values (e.g., it is exponentially distributed), then a natural strategy for the observer would be to wait for a signal to occur: at no point in time would a decision be made that $A = 0$, and that therefore no response is required. Suppose, however, that the foreperiod is bounded from above (e.g., it is uniformly distributed between f_0 and f_1), and A may assume only three values: zero, a large positive and a large negative ones ("large" meaning "perfectly detectable"). Then a reasonable strategy would be to wait for a signal only within the interval $(-F + f_0; -F + f_1)$, and at $t = -F + f_1$ to decide that $A = 0$ and no response is to be given. This example shows that the distribution function $\mathcal{T}_\Xi(i, t)$ may in fact be defined at $i = 3$ (a distribution of no-response decision times). If this distribution were known, the analysis would be formally reduced to that of a 3-alternative choice paradigm.

In general, however, $\mathbf{T}(\Xi)$ for no-response is unknown even if assumed to be defined. A Grice-representation for $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ in this case can always be constructed, but it cannot be constructed uniquely. Assuming mutually independent criteria, and rewriting (18) as

$$\widetilde{C}_{i,\Xi}(t) = \mathscr{C}_i^{-1}\left(1 - \exp\left\{-\int_{-\infty}^{t} \frac{d\mathcal{T}_\Xi(i, u)}{1 - \sum_{k=1}^{N-1} \mathcal{T}_\Xi(k, u) - \mathcal{T}_\Xi(N, u)}\right\}\right), \tag{21}$$

one can see that by choosing different no-response distributions $\mathcal{T}_\Xi(N, t)$ one can construct different Grice-representations for one and the same observable part of the distribution $\mathcal{T}_\Xi(i, t)$, $i = 1, \ldots, N-1$. The only constraint imposed on a choice of $\mathcal{T}_\Xi(N, t)$ is that it cannot exceed $1 - p(\Xi)$ as defined in (20).

By definition, $\mathcal{T}_\Xi(N, t) \to 1 - p(\Xi)$ if the label-and-time variable $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ is complete. However, a valid Grice-representation can be constructed even if $\mathcal{T}_\Xi(N, t)$ tends to a nonnegative value $q(\Xi) < 1 - p(\Xi)$. A substantive interpretation of this possibility is that the subject's strategy can be mixed: no-response can be a deliberate choice in some trials, and a result of indefinite waiting in other trials, with probability $1 - p(\Xi) - q(\Xi)$. To see that this interpretation is internally consistent, let the limit value for $\mathcal{T}_\Xi(N, t)$, $q(\Xi)$, decrease gradually from $1 - p(\Xi)$ to 0. The corresponding interpretation of no-response then changes from a deliberate choice in all cases, $q(\Xi) = 1 - p(\Xi)$, through mixed strategies, $q(\Xi) < 1 - p(\Xi)$, to indefinite waiting in all cases, $q(\Xi) = 0$. The latter has been considered in the beginning of this section: $\mathcal{T}_\Xi(N, t)$ is undefined, and for a given choice of $N-1$ criteria, $\{\mathbf{C}_1, \ldots, \mathbf{C}_{N-1}\}$, there is one and only one set of $N-1$ response processes, $\{\widetilde{C}_{1,\Xi}(t), \ldots, \widetilde{C}_{N-1,\Xi}(t)\}$, Grice-representing $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$. On the other hand, this situation corresponds to $\mathcal{T}_\Xi(N, t) = 0$ in (21). Substituting zero for $\mathcal{T}_\Xi(N, t)$ in (21), it is easy to verify that for any choice of $\mathbf{C}_N$, added to the same set $\{\mathbf{C}_1, \ldots, \mathbf{C}_{N-1}\}$, the variable $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ will be Grice-represented by the same $N-1$ response processes plus a process $\widetilde{C}_{N,\Xi}(t) = \inf\mathbf{C}_N$; $\text{Prob}\{\widetilde{C}_{N,\Xi}(t) \geq \mathbf{C}_N\} = 0$.

2.6. Independent termination times—interdependent processes.

A simple but important consequence of the Grice-representability theory for choice RTs is that whether the termination times for response processes are stochastically independent or interdependent has nothing to do with the lack or presence of interactions between the processes themselves. The stochastic in/ter/dependence of termination times $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$, as defined in (17), is entirely determined by the arbitrary choice of the criteria: the termination times are mutually independent if and only if the criteria are mutually independent. Since there are no empirical grounds for preferring one set of criteria over another (see section 2.3, the paragraph following Theorem 2.3.2), no empirically relevant meaning can be assigned to stochastic in/ter/dependence of $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$. The response processes themselves, on the other hand, do interact in the following simple sense. Consider two label-and-time variables, $[\mathbf{I}(\Xi_1), \mathbf{T}(\Xi_1)]$ and $[\mathbf{I}(\Xi_2), \mathbf{T}(\Xi_2)]$, corresponding to two different values of Ξ within the same choice family (i.e., the choice is between the same N alternatives in both cases). Suppose that the joint distributions $\mathcal{T}_{\Xi_1}$ $(i = 1, t)$ and $\mathcal{T}_{\Xi_2}$ $(i = 1, t)$ are different, but for all $i \neq 1$, the joint distributions $\mathcal{T}_{\Xi_1}(i, t)$ and $\mathcal{T}_{\Xi_2}(i, t)$ are identical. Let $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ be a set of mutually independent criteria, and consider the two corresponding sets of response processes, $\{C_{1,\Xi_1}(t), \ldots, C_{N,\Xi_1}(t)\}$ and $\{C_{1,\Xi_2}(t), \ldots, C_{N,\Xi_2}(t)\}$, computed according to (18). An inspection of (18) shows that under the assumptions made, $\widetilde{C}_{1,\Xi_1}(t)$ is different from $\widetilde{C}_{1,\Xi_2}(t)$, as one might readily expect, but for all $i \neq 1$, $\widetilde{C}_{i,\Xi_1}(t)$ are also different from $\widetilde{C}_{i,\Xi_2}(t)$. Indeed, for all $i \neq 1$, the subintegral function in (18),

$$\frac{d\mathcal{T}_{\Xi}(i, t)}{1 - \mathcal{T}_{\Xi}(t)} = \frac{d\mathcal{T}_{\Xi}(i, t)}{1 - \mathcal{T}_{\Xi}(1, t) - \sum_{2}^{N} \mathcal{T}_{\Xi}(k, t)},$$

is different for Ξ_1 and Ξ_2, because $\mathcal{T}_{\Xi_1}(1, t) \neq \mathcal{T}_{\Xi_2}(1, t)$, whereas all other components of the expression are the same. ("Paradoxically", the proof that $C_{1,\Xi}(t)$ must change in this situation as well is somewhat more technical.) Different response processes, therefore, change in an interdependent fashion. Put in more "physicalistic" terms, the (deterministic) response processes (deterministically) interact with each other. At the same time, the (stochastic) durations of the processes, $\{\mathbf{T}_1(\Xi), \ldots, \mathbf{T}_N(\Xi)\}$, are mutually independent for both Ξ_1 and Ξ_2, reflecting the fact that $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ are chosen to be stochastically independent. It is useful to emphasize here that the stochastic independence exists between potential, or virtual, termination times, rather than actual durations. The latter are, in a sense, perfectly synchronized in the Grice modelling scheme: all processes terminate as soon as one of them reaches its criterion.

It was shown in sections 1.8 and 2.4 that response processes are physically realizable because they are causally consistent with their input functions. As a simple corollary, it follows that the dynamic interaction between competing processes is also causally consistent. The interaction can be, therefore, viewed as an objective characteristic of a physical system. If the example just considered is modified by assuming that $\mathcal{T}_{\Xi_1}(i = 1, t)$ and $\mathcal{T}_{\Xi_2}(i = 1, t)$ differ after some moment t^*, but not before, then it is clear from (18) that none of the processes $C_{i,\Xi_1}(t)$ differs from $C_{i,\Xi_2}(t)$ before moment t^*, whether $i = 1$ or $i \neq 1$. In the general context of Theorem 2.3.1, the same conclusion is derived from observing that solutions of differential equations (A6) at t^* cannot be affected by the values of $\mathcal{T}_{\Xi}(i, t)$ after that moment.

The notion of a deterministic dynamic interaction might help to further reconcile one's intuition with the natural occurrence of incomplete time variables in the Grice modeling scheme (see Corollary 2.3.1). Faster developing processes may be thought of as inhibiting slowly developing ones, thereby preventing them from exceeding a certain level. As a result, an inhibited process can never win the who-is-first race if the corresponding criterion is set sufficiently high.

2.7. *False Alarms And Mixtures.*

In this and the next section, the Grice-representability of label-and-time variables will be illustrated on RT mixture models, with an emphasis on the notion of target-unrelated responses. Let $\mathbf{T}(F, \xi)$ represent simple RT to a target stimulus presented after a foreperiod F, and characterized by a set of factors ξ. It is often assumed (see Everitt & Hand, 1981, and Luce, 1986, for overviews) that $\mathbf{T}(F, \xi)$ is a mixture of two time variables:

$$\mathbf{T}(F, \xi) = \begin{cases} \mathbf{T}(F) \geq -F & \text{with probability } 1 - p(F, \xi); \\ \mathbf{T}(\xi) \geq 0 & \text{with probability } p(F, \xi). \end{cases} \tag{22}$$

Here, $\mathbf{T}(\xi)$ represents "true", target-initiated, responses, and $\mathbf{T}(F)$ represents "false alarms", or target-unrelated responses, usually assumed to be stochastically "fast": $\text{Prob}\{\mathbf{T}(F) \leq t\} > \text{Prob}\{\mathbf{T}(\xi) \leq t\}$. These assumptions are based on the common observation that RT values in some trials may be very small or even negative (counted from the target signal onset).

A mixture model is not, of course, the only conceptual framework for false alarms.

In fact, a straightforward application of Theorem 1.5 to $\mathbf{T}(F, \xi)$ suggests the following, simpler picture. A response process $\bar{C}_{F,\xi}(t)$ is always initiated at $t = -F$ (warning signal onset), and up to $t = 0$ (target stimulus onset) it develops as a function of time-since-warning only, $\bar{C}_{F,\xi}(t) = \bar{C}_0(t + F)$. If the criterion $\mathbf{C}$ is set at a level below $\bar{C}_{F,\xi}(0) = \bar{C}_0(F)$, the process terminates before $t = 0$, resulting in an obvious false alarm (negative RT). After $t = 0$ the course of the process is generally influenced by target signal characteristics, ξ, but it will depend not only on ξ but also on F, through the course of the process preceding target onset. The target stimulus does not initiate a new process, but acts as a *modifier* of an already developing process. As a result, there is no sharp demarcation between "true" responses and "false alarms" for positive RTs (see section 1.7 for an example). It is remarkable that this theoretically very interesting approach is suggested by merely describing the generation of pretarget and post-target responses in the Grice modeling language. A somewhat similar picture was proposed by Laming (1968, Axiom 6, p. 82) in the context of a random-walk model, but it did not receive a deserving development in the subsequent RT literature (see Luce, 1986, p. 145).

The closest analogue of a target-modified process to a target-initiated one is a response process that has the following structure when measured on a positive scale calibrated by a unit-exponentially distributed criterion $\mathbf{C}$:

$$\bar{C}_{F,\xi}(t) = \begin{cases} \bar{C}_0(F + t) & \text{if } t < r; \\ \bar{C}_0(F + r) + \bar{C}_\xi(t - r) & \text{if } t \geq r; \end{cases}$$

where $r \geq 0$. This process is not affected by ξ up to $t = r$, but afterwards it develops depending on ξ exclusively, if measured with respect to the level it has reached at $t = r$. Applying Lemma 1.4, one has for $t \geq r$:

$$\mathcal{T}_{F,\xi}(t) = 1 - \exp\{-\bar{C}_0(F + r) - \bar{C}_\xi(t - r)\}.$$

The conditional probability that $\mathbf{T}(F, \xi) < t$, given that no response ("false alarm") occurs before r, is

$$\mathcal{T}_{F,\xi}(t|\mathbf{T}(F, \xi) > r) = 1 - \exp\{-\bar{C}_\xi(t - r)\}.$$

This simple relation demonstrates the convenience of using exponentially distributed criteria (see section 1.6).

Let us return now to (22). The time variable $\mathbf{T}(F, \xi)$ in (22) can be redefined as a label-and-time variable $[\mathbf{I}(F, \xi), \mathbf{T}(F, \xi)]$, $i = 0, 1$, with the following distribution function:

$$\mathcal{T}_{F,\xi}(i, t) = \begin{cases} \mathcal{T}_F(t)(1 - p(F, \xi)) & \text{for } i = 0; \\ \mathcal{T}_\xi(t)p(F, \xi) & \text{for } i = 1; \end{cases} \tag{23}$$

where $\mathcal{T}_F(t)$ and $\mathcal{T}_\xi(t)$ are the distribution functions of $\mathbf{T}(F)$ and $\mathbf{T}(\xi)$, respectively. Below $-F$ and 0, respectively, the two functions equal zero. The situation becomes formally identical to a 2-alternative choice paradigm, and Theorems 2.3.1 and 2.3.2 are applicable. Choosing, for simplicity, unit-exponentially distributed independent criteria $\{\mathbf{C}_0, \mathbf{C}_1\}$, and combining (23) with (18), one gets, after transformations, the following two response processes:

$$\widetilde{C}_{F,(\xi)}(t) = \int_{-F}^{t} \frac{d\mathcal{T}_F(u)(1 - p(F, \xi))}{1 - \mathcal{T}_F(u)(1 - p(F, \xi)) - \mathcal{T}_\xi(u)p(F, \xi)},$$
$$\widetilde{C}_{\xi,(F)}(t) = \int_{0}^{t} \frac{d\mathcal{T}_\xi(u)p(F, \xi)}{1 - \mathcal{T}_F(u)(1 - p(F, \xi)) - \mathcal{T}_\xi(u)p(F, \xi)}, \quad (24)$$

where the subscript listed first in $\widetilde{C}_{F,(\xi)}$ and $\widetilde{C}_{\xi,(F)}$ shows the "principal" factor that determines each of the processes, and establishes correspondence between these processes and the time variables in (22). Since $\mathcal{T}_\xi(t) = d\mathcal{T}_\xi(t)/dt = 0$ for $t < 0$, the second process is target-initiated: $\widetilde{C}_{\xi,(F)} = 0$ for $t < 0$. The course of this process after $t = 0$, however, generally depends not only on target characteristics ξ, but also on the foreperiod F. The first process $\widetilde{C}_{F,(\xi)}$, develops alone between $-F$ and 0, and it is easy to verify that within this interval it does not depend on ξ: $\widetilde{C}_{F,(\xi)}(t) = \widetilde{C}_F(t + F)$ (causal consistency). Its course after $t = 0$, however, does generally depend on ξ. One can see that as the two processes develop together, after $t = 0$, they both depend on both F and ξ. A natural interpretation of this fact is that the two processes interact, as discussed in section 2.6. By reversing the order of derivations, one can verify that if

$$\widetilde{C}_{F,(\xi)}(t) = \widetilde{C}_F(t + F) \quad \text{for all } t \geq -F;$$
$$\widetilde{C}_{\xi,(F)}(t) = \widetilde{C}_\xi(t) \quad \text{for all } t \geq 0,$$

that is, if the two processes depend on their respective principal factors only (and thereby they do not interact), then the resulting $\mathbf{T}(F, \xi)$ generally cannot be decomposed according to (22).

2.8. Other Mixtures.

The analysis of (22) in the previous section did not utilize the assumption that false alarms are stochastically faster than true responses, $\mathcal{T}_F(t) > \mathcal{T}_\xi(t)$. It is not surprising, therefore, that virtually the same mixture model and the same Grice-representability analysis can also be applied to other situations in which some proportion of responses are assumed to be target-unrelated.

As discussed in section 2.2, analysis of simple RT to weak signals leads naturally to the concept of an incomplete time variable. A response process evoked (or modified) by a weak signal may never reach a preset value of its termination criterion. In such cases, since response is compulsory, it should be generated by an additional mechanism. One might think of this mechanism as a stochastically slow "clocking" variable that terminates a trial if no response has been given within a certain period counted from the warning signal onset. Formally, this corresponds to the following mixture model:

$$\mathbf{T}^*(F, \xi) = \begin{cases} \mathbf{T}(F, \xi) \geq -F & \text{with probability } q(F, \xi); \\ \mathbf{T}^*(F) \geq -F & \text{with probability } 1 - q(F, \xi). \end{cases} \quad (25)$$

Here $\mathbf{T}^*(F)$ is the "clocking" variable with a distribution function $\mathcal{T}^*_F(t)$. The time variable $\mathbf{T}(F, \xi)$, a "normal" RT with a distribution function $\mathcal{T}_{F,\xi}(t)$, is assumed to be stochastically faster than $\mathbf{T}^*(F)$, that is $\mathcal{T}^*_F(t) < \mathcal{T}_{F,\xi}(t)$. The Grice-representability analysis of this model leads to two (interacting) processes, competing for their preset criteria, $\{\mathbf{C}_1, \mathbf{C}_2\}$. Choosing the criteria again unit-exponentially distributed and mutually independent, one gets:

$$\bar{C}_{F,\xi}(t) = \int_{-F}^{t} \frac{d\mathcal{T}_{F,\xi}(u)q(F,\xi)}{1-\mathcal{T}^*_F(u)(1-q(F,\xi))-\mathcal{T}_{F,\xi}(u)q(F,\xi)};$$
$$\bar{C}^*_{F,(\xi)}(t) = \int_{-F}^{t} \frac{d\mathcal{T}^*_F(u)(1-q(F,\xi))}{1-\mathcal{T}^*_F(u)(1-q(F,\xi))-\mathcal{T}_{F,\xi}(u)q(F,\xi)}. \quad (26)$$

It would be repetitious to discuss these processes in greater detail. Instead, consider how these processes change as a result of a further decomposition of the time variables being modeled. For example, the time variable $\mathbf{T}(F, \xi)$ in the analyzed mixture may in turn be thought of as a mixture of target-initiated RTs with false alarms. If so, then $\mathbf{T}(F, \xi)$ in (25) is decomposed according to (22), which means that (25) transforms into the following mixture model:

$$\mathbf{T}^*(F,\xi) = \begin{cases} \mathbf{T}(F) \geq -F & \text{with probability } q(F,\xi)(1-p(F,\xi)); \\ \mathbf{T}(\xi) \geq 0 & \text{with probability } q(F,\xi)p(F,\xi); \\ \mathbf{T}^*(\overline{F}) \geq -F & \text{with probability } 1-q(F,\xi). \end{cases} \quad (27)$$

The corresponding distribution functions are $\mathcal{T}_F(t) > \mathcal{T}_\xi(t) > \mathcal{T}^*_F(t)$. Choosing, as before, unit-exponentially distributed mutually independent criteria $\{\mathbf{C}_0, \mathbf{C}_1, \mathbf{C}_2\}$, one finds three response processes that correspond to the three components of the mixture:

$$\bar{C}_{F,(\xi)}(t) = \int_{-F}^{t} \frac{q(F,\xi)(1-p(F,\xi))d\mathcal{T}_F(u)}{\mathcal{D}_{F,\xi}(u)};$$

$$\bar{C}_{\xi,(F)}(t) = \int_{0}^{t} \frac{q(F,\xi)p(F,\xi)d\mathcal{T}_\xi(u)}{\mathcal{D}_{F,\xi}(u)};$$

$$\bar{C}^*_{F,(\xi)}(t) = \int_{-F}^{t} \frac{(1-q(F,\xi))d\mathcal{T}^*_F(u)}{\mathcal{D}_{F,\xi}(u)};$$

where $\mathcal{D}_{F,\xi}(u)$ stands for

$$1-\mathcal{T}_F(u)q(F,\xi)(1-p(F,\xi))-\mathcal{T}_\xi(u)q(F,\xi)p(F,\xi)-\mathcal{T}^*_F(u)(1-q(F,\xi)).$$

One can easily verify that, in agreement with Lemma 2.4, $\bar{C}^*_{F,(\xi)}(t)$ here remains the same as in (26), whereas

$$\bar{C}_{F,\xi}(t) = \bar{C}_{F,(\xi)}(t) + \bar{C}_{\xi,(F)}(t).$$

This is a particular case of (19), and its remarkable simplicity is yet another reason for using exponentially distributed criteria.

The Grice-representability analysis of the mixture models for simple RT can be trivially generalized to incorporate mixture models for choice RT. For example, if the conditional RTs $[\mathbf{T}(F,\xi) \mid \mathbf{I}(F,\xi) = i]$ in the N-alternative choice RT, $i = 1, \ldots, N$, are assumed to be decomposable according to (27), then the problem will be formally reduced to that of Grice-representing a $3N$-alternative label-and-time variable.

Conclusion

Obviously, to fit empirical RT distributions, one has to make specific assumptions about the shapes of response processes and criterion distributions. In the Introduction such assumptions were referred to as "auxiliary", as opposed to the principal idea of modeling RT in terms of stimulation-dependent deterministic processes developing until they reach randomly preset stimulation-independent criteria. Analogously, the principal idea in a model constructed according to the McGill scheme is that of a stochastic process (e.g., a deterministic function plus stationary stimulation-independent noise) reaching a fixed criterion; this idea is complemented by assumptions concerning the shapes of the function and noise distribution. It is a natural tendency to view empirical tests and comparisons of such models, say, a Grice-scheme model versus a McGill-scheme one, as tests and comparisons of logical conjunctions of the principal assumptions (call them P) with "auxiliary" ones, A: P_{Grice} & A_{Grice} versus P_{McGill} & A_{McGill}. I have shown in this paper that this is a misleading picture. The "principal assumptions", P_{Grice} and P_{McGill}, are not assumptions at all. They are mathematical languages universally applicable to all conceivable families of RT distributions: they cannot be "tested", and they are equivalent. What is being tested and compared are the "auxiliary" assumptions only. One can always translate A_{Grice} into A^*_{McGill}, for example, and to present the situation as A^*_{McGill} versus A_{McGill}, two sets of specific assumptions formulated in a common (McGill-scheme) language.

A mathematical language, when fully formulated and identified as a language, can be assessed, of course, in terms of its simplicity, convenience, and heuristic value. It has been shown that certain conceptual difficulties, such as stochastic independence of processing times that coexists with deterministic interdependence of processes themselves, are resolved by merely stating them in the language of the Grice modeling scheme. Also, it has been shown that certain concepts that otherwise might have been considered artificial, or even inadmissible, such as that of an incomplete time variable, arise naturally within the framework of the Grice modeling scheme, and play a useful role. Finally, it has been shown that Grice-representation of RT suggests some interesting possibilities that otherwise may have been overlooked, such as target-modified response processes instead of the dichotomy of target-initiated versus target-unrelated responses. All this allows one to conclude that the Grice modeling scheme is a good choice for a mathematical language, especially given its remarkable conceptual simplicity.

This paper by no means covers the entire scope of applicability of the Grice modeling scheme. For instance, it has been assumed throughout the paper that the moment when a response process crosses its criterion coincides with the initiation of an observable response. In the case of the choice RT paradigm, this means that the response processes compete for their criteria "till the very end", as it is assumed in the continuous flow models (Eriksen & Schultz, 1979; see also Coles, Gratton, Bashore, Eriksen, & Donchin, 1985). As a result, any experimental factor selectively affecting RTs for a given response, say $\text{Prob}[\mathbf{I}(\Xi) = 1, \mathbf{T}(\Xi) > t_0]$, will generally change the course of all response processes after moment t_0, rather than that of $\mathscr{C}_{1,\Xi}(t)$ alone (see sections 2.4 and 2.6). It is reasonable to assume, however, that RT can be additively decomposed into a "choice time" component and a "post-choice time" component, so that certain experimental factors can affect one of these components, but not the other (e.g., factors affecting the "motor difficulty" of a given response may be assumed to affect the post-choice component only). Stated formally,

$$[\mathbf{I}(\Xi), \mathbf{T}(\Xi)] = [\mathbf{I}(\Xi_1), \mathbf{T}_{\text{"pre"}}(\Xi_1) + \mathbf{T}_{\text{"post"}}(\Xi_2)],$$

where the two sets, Ξ_1 and Ξ_2, may intersect but generally do not coincide. The Grice modeling scheme should be applied to such a decomposition as follows. First, the label-and-time variable $[\mathbf{I}(\Xi_1), \mathbf{T}_{\text{"pre"}}(\Xi_1)]$ is modeled by competing response processes, as described in section 2.3. The first response process that crosses its criterion determines the value of $\mathbf{I}(\Xi_1)$ (response choice) and initiates the postchoice stage. The (conditional) time variables $[\mathbf{T}_{\text{"post"}}(\Xi_2) \mid \mathbf{I}(\Xi_1) = i]$ are modeled separately for different values of i, as described in section 1.5. This approach can be generalized further to incorporate multiple stages involving different forms of choice.

Our discussion of the Grice modeling scheme would not be complete without considering the possibility that certain experimental manipulations may affect criteria rather than response processes (see Assumption C in the Introduction). For example, in his 1972 paper Grice analyzes several RT distributions under the assumption that they all correspond to a single response process but to different criterion distributions. It was shown in this paper that this possibility is conceptually redundant: once response processes are assumed to vary depending on some external factors, all other factors, including experimental instructions, can also be thought to affect response processes only, coupled with a fixed (but arbitrarily chosen) criterion distribution. Moreover, there is a logical difficulty associated with the idea. Recall that numerical values of a response process are only defined with respect to a given criterion distribution. If the latter changes, the numerical values should change accordingly to represent the same, unchanged, process. The analogy with ratio scales mentioned in section 2.3 helps to clarify the issue. Let processes $C_\Xi(t)$ be defined on a ratio scale. What would it mean to say that a certain experimental manipulation affects the scaling factor rather than the processes themselves? Obviously, the meaning is that all the processes undergo one and the same similarity transformation, new $C_\Xi(t) = k$ times old $C_\Xi(t)$, which is convenient to present as remeasuring the "old" processes on a scale whose unit of measurement is k times smaller than before. Notice that if different processes underwent similarity transformations with different coefficients, this description would not be possible.

Returning to the case presented here, some external manipulations (say, experimental instructions) may be assumed to change all response processes (indexed by the values of other experimental factors) by one and the same monotonic transformation, $\mathcal{R}$, so that for any Ξ, new $C_\Xi(t) = \mathcal{R}(\text{old } C_\Xi(t))$. If this transformation is continuous and strictly increasing, it is convenient to present this situation as a transformation of the criterion distribution $\mathcal{C}$ into $\mathcal{R}^{-1}(\mathcal{C})$, while the processes retain their "old" numerical values. Due to Lemma 1.4, this is possible if and only if

$$\text{new } \mathcal{T}_\Xi(t) = \mathcal{R}^{-1}(\text{old } \mathcal{T}_\Xi(t)),$$

for all time variables $\mathbf{T}(\Xi)$. This is, of course, an empirically falsifiable assumption. It holds if and only if the "old" and "new" distribution functions $\mathcal{T}_\Xi(t)$ satisfy Schweickert's rectangle condition (Schweickert, 1985, Theorem 1b):

$$\text{old } \mathcal{T}_\Xi(t_1) < \text{old } \mathcal{T}_{\Xi'}(t_2) \quad \text{iff} \quad \text{new } \mathcal{T}_\Xi(t_1) < \text{new } \mathcal{T}_{\Xi'}(t_2),$$

for any Ξ, Ξ', t_1, t_2. This discussion of the criterion distribution changes can be trivially generalized to choice RT and disjunctive RT distribution families.

Appendix: Proofs

Lemma 1.3.1. The generating condition $C_\Xi(t) \geq \mathbf{C}$ in (1) is equivalent to $C_\Xi(t) - \mathbf{C} + c_0 \geq c_0$, where c_0 is an arbitrary constant. Consider a stationary random process

$\mathbf{N}(t)$ whose distribution at any moment t coincides with that of $c_0 - \mathbf{C}$. Then $C_{\Xi}(t) \geq \mathbf{C}$ iff $C_{\Xi}(t) + \mathbf{N}(t) \geq c_0$, and $\inf\{t\colon C_{\Xi}(t) \geq \mathbf{C}\} = \inf\{t\colon C_{\Xi}(t) + \mathbf{N}(t) \geq c_0\}$.

Lemma 1.3.2. If $C_{\Xi}(t)$ and $\mathbf{C}$ in (1) are not strictly positive, they can be redefined by putting new $C_{\Xi}(t) = \exp(\text{old } C_{\Xi}(t))$, and new $\mathbf{C} = \exp(\text{old } \mathbf{C})$. If $\inf(\text{new } \mathbf{C}) = 0$ then the same transformation can be applied once more. Then the generating condition $C_{\Xi}(t) \geq \mathbf{C}$ is equivalent to $C_{\Xi}(t)\mathbf{N}(t) \geq c_0$, where c_0 is an arbitrary positive constant, and $\mathbf{N}(t)$ is a stationary random process whose distribution at any moment t coincides with that of $c_0/\mathbf{C}$.

Lemma 1.4. Indeed, the following events are equivalent by definition, (3):

$$\bar{C}_{\Xi}(t) \geq \mathbf{C} \quad \text{iff} \quad C_{\Xi}(u) \geq \mathbf{C} \text{ for some } u \leq t \quad \text{iff} \quad \inf\{u\colon C_{\Xi}(u) \geq \mathbf{C}\} \leq t.$$

But $\mathscr{C}(\bar{C}_{\Xi}(t)) = \text{Prob}\{\bar{C}_{\Xi}(t) \geq \mathbf{C}\}$, and from (1),

$$\mathscr{T}_{\Xi}(t) = \text{Prob}\{\mathbf{T}(\Xi) \leq t\} = \text{Prob}\{\inf\{u\colon C_{\Xi}(u) \geq \mathbf{C}\} \leq t\}.$$

Theorem 1.5. Since $\mathscr{C}(c)$ is both strictly increasing and continuous, the quantile function $\mathscr{C}^{-1}(p)$ is a strictly increasing mapping from $0 < p < 1$ to the c-axis, uniquely defined for all p within this interval. Then the function $\mathscr{C}^{-1}(\mathscr{T}_{\Xi}(t))$ is defined for all t such that $0 < \mathscr{T}_{\Xi}(t) < 1$, and it is nondecreasing because $\mathscr{T}_{\Xi}(t)$ is a nondecreasing mapping from the t-axis to (0, 1). Therefore $\mathscr{C}^{-1}(\mathscr{T}_{\Xi}(t))$ is an HRL-form of some response process (e.g., of itself), and can be denoted by $\bar{C}_{\Xi}(t)$. Due to Lemma 1.4, the distribution function generated by $\bar{C}_{\Xi}(t)$ coupled with $\mathbf{C}$ is $\mathscr{C}(\bar{C}_{\Xi}(t))$. But

$$\mathscr{C}(\bar{C}_{\Xi}(t)) = \mathscr{C}(\mathscr{C}^{-1}(\mathscr{T}_{\Xi}(t))),$$

and since $\mathscr{C}$ is strictly increasing and continuous,

$$\mathscr{C}(\mathscr{C}^{-1}(\mathscr{T}_{\Xi}(t))) = \mathscr{T}_{\Xi}(t),$$

which is the distribution function for $\mathbf{T}(\Xi)$.

Lemma 1.5. Let $\mathscr{C}(c)$ have a discontinuity of the first kind at some value c_0, that is $\lim \mathscr{C}(c) = p < 1$ as $c \to c_0 - 0$, but $\mathscr{C}(c) = p + p_0 \leq 1$, $p_0 > 0$. Let a time-dimensioned variable $\mathbf{T}$ have a continuous distribution function $\mathscr{T}(t)$ such that $\mathscr{T}(t) \geq p$ beginning with some value of t. (Such a random variable, obviously, exists.) If there is a response process $C(t)$ that generates $\mathbf{T}$ when coupled with $\mathbf{C}$, then $t_0 = \inf\{t\colon C(t) \geq c_0\}$ should exist as a finite value: otherwise $C(t)$ would always be less than c_0 (or undefined), and $\mathscr{T}(t) = \mathscr{C}(\bar{C}(t))$ would be less than p (or undefined) for any t, contrary to the agreement above. Then $\mathscr{T}(t) \geq \mathscr{C}(c_0) = p + p_0$ for any $t > t_0$, whereas for any $t < t_0$, $\mathscr{T}(t) < p$. Since $p_0 > 0$, this contradicts the assumption that $\mathscr{T}(t)$ is continuous.

Lemma 1.8. Due to (5), there is a point-to-point correspondence between $\bar{C}_{\Xi}(t)$ and $\mathscr{T}_{\Xi}(t)$: $\bar{C}$ depends on t and on Ξ only through its dependence on $\mathscr{T}_{\Xi}(t)$.

Lemma 1.9.1. From Lemma 1.4 and Theorem 1.5, $\mathbf{T}(\Xi^*, b)$ is generated by a response process $C_{\Xi^*,b}(t)$ coupled with $\mathbf{C}$, iff

$$\mathscr{C}^{-1}(\mathscr{T}_{\Xi^*}(t|\mathbf{B} = b)) = \bar{C}_{\Xi^*,b}(t), \tag{A1}$$

and

$$\mathcal{T}_{\Xi^*}(t|\mathbf{B} = b) = \mathcal{C}(\bar{C}_{\Xi^*,b}(t)). \tag{A2}$$

At the same time, $\mathbf{T}(\Xi^*)$ is generated by a response process $C_{\Xi^*}(t)$ coupled with the same $\mathbf{C}$, iff

$$\mathcal{C}^{-1}(\mathcal{T}_{\Xi^*}(t)) = \bar{C}_{\Xi^*}(t), \tag{A3}$$

and

$$\mathcal{T}_{\Xi^*}(t) = \mathcal{C}(\bar{C}_{\Xi^*}(t)). \tag{A4}$$

Substituting (A1) and (A3) in (9) on gets (8), substituting (A2) and (A4) in (8) one gets (9).

Lemma 1.9.2. Due to Lemma 1.9.1, it is sufficient to show that (10) implies (9). If (10) holds, then for $u \leq t$,

$$\sup\mathcal{C}(C_{\Xi^*}(u)) = \sup \int_b \mathcal{C}(C_{\Xi}(u))\ d\mathcal{B}(b).$$

But $\sup\mathcal{C}(C_{\Xi^*}(u)) = \mathcal{C}(\sup C_{\Xi^*}(u)) = \mathcal{C}(\bar{C}_{\Xi^*}(t))$, whereas

$$\sup \int_b \mathcal{C}(C_{\Xi}(u))\ d\mathcal{B}(b) = \int_b \sup\mathcal{C}(C_{\Xi}(u))\ d\mathcal{B}(b)$$

$$= \int_b \mathcal{C}(\sup C_{\Xi}(u))\ d\mathcal{B}(b) = \int_b \mathcal{C}(\bar{C}_{\Xi}(t))\ d\mathcal{B}(b).$$

Theorem 2.2. The proof is essentially the same as for Theorem 1.5. The function $\mathcal{C}^{-1}(\mathcal{T}_{\Xi}(1, t))$ is a nondecreasing mapping from $(0, p(\Xi))$ into the c-axis. Therefore $\mathcal{C}^{-1}(\mathcal{T}_{\Xi}(1, t))$ is an HRL-form of some response process $C_{\Xi}(t)$, and can be denoted by $\bar{C}_{\Xi}(t)$. Applying Lemma 1.4 (which did not require that the time variable generated be complete), one can see that $\bar{C}_{\Xi}(t)$ coupled with $\mathbf{C}$ generates a time variable with the distribution function $\mathcal{C}(\bar{C}_{\Xi}(t))$. Substituting $\mathcal{C}^{-1}(\mathcal{T}_{\Xi}(1, t))$ for $\bar{C}_{\Xi}(t)$, one proves that the generated distribution function is $\mathcal{T}_{\Xi}(1, t)$. This means that

$$[\mathbf{I}(\Xi) = 1 \text{ and } \mathbf{T}(\Xi) = t] \quad \text{iff} \quad \inf\{u\colon C_{\Xi}(u) \geq \mathbf{C}\} = t.$$

But $C_{\Xi}(u) \geq \mathbf{C}$ is possible at some value of u iff $\mathbf{C} \leq \sup C_{\Xi}(t)$. Otherwise the infimum, and thereby $\mathbf{T}(\Xi)$, are undefined, and $\mathbf{I}(\Xi)$ in $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$ can be set equal to 2. Observe that

$$\text{Prob}\{\mathbf{C} \leq \sup C_{\Xi}(t)\} = \text{Prob}\{\mathbf{C} \leq \sup\bar{C}_{\Xi}(t)\}$$

$$= \text{Prob}\{\mathbf{C} \leq \mathcal{C}^{-1}(p(\Xi))\} = p(\Xi).$$

Theorem 2.3.1. The proof is greatly simplified by the following transformations of the variables. First, we transform every value of t into $\bar{t} = \Phi_{\Xi}(t)$, where Φ_{Ξ} is a strictly increasing continuous mapping of $(\inf\mathbf{T}(\Xi), \sup\mathbf{T}(\Xi))$ onto $(0, 1)$. Recall from section 2.1, that $\mathcal{T}_{\Xi}(i, t)$ is assumed to be continuously differentiable with respect to a strictly increasing continuous function $\psi_{\Xi}(t)$ everywhere, except (possibly) at some points $\psi_{\Xi}(t_1) < \cdots < \psi_{\Xi}(t_m)$. We choose the transformation $\bar{t} = \Phi_{\Xi}(t)$ so that for all $\mathcal{T}_{\Xi}(i, t) = \bar{\mathcal{T}}_{\Xi}(i, \bar{t})$, $i = 1, \ldots, N$, the derivatives $d\bar{\mathcal{T}}_{\Xi}(i, \bar{t})/d\bar{t}$ exist and are

continuous on the adjacent *closed* intervals $[0, \bar{t}_1]$, $[\bar{t}_1, \bar{t}_2]$, ..., $[\bar{t}_m, 1]$, where $\bar{t}_1 < \cdots < \bar{t}_m$ correspond to the moments of discontinuity $\psi_\Xi(t_1) < \cdots < \psi_\Xi(t_m)$ for $d\mathcal{T}_\Xi(i, t)/d\psi_\Xi(t)$. One such transformation is

$$\bar{t} = \Phi_\Xi(t) = \frac{U(\psi(t)) + \sum_1^N \mathcal{T}_\Xi(i, t)}{N + 1},$$

where $U(\psi(t))$ is an arbitrary distribution function strictly increasing and continuously differentiable with respect to $\psi_\Xi(t)$ on $(\inf\mathbf{T}(\Xi), \sup\mathbf{T}(\Xi))$. One can verify that the distribution functions $\bar{\mathcal{T}}_\Xi(i, \bar{t})$ are continuous on their domain, (0, 1), and based on limit considerations, one can put $\bar{\mathcal{T}}_\Xi(i, 0) = 0$, $\bar{\mathcal{T}}_\Xi(i, 1) = 1$. One can also verify that

$$0 < \frac{d\bar{\mathcal{T}}_\Xi(i, \bar{t})}{d\bar{t}} = \frac{(N + 1)d\mathcal{T}_\Xi(i, t)}{dU(\psi(t)) + \sum_1^N d\mathcal{T}_\Xi(j, t)} < (N + 1).$$

It is clear from this expression that $d\bar{\mathcal{T}}_\Xi(i, \bar{t})/d\bar{t}$ can only have discontinuities of the first kind (finite jumps), and that these may only occur at the (possible) discontinuity points $\bar{t}_1, \ldots, \bar{t}_m$. The limits of $d\bar{\mathcal{T}}_\Xi(i, \bar{t})/d\bar{t}$, as $\bar{t}$ approaches 0 or 1, are finite and can be viewed as the values of the derivative at 0 and 1, respectively. Both $\bar{\mathcal{T}}_\Xi(i, \bar{t})$ and $d\bar{\mathcal{T}}_\Xi(i, \bar{t})/d\bar{t}$, therefore, are defined on [0, 1].

The second transformation is that of the criteria. We transform every criterion $\mathbf{C}_i$ into $\bar{\mathbf{C}}_i = \mathcal{C}_i(\mathbf{C}_i)$. The marginal distributions $\mathcal{C}_i$ of $\{\bar{\mathbf{C}}_1, \ldots, \bar{\mathbf{C}}_N\}$ are uniform between 0 and 1: $\mathcal{C}_i(\bar{c}_i) = \bar{c}_i$. Obviously,

$$\text{Prob}\{\mathbf{C}_j > c_j \text{ for all } j \neq i | \mathbf{C}_i = c_i\} = \text{Prob}\{\bar{\mathbf{C}}_j > \bar{c}_j \text{ for all } j \neq i | \bar{\mathbf{C}}_i = \bar{c}_i\}.$$

or, in the notation adopted,

$$\bar{\mathcal{C}}_i^-(\bar{c}_1, \ldots, \bar{c}_N) = \mathcal{C}_i^-(c_1, \ldots, c_N).$$

It follows from the admissibility conditions that all $\bar{\mathcal{C}}_i^-(\bar{c}_1, \ldots, \bar{c}_N)$ are positive and locally Lipschitzean with respect to all arguments on their domain $(0, 1)^N$. Based on limit considerations, the values of these conditionals can be extended to the domain boundaries (retaining all admissibility properties). In particular, $\bar{\mathcal{C}}_i^-(\bar{c}_1, \ldots, \bar{c}_N) = 1$ when all its arguments are zero, and $\bar{\mathcal{C}}_i^-(\bar{c}_1, \ldots, \bar{c}_N) = 0$ iff some of the arguments equal 1.

To simplify the notation, *the upper bars will be dropped from the notation for the transformed variables and functions*, and it will be assumed instead that all response processes are described in transformed time (denoted by t rather than $\bar{t}$) and on the transformed scales (c_i rather than $\bar{c}_i$). One simply has to remember that each response process $C_{i,\Xi}(t)$ thus described will eventually have to be transformed back into real time and "real" (i.e., calibrated by the initially assumed criterion distribution) scale of values:

$$\text{"real"}C_{i,\Xi}(\text{real } t) = \mathcal{C}_i^{-1}(C_{i,\Xi}(\Phi_\Xi^{-1}(t)). \tag{A5}$$

Our goal now is to find continuous nondecreasing functions $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$ for which

$$[\mathbf{I}(\Xi) = i,\ \mathbf{T}(\Xi) = t]$$

iff

$$[\mathbf{C}_j > C_{j,\Xi}(t) \text{ for } j \neq i,\ \mathbf{C}_i = C_{i,\Xi}(t),\ \mathbf{C}_i > C_{i,\Xi}(u) \text{ for } u < t].$$

This is equivalent to requiring that

$$\text{Prob}\{\mathbf{I}(\Xi) = i,\ t - \mathrm{d}t < \mathbf{T}(\Xi) \leq t\}$$

equals

$$\text{Prob}\{\mathbf{C}_j > C_{j,\Xi}(t) \text{ for } j \neq i, \quad \text{and} \quad \mathscr{C}_i(C_{i,\Xi}(t - \mathrm{d}t)) < \mathscr{C}_i(\mathbf{C}_i) \leq \mathscr{C}_i(C_{i,\Xi}(t))\}.$$

Expressing the joint probability as a product of the corresponding conditional and marginal, and recalling that the marginals are uniform between 0 and 1, we come to the following differential equation:

$$\mathrm{d}\mathscr{T}_\Xi(i,\ t) = \mathscr{C}_i^-(C_{1,\Xi},\ \ldots,\ C_{N,\Xi})\mathrm{d}C_{i,\Xi},\ i = 1,\ \ldots,\ N,$$

or, in a normal form,

$$\frac{\mathrm{d}C_{i,\Xi}}{\mathrm{d}t} = \frac{\dfrac{\mathrm{d}\mathscr{T}_\Xi(i,\ t)}{\mathrm{d}t}}{\mathscr{C}_i^-(C_{1,\Xi},\ \ldots,\ C_{N,\Xi})},\ i = 1,\ \ldots,\ N, \tag{A6}$$

subject to the initial condition

$$C_{i,\Xi}(t) = 0 \text{ at } t = 0,\ i = 1,\ \ldots,\ N. \tag{A7}$$

(In the original, untransformed coordinates, the initial condition is $C_{i,\Xi}(t) \to \inf \mathbf{C}_i$ as $t(\Xi) \to \inf \mathbf{T}(\Xi)$.)

Consider the following series of $(N + 1)$-dimensional rectangles:

$$[t_k \leq t \leq t_{k+1}] \times [0 \leq C_1 < 1) \ldots \times [0 \leq C_N < 1),$$

$$k = 0,\ 1,\ \ldots,\ m,\ m + 1, \tag{A8}$$

where $t_0 = 0, t_1, \ldots, t_m$ are the (possible) discontinuity points for $\mathrm{d}\mathscr{T}_\Xi(i,\ t)/\mathrm{d}t$, and $t_{m+1} = 1$. Within each of these rectangles the right-hand sides of (A6) are continuous in t (by construction) and locally Lipschitzean and all C-arguments. The latter follows from observing that

$$\left| \frac{\dfrac{\mathrm{d}\mathscr{T}_\Xi(i,\ t)}{\mathrm{d}t}}{\mathscr{C}_i^-(c_1 + \Delta c_1,\ \ldots,\ c_N + \Delta c_N)} - \frac{\dfrac{\mathrm{d}\mathscr{T}_\Xi(i,\ t)}{\mathrm{d}t}}{\mathscr{C}_i^-(c_1,\ \ldots,\ c_N)} \right| \leq M \sum_1^N |\Delta c_i|,$$

for any two points $(t, c_1, \ldots, c_N)$ and $(t, c_1 + \Delta c_1, \ldots, c_N + \Delta c_N)$ within a closed subregion of (A8). Indeed, denoting by L the Lipschitz constant for $\mathscr{C}_i^-(C_{1,\Xi}, \ldots, C_{N,\Xi})$ in this sub-region (which exists due to the admissibility assumptions), one can put

$$M = \frac{L}{\min \mathscr{C}_i^-(c_1, \ldots, c_N)^2} \max \frac{\mathrm{d}\mathscr{T}_\Xi(i,\ t)}{\mathrm{d}t},$$

where the denominator is finite and positive because $\mathscr{C}_i^-(c_1, \ldots, c_N)$ is positive and continuous (due to the admissibility assumptions) and the subregion is closed; analogously, the maximum of the derivative exists and is nonnegative.

Consider now the first in the series of rectangles (A8), the one with the time base $[0, t_1]$. Since the right-hand sides of (A6) are continuous in t and locally Lipschitzean in all C-arguments, we can apply the Picard theorem for ordinary differential equations, combined with an appropriate extension theorem (see, e.g., Matveev, 1974, Cronin, 1980), to get the following result: system (A6, A7) has a unique maximal solution $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$ within the rectangle. More specifically, there are N unique functions $C_{i,\Xi}(t)$ that start at the origin,

$$(t = 0,\ C_{1,\Xi}(0) = 0, \ldots, C_{N,\Xi}(0) = 0),$$

develop according to (A6), and end at some of the "opposite" boundaries of the rectangle,

$$t = t_1 \text{ or } C_{1,\Xi}(u) \to 1 \text{ or } \ldots \text{ or } C_{N,\Xi}(u) \to 1 \text{ as } u \to t \leq t_1,$$

"or" being nonexclusive. By construction, for any moment t within this interval, the event $[\mathbf{I}(\Xi) = i, \mathbf{T}(\Xi) = t]$ occurs iff $[\mathbf{C}_j > C_{j,\Xi}(t)$ for $j \neq i$ and $\mathbf{C}_i = C_{i,\Xi}(t)$ and $\mathbf{C}_i > C_{i,\Xi}(u)$ for $u < t]$. It follows then that the values of $C_{i,\Xi}(t)$ are such that

$$\text{Prob}\{\mathbf{T}(\Xi) \leq t\} = 1 - \text{Prob}\{\mathbf{C}_i > C_{i,\Xi}(t) \text{ for all } i\},$$

so that $\text{Prob}\{\mathbf{T}(\Xi) \leq t\} = 1$ iff at least one of $C_{i,\Xi}(t) = 1$. From this we conclude that the "opposite" boundaries of the rectangle, at which the solutions $C_{i,\Xi}(t)$ end, may only be of one of the following two kinds:

1. $t_1 = 1$ and $(C_{1,\Xi}(t) \to 1$ or $\ldots$ or $C_{N,\Xi}(t) \to 1)$ as $t \to 1$;

or

2. $t_1 < 1$ and $C_{1,\Xi}(t_1) < 1$ and $\ldots C_{N,\Xi}(t_1) < 1$.

The first case occurs iff there are no discontinuity points between 0 and 1 (i.e., $m = 0$), and then system (A6, A7) is solved. In the second case, we fix the point (b) as the initial condition for the next step, replacing (A7). Then we seek the solution of (A6), subject to this new initial condition, in the next rectangle of series (A8), this time with the time base $[t_1, t_2]$. Obviously, everything said about the previous rectangle applies here as well, and we continue in this process until we come to the last rectangle, with the time base $[t_m, 1]$, and reach the final endpoint there. This endpoint should necessarily be of Type 1 above: at least one of the functions $C_{i,\Xi}(t)$ approaches 1 (the upper limit of $\mathbf{C}_i$) as $t \to 1$. (Note, however, that as many as $N-1$ remaining functions may end up below 1.)

Observe now that since the right-hand expressions in system (A6) are never negative, the functions $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$ are all nondecreasing. Therefore they can be considered HRL-forms of some response processes. A function $C_{i,\Xi}(t)$ is nonincreasing within some interval iff $\mathrm{d}\mathscr{T}_\Xi(i, t)/\mathrm{d}t = 0$ within this interval. These properties do not change after we apply transformation (A5) to express the solutions in original coordinates. This completes the proof.

(The local Lipschitz condition serves to insure the uniqueness of the solution of (A6, A7), and this is the main reason for including it in the admissibility conditions for criteria. One might, therefore, consider replacing the local Lipschitz condition with stronger assumptions, such as piecewise continuous differentiability, or weaker as-

sumptions, such as the local Osgood condition for differential equations (see, e.g., Petrovski, 1964).)

Theorem 2.3.2. Assume that the variables have been transformed as in the previous theorem. The conditionals for mutually independent criteria (uniformly distributed between 0 and 1) are, simply,

$$\mathscr{C}_i^-(c_1, \ldots, c_N) = \frac{\prod_{j=1}^{N}(1 - c_j)}{1 - c_i}. \tag{A9}$$

They are Lipschitzean (in fact, globally, with the Lipschitz constant equal to 1), because they are continuously differentiable in all arguments, with the values of the derivatives being -1 or 0 only. Theorem 2.3.1 is, therefore, applicable, and a unique solution $\{C_{1,\Xi}(t), \ldots, C_{N,\Xi}(t)\}$ exists. Combining (A9) with (A6), and observing that

$$\prod_{j=1}^{N}(1 - C_{j,\Xi}(t)) = 1 - \mathcal{T}_\Xi(t),$$

we derive, after elementary transformations,

$$\frac{\dfrac{dC_{i,\Xi}}{dt}}{1 - C_{i,\Xi}} = -\frac{d\log(1 - C_{i,\Xi})}{dt} = \frac{d\mathcal{T}_\Xi(i, t)/dt}{1 - \mathcal{T}_\Xi(t)},$$

from which (18) is obtained by integration followed by transformations (A5).

Lemma 2.4. Observe that

$$\mathcal{T}_\Xi(i^*, t) = \sum_{i \in L_i^*} \mathcal{T}_\Xi(i, t),$$

whereas the marginal distributions $\mathcal{T}_\Xi(t)$ for the two label-and-time variables, $[\mathbf{I}^*(\Xi), \mathbf{T}^*(\Xi)]$ and $[\mathbf{I}(\Xi), \mathbf{T}(\Xi)]$, are identical. Equation (19) now immediately follows from (18), after elementary transformations.

References

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89*, 599–607.

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.

Cronin, J. (1980). *Differential equations: Introduction and qualitative theory*. New York: Marcel Dekker.

Dzhafarov, E. N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology, 36*, 235–268.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics, 25*, 249–263.

Everitt, B., & Hand, D. J. (1981). *Finite mixture distributions*. New York: Chapman and Hall.

Feller, W. (1968). *An introduction to probability theory and its applications, Vol. 1*. New York: Wiley.

Green, D. M., & Luce, R. D. (1974). Timing and counting mechanisms in auditory discrimination and reaction time. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds), *Contemporary developments in Mathematical Psychology, Vol. 2* (pp. 372–415). San Francisco: Freeman.

Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review, 75*, 359–373.

Grice, G. R. (1972). Application of a variable criterion model to auditory reaction time as a function of the type of catch trial. *Perception & Psychophysics, 12*, 103–107.

Grice, G. R., Canham, L., & Boroughs, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics, 35*, 451–463.

Grice, G. R., Nullmeyer, R., & Spiker, V. A. (1982). Human reaction time: Toward a general theory. *Journal of Experimental Psychology: General, 111*, 135–153.

LaBerge, D. A. (1962). A recruitment theory of simple behavior. *Psychometrica, 27*, 375–396.

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.

Link, S. W. (1975). The relative judgement theory of two-choice response time. *Journal of Mathematical Psychology, 12*, 114–135.

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

Marley, A. A. J. (1992). A selective review of recent characterizations of stochastic choice models using distribution and functional equation techniques. *Mathematical Social Sciences, 23*, 5–29.

Marley, A. A. J., & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risk interpretation. *Journal of Mathematical Psychology, 36*, 1–20.

Matveev, N. M. (1974). *Methods of integration of ordinary differential equations*. Minsk: Vysheishaia Shkola. (in Russian)

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287–330.

McGill, W. (1963). Stochastic latency mechanisms. In R. D. Luce & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. 1* (pp. 309–360). New York: Wiley.

Pacut, A. (1977). Some properties of threshold models of reaction latency. *Biological Cybernetics, 28*, 63–72.

Petrovski, I. G. (1964). *Lectures on theory of ordinary differential equations*. Moscow: Nauka. (in Russian)

Pike, R. (1973). Response latency models for signal detection. *Psychological Review, 80*, 53–68.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Schweickert, R. (1985). Separable effects of factors on activation functions in discrete and continuous models: d' and evoked potentials. *Psychological Bulletin, 106*, 318–328.

Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology, 14*, 219–238.

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

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13*, 37–58.

Manuscript received 12/26/91
Final version received 5/22/92