URL for data and model code: http://matt.colorado.edu/universality

Unfalsifiability and Mutual Translatability of Major Modeling Schemes for Choice Reaction Time

Matt Jones
University of Colorado, Boulder

Ehtibar N. Dzhafarov
Purdue University

Much current research on speeded choice utilizes models in which the response is triggered by a stochastic process crossing a deterministic threshold. This article focuses on two such model classes, one based on continuous-time diffusion and the other on linear ballistic accumulation (LBA). Both models assume random variability in growth rates and in other model components across trials. We show that if the form of this variability is unconstrained, the models can exactly match any possible pattern of response probabilities and response time distributions. Thus, the explanatory or predictive content of these models is determined not by their structural assumptions, but rather by distributional assumptions (e.g., Gaussian distributions) that are traditionally regarded as implementation details. Selective influence assumptions (i.e., which experimental manipulations affect which model parameters) are shown to have no restrictive effect, except for the theoretically questionable assumption that speed-accuracy instructions do not affect growth rates. The second contribution of this article concerns translation of falsifiable models between universal modeling languages. Specifically, we translate the predictions of the diffusion and LBA models (with their parametric and selective influence assumptions intact) into the Grice modeling framework, in which accumulation processes are deterministic and thresholds are random variables. The Grice framework is also known to reproduce any possible pattern of response probabilities and times, and hence it can be used as a common language for comparing models. It is found that only a few simple properties of empirical data are necessary predictions of the diffusion and LBA models.

Keywords: choice reaction time, diffusion model, linear ballistic accumulator, Grice framework, model falsifiability

---

Matt Jones, Department of Psychology and Neuroscience, University of Colorado; Ehtibar N. Dzhafarov, Department of Psychological Sciences, Purdue University. This research was supported by AFOSR grant FA9550-10-1-0177 to MJ and NSF grant SES-1155956 to END. We thank Jerome Busemeyer, Andrew Heathcote, Roger Ratcliff, Rich Shiffrin, and Jim Townsend for helpful comments on this work, and Roger Ratcliff for sharing his data. Correspondence regarding this article should be addressed to MJ at mcj@colorado.edu.

1. INTRODUCTION

A number of theoretical frameworks have been developed for modeling choice reaction time (RT) in terms of certain psychological processes developing until they reach threshold levels, or boundaries. Such processes may be thought of as evidence accumulation or response readiness, but we use the neutral term *response processes*. In most contemporary models, the response processes are stochastic, and the decision thresholds are fixed under given observation conditions. In this article we focus on two classes of such models: the linear ballistic accumulator (LBA; Brown & Heathcote, 2008) and the class of diffusion models, which includes the Wiener diffusion model (Ratcliff, 1978) and the Ornstein-Uhlenbeck (OU) model (Busemeyer & Townsend, 1993).

As an alternative, Grice (1968, 1972; Grice, Cahnam & Boroughs, 1984; Grice, Nullmeyer, & Spiker, 1982) proposed a class of models in which the response processes are deterministic, and the thresholds are stochastically sampled for each trial. Grice's formulation involved specific assumptions about the shape of the response processes (linear or negatively accelerated exponential) and the distributions of the thresholds (independent normal). These assumptions are clearly of a "technical" nature rather than part of Grice's "main idea"—they are introduced to fit empirical data and are supposed to be freely modifiable if needed. Thus, the linear response processes posited in Grice (1968) were replaced with the negatively accelerated exponential ones in Grice's later publications, and this was not presented as changing the essence of the model. Dzhafarov (1993) showed, however, that with these technical assumptions removed, the Grice framework can match any pattern of response probabilities and RTs, across any set of stimuli and other experimental conditions (this result is recapitulated below as Theorem 1). This mathematical result means that the Grice architecture (deterministic response processes racing to probabilistically chosen thresholds) is not a falsifiable model but rather a universal modeling language. Its only testable aspects are the "technical" constraints that may be imposed on its constituents.

The distinction between universal theoretical languages and empirically falsifiable models formulated in these languages is the leitmotif of this article. A comparison of competing models is often thought of as a contrast of their main ideas. This is incorrect, however, if the competing main ideas can be shown to be universally applicable theoretical languages.[1] In that case they are equivalent, and one can always translate a model formulated in one

[1] RT distributions are usually assumed to satisfy some regularity conditions, such as being representable by probability density functions, and the universality of a theoretical language may be confined to such regularity conditions. This should not prevent one from calling the language "universal" insofar as the regularity conditions in question are considered firmly established.

language into the other. The contrast in such a case occurs between the technical assumptions of the models only, such as the common assumption of normality in the distribution of growth rates of each response process.

To clarify this argument by a simple analogy, any RT distribution can be represented by a cumulative distribution function $F(t) = \Pr[RT \leq t]$, log survival function $S(t) = -\log \Pr[RT > t]$, odds ratio function $H(t) = \Pr[RT \leq t] / \Pr[RT > t]$, etc. Under the commonly accepted assumption of differentiability, one can add to this list density functions $f(t) = \mathrm{d}F(t)/\mathrm{d}t$, hazard functions $h(t) = f(t)/(1 - F(t))$, reverse hazards $s(t) = f(t)/F(t)$, etc. Due to the simplicity and familiarity of these representations, their universality and their mutual equivalence are obvious. Using one representation over another is merely a matter of convenience. However, if one makes a simplifying assumption within one of these representations, for example that $F(t)$ is a linear function on some interval $a < t < b$, and someone else makes the simplifying assumption that $S(t)$ is a linear function on the same interval, then the two descriptions become falsifiable models, one of which may very well be more correct than the other.

This simple analogy also illustrates that non-falsifiability need not mean deficiency or "emptiness" of one's construction. A modeling language might be universal but nevertheless valuable in enabling formulation of falsifiable models that otherwise could not be stated. Using our analogy, one cannot make the falsifiable statement that $F(t)$ is linear on some interval without first introducing the universally applicable notion of a cumulative distribution function. Problems only arise when one does not distinguish a descriptive language from a model formulated in that language. In some fields of psychology this distinction may not be easy to achieve, but in areas amenable to rigorous mathematical formulations (such as the models of RT analyzed in this article) it is relatively straightforward.

The present article pursues two goals. First, we prove that both the LBA and diffusion models, when stripped of their "technical" assumptions, are non-falsifiable modeling languages, on a par with the Grice modeling framework (see Section 2). Specifically, both the LBA and diffusion models assume random trial-by-trial variability in the growth rates and starting points of their response processes, respectively following Gaussian and uniform distributions. These distributions were chosen for mathematical convenience rather than theoretical reasons. The models also assume the growth-rate distribution is invariant with respect to certain experimental manipulations, although we argue that this assumption is poorly motivated (see Section 3). The primary results of this article are proofs that, if the constraints on the distributions of growth rates are removed, the diffusion and LBA models become universal: Any pattern of response probabilities and RT distributions, across any set of stimuli and experimental conditions, can be fit exactly. In certain nonlinear ballistic accumulator models (including that of Brown & Heathcote, 2005, except for one additional

technical detail), universality can also arise from start-point variability, if that distribution is unconstrained. These unfalsifiability results for the generalized models, with distributional constraints removed, imply that the predictive content of the standard models is fully determined by their assumptions about these distributions. This conclusion does not apply to stochastic accumulation models that exclude between-trial variability in growth rates or starting points (e.g., Link & Heath, 1975; Palmer, Huk, & Shadlen, 2005), but it has significant implications for models that do.

Our second goal is to develop methods for *translating* falsifiable models between universal modeling languages. When two languages are universal, any constrained model (e.g., defined by parametric restrictions on one or more of its components) expressed in one language can be re-expressed as an equivalent model in the other language. Here we translate the diffusion and LBA models (with all their technical assumptions included) into the Grice modeling language, by deriving Grice models that make the same predictions (see Section 4). The specific goal is to determine the role that the technical assumptions of the LBA and diffusion models play in explaining empirical data, and to develop an understanding of the predictive constraints these assumptions impose that is not tied to one modeling framework. The free parameters of the original models are systematically varied to investigate how they affect the course of the Grice response processes. Grice representations are then derived for the diffusion and LBA models with their parameters fit to empirical data, as well as for the empirical data themselves (taken from Ratcliff, Thapar, & McKoon, 2001), to understand what aspects of the data the models capture, and how. These analyses demonstrate that translating technical assumptions (i.e., specific models) between different modeling languages can provide greater insight into their logical content, as well as into the patterns of behavior those assumptions can explain.

In addition to their parametric distributional assumptions, the LBA and diffusion models make two *selective influence* assumptions about how model parameters can depend on different aspects of stimulus and observation conditions. First, it is assumed that variation of stimulus difficulty (e.g., perceptual discriminability or lexical frequency) affect only the growth rates of the response processes and not other parameters, such as threshold values and starting points. Second, it is assumed that manipulations of speed-accuracy instructions, usually held constant across a block of trials, affect only the threshold parameter (and start-point variability in the LBA). We consider the implications of these assumptions in Section 3. The first selective influence assumption has no impact on our universality results (i.e., the models are still unfalsifiable with this assumption), whereas the second assumption introduces predictive constraints and hence makes the models falsifiable. However, whereas the justification commonly offered for the first selective influence assumption is logically compelling, we argue that the justification for the second is not. We also consider the im-

plications of selective influence for mutual translatability between modeling languages, and demonstrate that the notion of selective influence *per se* cannot be used to distinguish between equivalent languages. This follows from the fact that when a model is translated from one modeling language into another, the precise parameterization of the original model can be carried over to the new language, with all assumptions of selective influences included. The selective influence assumptions therefore are never "lost in translation," even though the parameters in question may change their interpretations.

To prevent misunderstanding, we are not proposing to replace the different modeling languages investigated in this article with just one of them (say, the Grice framework). Universality and mutual equivalence still allow for one theoretical description to be more convenient or more transparent than another. Returning to our simple analogy, some properties of RTs may be more apparent when represented by hazard functions than by distribution functions (Luce, 1986). Despite the equivalence among the diffusion, LBA, and Grice modeling languages, as well as the variety of modifications and simplifications of these languages readily suggesting themselves, it is possible that one of these is more convenient than others in a specific context or in view of a specific goal. For example, the LBA and Grice frameworks assume a separate process for each response, with a response triggered once its process reaches an absolute threshold. This stopping rule is equally simple regardless of the number of response options. In contrast, the diffusion model uses a relative stopping rule that admits a simpler formulation for the case of two responses (a single process representing the difference between the responses, with two constant thresholds) but that becomes more complex in many-alternative tasks (Bogacz & Gurney, 2007; Draglia, Tartakovsky, & Veeravalli, 1999; Roe, Busemeyer, & Townsend, 2001). As we elaborate in Section 5, consideration of neurophysiological or other process-level data, in addition to behavior, may also lead to advantages of one framework over another, although incorporation of such data requires additional assumptions about how abstract model components map onto physical variables. Convenience in a model (mathematical simplicity, ease of mapping special terms into colloquial or traditional ones) is a legitimate consideration provided one does not confuse it with truth about the system being modeled.

2. UNIVERSALITY OF MODELS OF SPEEDED CHOICE

This article focuses on speeded choice tasks, in which on each trial the subject selects one of n response options. We index variation in stimulus values by s, and variation in other experimental factors, such as speed versus accuracy instructions, by c, generically referring to values of c as (experimental, or observation) conditions. In the paradigms we consider, s always varies from trial to trial, whereas c is constant across a block of trials. Because

of this we also refer to s and c as trial-level and block-level manipulations, leaving open the question of whether it is the blocking scheme or the nature of each experimental factor that determines which theoretical parameters it influences. For each choice of s and c, the subject's behavior can be summarized by a joint distribution over the chosen response, r, and the response time, t. Formally, this (joint) *response-and-time (R&T) distribution* can be defined as

$$G^{s,c}(r,t) = \Pr\left[\text{response} = r \text{ and RT} \leq t | s, c\right]. \tag{2.1}$$

The marginal distributions for the joint R&T distribution determine the response probabilities

$$G^{s,c}(r,\cdot) = \Pr\left[\text{response} = r | s, c\right] = \lim_{t\to\infty} G^{s,c}(r,t) \tag{2.2}$$

and the overall RT distribution

$$G^{s,c}(\cdot,t) = \Pr\left[\text{RT} \leq t | s, c\right] = \sum_{r=1}^{n} G^{s,c}(r,t). \tag{2.3}$$

In the following we tacitly assume that the domain of t is $[0,\infty)$, that is, all finite nonnegative real numbers. The analysis can be easily extended to include negative values for t (i.e., premature responses), but we need not do this as premature responses are rarely observed in the choice RT paradigm.

With little loss of generality, we assume that $G^{s,c}(r,t)$ is differentiable in t, which allows us to define the (joint) *R&T density function*,

$$g^{s,c}(r,t) = \frac{\mathrm{d}G^{s,c}(r,t)}{\mathrm{d}t}. \tag{2.4}$$

Under this assumption, we can also define the *joint hazard function for R&T* as the probability density of response r occurring at time t, conditioned on no response having occurred before t (Marley & Colonius 1992; Townsend 1976):

$$h^{s,c}(r,t) = \begin{cases} \frac{g^{s,c}(r,t)}{1-G^{s,c}(\cdot,t)} & \text{if } t < t^{s,c}_{\max} \\ \\ 0 & \text{if } t \geq t^{s,c}_{\max}, \end{cases} \tag{2.5}$$

where $t^{s,c}_{\max}$ is the maximal RT:

$$t^{s,c}_{\max} = \min\left\{t : G^{s,c}(\cdot,t) = 1\right\}, \tag{2.6}$$

which is understood to be ∞ if $G^{s,c}(\cdot,t) < 1$ for all t.

With reference to the example in the Introduction, the R&T distribution, R&T density function, and joint hazard function are all equivalent representations of the same information. It will be useful in what follows to switch freely among these alternate representations.

2.1. The Grice Framework and Independent Race Models

Consider now the general class of *race models*, in which each response r is associated with a process $R_r^{s,c}(t)$, generally a random process, and a threshold $\theta_r^{s,c} > R_r^{s,c}(0)$, generally a random variable. As indicated by the superscripts, both the processes and the thresholds may in general depend on the stimulus (s) and experimental condition (c). On each trial, the processes race to their respective thresholds. The first process to reach its threshold determines which response is emitted, and the time taken to reach that threshold equals the RT.

For each response r, define the *first-passage time*, $T_r^{s,c}$, as the time it would take $R_r^{s,c}(t)$ to cross $\theta_r^{s,c}$ for the first time, ignoring the other process-threshold pairs:

$$T_r^{s,c} = \min_t \left\{ t : R_r^{s,c}(t) \geq \theta_r^{s,c} \right\}. \tag{2.7}$$

Then the response observed in a given trial, r, is given by

$$r = \arg\min_i \left\{ T_i^{s,c} : i = 1, \ldots, n \right\}, \tag{2.8}$$

where arg min with index i indicates the value of i (here, one of the n responses) for which the first-passage time $T_i^{s,c}$ is shortest. The corresponding RT is given by

$$\mathrm{RT} = \min \left\{ T_i^{s,c} : i = 1, \ldots, n \right\} = T_r^{s,c}. \tag{2.9}$$

This representation is useful for modeling response choice and RT in a way that avoids commitments to the nature of the underlying processes (e.g., Townsend & Altieri, 2012; Townsend, Houpt, & Silbert, 2012).

Any vector of process-threshold pairs

$$M = \left((R_1^{s,c}(t), \theta_1^{s,c}), \ldots, (R_n^{s,c}(t), \theta_n^{s,c}) \right)$$

thus defines a *race model*. Equations 2.8 and 2.9 determine the R&T distribution predicted by M for each stimulus and experimental condition. If this joint distribution is equal to $G^{s,c}$, then we say that M is a *race representation* for $G^{s,c}$. If the crossing times $T_1^{s,c}, \ldots, T_n^{s,c}$ are mutually independent for each stimulus and condition, then we refer to M as an *independent race model*, and we say that M is an *independent race representation* for $G^{s,c}$.

The Grice framework (Grice, 1968, 1972) is a class of race models in which the response processes are all deterministic, and the distribution of thresholds is independent of stimulus and condition. The thresholds $\theta_1, \ldots, \theta_n$ are sampled from this fixed joint distribution before each trial, and the stimulus (with onset at $t = 0$) evokes n deterministic processes racing to their respective thresholds (see Figure 1A). This class of models is seemingly quite

simple: given a distribution for the thresholds, the only flexibility is in the choice of the (deterministic) response process for each response and experimental condition. Nevertheless, it is shown in Dzhafarov (1993) that the Grice framework is universal, meaning that a Grice representation can be constructed for any family of R&T distributions $G^{s,c}(r,t)$. Furthermore, the joint distribution defining the thresholds can be chosen arbitrarily (except for a mild technical constraint given below). This result is formalized as follows (see Dzhafarov, 1993, for proof).

Theorem 1 (Universality of Grice framework; Dzhafarov, 1993)**.** *Let $(\theta_1,\ldots,\theta_n)$ be a vector of random thresholds, with any joint distribution possessing a nonzero density everywhere on its domain $(\inf\theta_1,\sup\theta_1)\times\ldots\times(\inf\theta_n,\sup\theta_n)$.*[2] *Then for any family of R&T distributions $G^{s,c}(r,t)$, there exist deterministic response processes $R_r^{s,c}$, such that*

$$\left(\left(R_1^{s,c}(t),\theta_1\right),\ldots,\left(R_n^{s,c}(t),\theta_n\right)\right)$$

is a race representation for $G^{s,c}(r,t)$.

In other words, if no restrictions are placed on the response processes (e.g., requiring them to come from some parametric family), then the Grice framework is a universally applicable modeling language rather than an empirically falsifiable model. Moreover, because the threshold distribution is arbitrary, one can fix this distribution in advance and still have a universal modeling language.

If the thresholds are chosen to be mutually independent (making the Grice model an independent race model), then the Grice representation given by Theorem 1 can be analytically derived, as shown in Theorem 2.

Theorem 2 (Grice framework with independent thresholds; Dzhafarov, 1993)**.** *Let $\theta_1,\ldots,\theta_n$ in Theorem 1 be mutually independent, and let $h^{s,c}(r,t)$ be the joint hazard function associated with $G^{s,c}(r,t)$. Then the response processes in the Grice representation of $G^{s,c}(r,t)$ are given by*

$$R_r^{s,c}(t)=\Theta_r^{-1}\left(\Pr\left[T_r^{s,c}\le t\right]\right), \tag{2.10}$$

for all $t<t_{\max}^{s,c}$, where Θ_r is the (strictly increasing) cumulative distribution function for θ_r, and $T_r^{s,c}$ is the model's first-passage time, with distribution function satisfying

$$\Pr\left[T_r^{s,c}\le t\right]=1-\exp\left(-\int_0^t h^{s,c}(r,\tau)\,\mathrm{d}\tau\right) \tag{2.11}$$

[2] This condition is unnecessarily strong, but sufficiently innocuous for present purposes. In Dzhafarov (1993) the theorem is proved under weaker constraints. Given a random variable x (here, one of θ_i), $\inf x$ and $\sup x$ denote the lower and upper boundaries for the set of this variable's values, these boundaries not being necessarily included in this set.

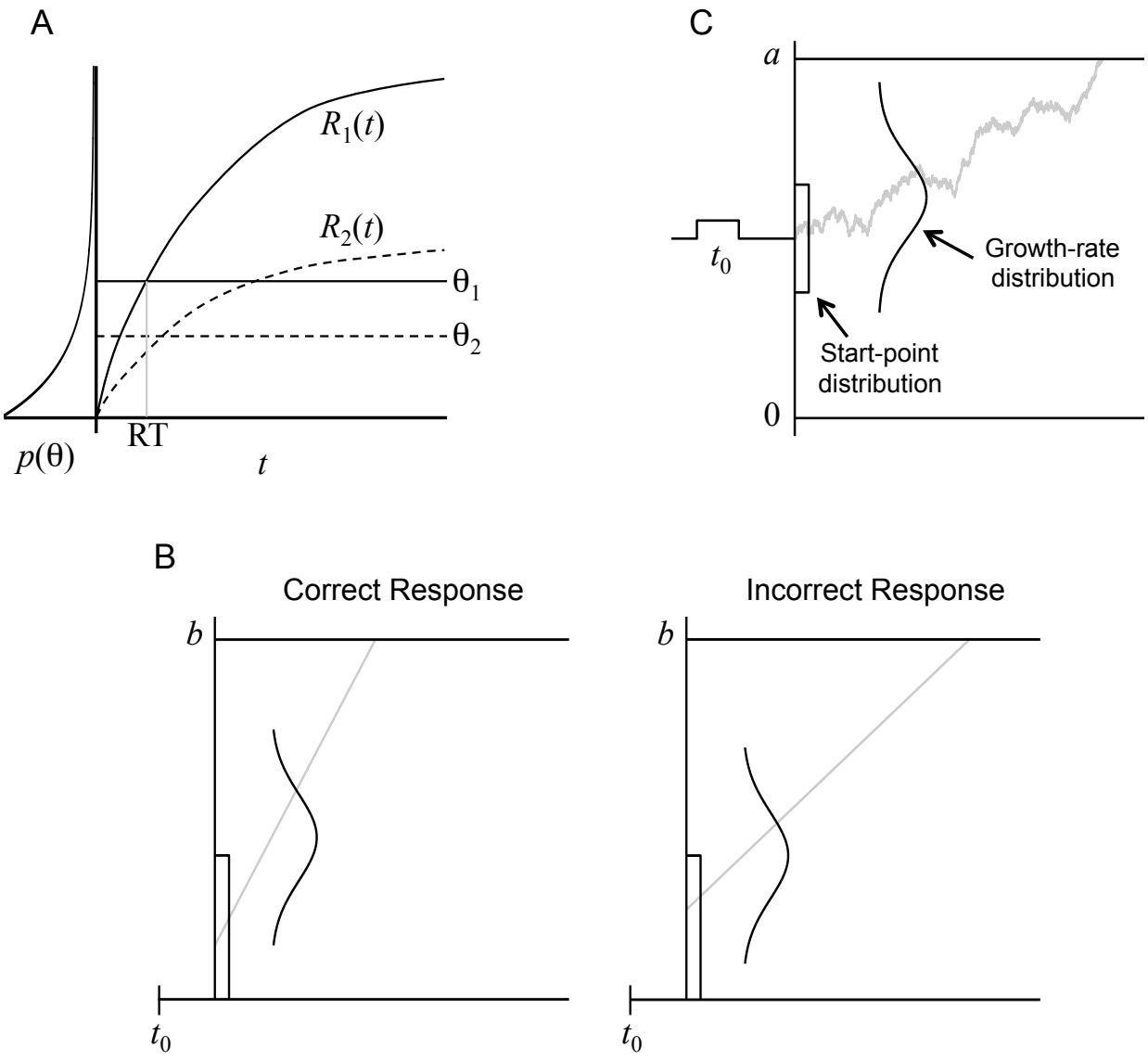

Figure 1: Illustrations of three modeling frameworks for two-alternative speeded choice. A: Grice modeling framework. Before stimulus onset, the thresholds (θ_1 and θ_2) are probabilistically sampled for each response. The threshold distributions can be arbitrarily chosen, and here they are taken to be independently and identically distributed, according to the exponential density function $p(\theta)$ shown at left. Stimulus onset (at time $t = 0$) triggers deterministic response processes $R_1(t)$ and $R_2(t)$. The first process to reach its threshold (here, R_1) determines the response and the response time (RT). B: Linear ballistic accumulator (LBA). Response processes are linear, racing to a common threshold (b). The growth rate and starting point of each process are sampled from Gaussian and uniform distributions, respectively. Nondecision time (t_0) is added to the time of the winning process to determine RT. C: Diffusion model. A single stochastic process evolves until it reaches either decision threshold (0 for response 1 or a for response 2). The growth rate and starting point are sampled from Gaussian and uniform distributions, respectively. Nondecision time t_0 is sampled from a uniform distribution.

for all $t < t^{s,c}_{\max}$.

Equation 2.11 is significant because it turns out to be a necessary and sufficient condition for any independent race representation (not just independent-threshold Grice representations), as stated in Theorem 3.

Theorem 3 (Universality of independent race models)**.** *Let* $G^{s,c}(r,t)$ *be a family of R&T distributions, with associated joint hazard functions* $h^{s,c}(r,t)$*. Let*

$$M = \left(\left(R^{s,c}_1(t),\theta_1\right),\ldots,\left(R^{s,c}_n(t),\theta_n\right)\right)$$

be an independent race model, with first-passage times $(T^{s,c}_1,\ldots,T^{s,c}_n)$*. Then* M *is an independent race representation for* $G^{s,c}$ *if and only if Equation 2.11 holds for* $t < t^{s,c}_{\max}$.

If one disregards minor technical details and logical subtleties in formulation, this theorem was proved by Marley & Colonius (1992) and, in a different context, by Townsend (1976). For completeness (and due to the details and subtleties just mentioned), we provide a proof in Appendix A. The proof hinges on the fact that in any independent race model representing $G^{s,c}$, the joint hazard function $h^{s,c}(r,t)$ coincides with the individual hazard functions $\nu^{s,c}_r(t)$ for the first-passage times $T^{s,c}_r$:

$$h^{s,c}(r,t) = \nu^{s,c}_r(t)\,. \tag{2.12}$$

Theorem 3 implies that the class of independent race models is not a falsifiable model but a universal modeling language. As long as one can define for each response a process-threshold pair with the right distribution of first-passage times, any R&T distribution can be perfectly fit. Thus the requirement of independent first-passage times, taken alone, imparts no restriction on model predictions. Although there are models of speeded choice in which the first-passage times are nonindependent (because of interactions among the processes, as in Usher & McClelland, 2001, or mutual interdependence of thresholds, as in Dzhafarov, 1993), the question can never be empirically decided, at least from RT and choice data alone. Furthermore, any model class within the family of independent race models only has predictive power insofar as the individual processes are restricted in their first-passage distributions. This conclusion is used below to prove universality of the LBA when its growth-rate distributions are unconstrained (see Section 2.3).

As stated in Marley and Colonius (1992), an important caveat to representing $G^{s,c}(r,t)$ through the competing random times $T^{s,c}_1,\ldots,T^{s,c}_n$ is that Equation 2.11 may not define a proper probability distribution for some values of r, because $\Pr\left[T^{s,c}_r \le t\right]$ may be bounded by some value less than 1 for all values of t. In other words, the first-passage times $T^{s,c}_r$ must be viewed as distributed on the extended set of nonnegative reals, $[0,\infty]$, with infinity a possible value for $T^{s,c}_r$, attainable with a nonzero probability. Such random variables are called

incomplete (Dzhafarov, 1993), *improper* (Marley & Colonius, 1992), or *defective* (Feller, 1968). Although completeness of first-passage times has been taken as a defining requirement in some past analyses of race models (Townsend, 1976), we hold that incompleteness of individual first-passage times is not problematic because it has a natural cause and interpretation: a threshold can be set too high for a process ever to cross it (Dzhafarov, 1993). As long as $T_r^{s,c}$ is complete for at least one r, the observed RT, equal to $\min\{T_1^{s,c}, \ldots, T_n^{s,c}\}$, will be complete as well (i.e., a response will occur on every trial).

Allowing for some response processes to have incomplete first-passage distributions is necessary for Theorem 3 to hold. This is because there exist R&T distributions that cannot be represented by an independent race model if all processes are required eventually to reach threshold with probability 1. This can be seen by constructing a model in which some first-passage distributions are incomplete, and showing that the R&T distribution predicted by this model has no other independent race representation in which the first-passage distributions are all complete. This result (proved in Appendix A) will be useful in the next two subsections in determining conditions for certain model classes to be universal.

Theorem 4 (Incomplete termination times)**.** *For any subset* $\{r_1, \ldots, r_m\}$ *of the response options* $\{1, \ldots, n\}$*, where* $0 \leq m < n$*, there exists an R&T distribution* $G(r,t)$ *such that any independent race model generating* G *has first-passage distributions that are incomplete for* $r_1, \ldots, r_m$ *and that are complete for the remaining responses.*

Note that the choice of the subset $\{r_1, \ldots, r_m\}$ in this theorem can be different for different conditions and stimuli. Some models (including Grice et al., 1982; McClelland, 1979; and the LBA, as discussed below) allow for a *global incompleteness* of RT, a positive probability that no response is given in a trial. Depending on ramifications, this property may be viewed as a construction deficiency or as the reflection of a true empirical phenomenon. In the latter case, we may allow for $m = n$ in Theorem 4.

2.2. Universal Variants of the Grice Framework

Given a Grice model $M = ((R_1^{s,c}(t), \theta_1), \ldots, (R_n^{s,c}(t), \theta_n))$ (with deterministic response processes and stochastic thresholds), one can define equivalent models in which the thresholds are deterministic and all stochasticity resides in the response processes. The new models are equivalent to M in the sense of accounting for (i.e., generating) precisely the same R&T distributions.

In view of the discussion of the LBA and diffusion models below, it is especially interesting to consider the immediately obvious equivalents of a Grice model in which the thresholds are deterministic, and each response process has a fixed shape (for any given stimulus and

condition), but the growth rates or starting points vary randomly across trials (see Figure 2). For example, let γ be any strictly increasing nonnegative function, define a random growth rate for each response process by

$$k_r = \frac{1}{\gamma(\theta_r)}$$

(under our assumptions, $\Pr[k_r < \infty] = 1$), and define the (deterministic) shape of each process by

$$\tilde{R}_r^{s,c}(t) = \gamma\left(R_r^{s,c}(t)\right).$$

Then the model defined by

$$M' = \left(\left(k_1\tilde{R}_1^{s,c}(t), \tilde{\theta}_1 = 1\right), \ldots, \left(k_n\tilde{R}_n^{s,c}(t), \tilde{\theta}_n = 1\right)\right) \tag{2.13}$$

is equivalent to the Grice model M. The equivalence follows from the fact that

$$k_r\tilde{R}_r^{s,c}(t) \geq 1 \text{ if and only if } R_r^{s,c}(t) \geq \theta_r.$$

Analogously, let z_r be a random starting point (i.e., offset) for process r, defined by

$$z_r = 1 - \gamma(\theta_r).$$

Then the model defined by

$$M'' = \left(\left(z_1 + \tilde{R}_1^{s,c}(t), \tilde{\theta}_1 = 1\right), \ldots, \left(z_n + \tilde{R}_n^{s,c}(t), \tilde{\theta}_n = 1\right)\right), \tag{2.14}$$

is also equivalent to the Grice model M. The equivalence here follows from

$$z_r + \tilde{R}_r^{s,c}(t) \geq 1 \text{ if and only if } R_r^{s,c}(t) \geq \theta_r.$$

Because of the universality of the Grice framework, Equations 2.13 and 2.14 also define universal modeling languages, provided the functions $\tilde{R}_r^{s,c}(t)$ are treated as free in fitting R&T distributions. Moreover, the choice of the distribution for k_r or z_r is arbitrary (because the distribution of θ_r in the Grice representation is), except for the requirement of nonvanishing density on an appropriately chosen open domain. We summarize these observations in the following two theorems.[3]

Theorem 5. *Let $k_1, \ldots, k_n$ be stochastically independent nonnegative random variables, each k_r having a nonzero density on its domain $(\inf k_r, \sup k_r)$. Then for any family of R&T*

[3] Theorems 5 and 6 respectively correspond to Lemmas 1.3.2 and 1.3.1 in Dzhafarov (1993). Unfortunately, it is erroneously stated in those lemmas that k_r and z_r (using our present notation) can be defined as random processes $k_r(t)$ and $z_r(t)$ rather than within-trial constants.

distributions $G^{s,c}(r,t)$*, there exist deterministic, nondecreasing, and nonnegative functions* $\tilde{R}_r^{s,c}(t)$ *such that*

$$\left(\left(k_1\tilde{R}_1^{s,c}(t),1\right),\ldots,\left(k_n\tilde{R}_n^{s,c}(t),1\right)\right)$$

is an independent race representation for $G^{s,c}(r,t)$*.*

Theorem 6. *Let* $z_1,\ldots,z_n$ *be stochastically independent random variables, each* z_r *having a nonzero density on its domain* $(\inf z_r, \sup z_r)$*, with* $\max_r \sup z_r \leq 1$*. Then for any family of R&T distributions* $G^{s,c}(r,t)$ *there exist deterministic, nondecreasing, and nonnegative functions* $\tilde{R}_r^{s,c}(t)$*, such that*

$$\left(\left(z_1+\tilde{R}_1^{s,c}(t),1\right),\ldots,\left(z_n+\tilde{R}_n^{s,c}(t),1\right)\right)$$

is an independent race representation for $G^{s,c}(r,t)$*.*

These straightforward results show that the essence of the Grice framework is not in the assumption of deterministic processes and random thresholds. Rather, its essence is in how the difference between thresholds and processes, $\theta_r - R_r^{s,c}(t)$, is separated into deterministic and random components. In the Grice representation and in the equivalent representations of Theorems 5 and 6, the deterministic components (i.e., the shapes of the response processes) are stimulus- and condition-dependent and free to vary in fitting data, whereas the random parts of the models (thresholds, growth rates, or starting points, respectively) have fixed (and essentially arbitrary) distributions. In contrast, in the LBA modeling language considered next, the randomly varying growth rates are stimulus-dependent and free to vary, whereas the shape of the processes is fixed (chosen to be linear).

2.3. Universality of the Linear Ballistic Accumulator

The LBA is a stochastic race model that has been recently proposed as a model of human choice RT (Brown & Heathcote, 2008; Donkin, Averell, Brown, & Heathcote, 2009; Donkin, Brown, Heathcote, & Wagenmakers, 2011). One advantage that has been claimed for the LBA is its simplicity, relative to other models of choice RT (e.g., Ratcliff, 1978; Usher & McClelland, 2001). Nevertheless, we demonstrate that, aside from a selective influence assumption and two distributional assumptions that have received little attention or justification, the LBA is universal, in that it can reproduce any possible R&T distribution. Therefore the predictive constraints in the LBA derive entirely from these assumptions.

Each response process $R_r^{s,c}(t)$ in the LBA, once initiated at the start of a trial, deterministically follows a simple linear function,

$$R_r^{s,c}(t) = z_r^c + k_r^s t, \tag{2.15}$$

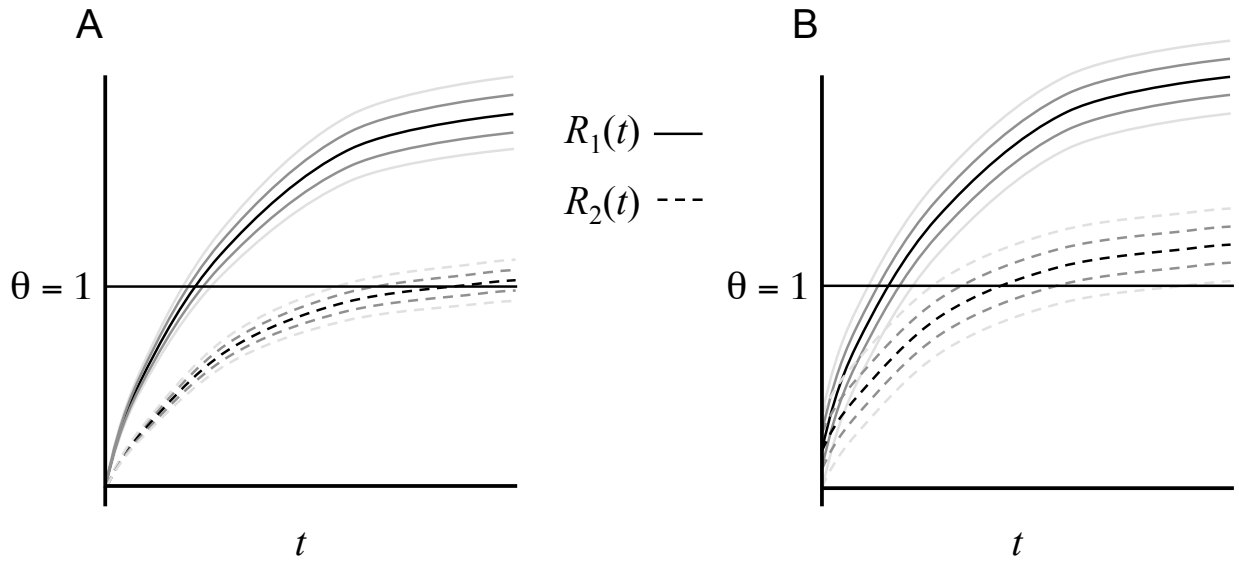

Figure 2: Illustration of Grice-equivalent frameworks of Theorems 5 and 6, based on the Grice model diagrammed in Figure 1A. In both models, the thresholds and the shapes of the response processes are fixed. Random variability resides in the growth rates (A) or starting points (B).

and the processes race to a common deterministic threshold, $\theta_r^{s,c} = b^c$. Stochasticity in the LBA comes from two sources: variability in the starting points of the processes (z_r^c) and variability in their growth rates (k_r^s). Both of these variables are sampled independently on each trial for each response process. Thus, the LBA is a subclass of independent race models.

The starting points are sampled from a uniform distribution ranging from 0 to A^c, where A^c is a free parameter constrained to be less than the common threshold b^c:

$$z_r^c \sim U(0, A^c), \tag{2.16}$$

where $\sim$ stands for "is distributed as." Note that this start-point variability is equivalent to independent variability in the threshold for each response (i.e., setting $\theta_r^c = b^c - z_r^c$ and constraining the starting points to zero).

The growth rates are sampled from Gaussian distributions, with a common fixed variance and means that vary across responses and stimuli:

$$k_r^s \sim N\left(v_r^s, \eta^2\right). \tag{2.17}$$

The choice of Gaussian distribution implies there is a nonzero probability of a negative growth rate, in which case the process will never terminate. The first-passage distributions are thus incomplete random variables for all responses. This also implies global incompleteness (i.e., a nonzero probability that no response will be given), although Brown and

Table 1: LBA model parameters

Parameter	Meaning	Associated Distribution
b^c	Common threshold for all response processes	
A^c	Upper limit on range of starting points	Uniform for starting point z_r^c
v_r^s	Mean growth rate of response process r	Gaussian for growth rate k_r^s
η	Standard deviation of growth rate for all response processes	Gaussian for growth rate k_r^s
t_0	Nondecision time	

Heathcote (2008) report that, in fits to empirical data, the probability of all growth rates being negative is typically less than 0.5%.

Finally, the LBA assumes a nondecision time, t_0, which is constant across trials and conditions and is added to the first-passage time of the winning process to determine total RT on each trial. Table 1 summarizes the parameters of the LBA model. As indicated by the superscripts, the threshold and start-point variability are assumed to depend on condition (e.g., a higher threshold for accuracy vs. speed instructions), and the mean growth rate is assumed to depend on the stimulus (e.g., a higher mean for the correct response for easier stimuli).

Random variability in starting points and in growth rates is critical to the LBA's ability to match empirical data (Brown & Heathcote, 2008). Start-point variability allows the model to predict rapid incorrect responses, whereas growth-rate variability allows it to predict slow incorrect responses. The first mechanism is more important in tasks with short RTs (e.g., under high discriminability or speed instructions), where the starting points can be close to the threshold, whereas the second is more important in tasks with long RTs (e.g., under low discriminability or accuracy instructions). Thus, these mechanisms can together produce error RTs that are shorter (on average) than correct RTs in easy tasks but that are longer on difficult tasks, an empirical pattern that has proven especially challenging for models of speeded choice (Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999).

Despite the centrality of these two sources of variability to the LBA model's predictions, the specific assumptions regarding the shapes of the distributions—uniform for starting points, Gaussian for growth rates—has received little discussion or theoretical justification. Indeed, Brown and Heathcote (2008, p. 160) state that they "chose the normal distribution for practical reasons, because it is both tractable and conventional." The normal distribution facilitates analytic derivations of the LBA's predictions, and it has been commonly adopted for the distribution of growth rates in other models as well (e.g., Ratcliff, 1978). The same considerations apply to the uniform distribution for starting points.

If these distributional assumptions are only a matter of convenience and tradition, then

they should not be considered a critical part of the psychological theory. Therefore, we define the *general LBA* (gLBA) as the LBA model without these distributional constraints. We also set aside the selective influence constraint on the growth rate, allowing it to depend on block-level manipulations in addition to the stimulus (see Section 3 for further discussion). Hence we denote the gLBA's growth rates $k_r^{s,c}$.

The gLBA is a significantly broader model class than the LBA, because it replaces the LBA's A^c, v_r^s, and η parameters with complete (nonparametric) flexibility in the start-point and growth-rate distributions. Nevertheless, the gLBA maintains all of the structural commitments of the LBA: linear response processes, a common fixed threshold, and independent random variability in starting values and growth rates for all processes. Building on the results of Section 2.1, it is easy to show that the gLBA can match any family of R&T distributions. In fact, start-point variability is unnecessary for this result; random variability in growth rates alone is sufficient.

Theorem 7 (Universality of general LBA)**.** *Let $G^{s,c}(r,t)$ be any family of R&T distributions. Then one can fix starting points z_r^c at zero and threshold b^c at an arbitrary positive value b for all c, and find a set of random growth rates $k_r^{s,c}$ under which the gLBA perfectly reproduces $G^{s,c}$. Furthermore, this can be done for any choice of t_0 less than the minimal possible RT (i.e., with $G^{s,c}(\cdot,t_0)=0$ for all s and c).*

The proof of this result (given in Appendix A) is based on the fact that the first-passage time of each response process is a function of its growth rate:

$$T_r^{s,c} = \begin{cases} \frac{b}{k_r^{s,c}} + t_0 & \text{if } k_r^{s,c} > 0 \\ \infty & \text{if } k_r^{s,c} \leq 0. \end{cases} \tag{2.18}$$

It is thus straightforward to define a distribution on $k_r^{s,c}$ so that $T_r^{s,c}$ takes on any desired distribution (bounded below by t_0). In particular, the growth-rate distributions can be chosen so that the first-passage distributions satisfy Equation 2.11, which by Theorem 3 implies the model will reproduce $G^{s,c}$. In other words, flexibility of growth-rate distributions gives the gLBA complete flexibility in its first-passage distributions, and from Theorem 3 this is sufficient to make the model universal.

This universality property can easily be extended to cover variants of the ballistic accumulator framework in which the linear form of the LBA's response processes is replaced by other forms (e.g., Brown & Heathcote, 2005). Specifically, let $L_r(t)$ be any family of strictly increasing, continuous, deterministic functions with $L_r(0)=0$, and define a model in which response processes evolve according to

$$R_r^{s,c}(t) = k_r^{s,c} L_r(t),$$

governed by random growth rates $k_r^{s,c}$ and racing to a common threshold b. Then the growth-rate distribution for each response can again be chosen to generate any desired first-passage distribution. Therefore any model of this form is universal, for any *a priori* choices of L_r and b. We state this result formally as follows (see Appendix A for proof).

Theorem 8 (Universal ballistic models with random growth rates)**.** *Let b be any positive number, and $L_1(t), \ldots, L_n(t)$ be strictly increasing continuous functions with $L_r(0) = 0$. Then for any family of R&T distributions $G^{s,c}(r,t)$, there exist random variables $k_1^{s,c}, \ldots, k_n^{s,c}$ such that*

$$((k_1^{s,c} L_1(t), b), \ldots, (k_n^{s,c} L_n(t), b))$$

is an independent race representation for $G^{s,c}(r,t)$.

The universality of the gLBA can be seen as a special case of this result, by taking $L_r(t) = \max\{0, t - t_0\}$.

Finally, instead of assuming random variability in the growth rates (with fixed starting points), one can assume random variability in the starting points (with fixed growth rates). This framework is also universal, subject to the caveat below regarding the shapes of the response processes.

Theorem 9 (Universal ballistic models with random starting points)**.** *Let b be any positive number, and let $L_1(t), \ldots, L_n(t)$ be strictly increasing continuous functions with $L_r(0) = 0$ and $\lim_{t\to\infty} L_r(t) \leq b$. Then for any family of R&T distributions $G^{s,c}(r,t)$, there exist random variables $z_1^{s,c}, \ldots, z_n^{s,c}$ with $0 \leq z_r^{s,c} < b$ such that*

$$((L_1(t) + z_1^{s,c}, b), \ldots, (L_n(t) + z_n^{s,c}, b))$$

is an independent race representation for $G^{s,c}(r,t)$.

The proof of Theorem 9 (given in Appendix A), relies on the same principle as Theorems 7 and 8: each response process can be made to take on any desired first-passage distribution, in this case through appropriate choice of the start-point distribution. Note that the condition $\lim_{t\to\infty} L_r(t) \leq b$ is critical for this construction. If this constraint does not hold — that is, $L_r(t) > b$ for some r and t — then the response process $L_r(t) + z_r^{s,c}$ will cross b at some finite time regardless of the starting point $z_r^{s,c}$. Consequently, the first-passage time $T_r^{s,c}$ will be a complete random variable, no matter how $z_r^{s,c}$ is distributed.[4] Thus by Theorem 4, the representation will not be universally applicable. For example, the LBA (with linear

[4] If one were to allow negative starting points, the condition in question would change to $L_r(t)$ being bounded from above (not necessarily by b). Other straightforward modifications of this theorem and of the previous one are mentioned in Appendix A.

response processes) cannot be made universal solely through start-point variability, because in that model $\lim_{t\to\infty} L_r(t) = \infty$ for all responses. However, other ballistic accumulator models can be made universal through start-point variability alone, provided their response processes have finite limits (e.g., $L_r(t) = 1 - e^{-t}$).[5]

As the proofs of the preceding theorems show, universality of the gLBA and its variants is a straightforward mathematical fact. However, its implications for the standard LBA seem to have been overlooked. Specifically, this result implies that the predictive power of the standard LBA lies in its assumptions regarding growth-rate and start-point distributions (which have heretofore been treated as implementation details), together with the selective influence assumptions regarding how model parameters can vary across stimuli and experimental conditions. We return to the implications of these latter assumptions in Sections 3 and 4.

2.4. Universality of Diffusion Models

We now consider a different class of RT models, in which response processes exhibit within-trial stochasticity, derived from Brownian motion, in addition to between-trial stochasticity in starting points and growth rates. Two such models are considered here: the Wiener diffusion model (Ratcliff, 1978) and the OU model (Busemeyer & Townsend, 1993), which we collectively refer to as diffusion models. In addition to the presence of diffusion, the diffusion models considered here (for binary choice tasks) can be contrasted with the LBA model in that they assume a single response process that can evolve in positive or negative directions until it reaches either an upper or a lower threshold. The model can be equivalently formulated with a separate process for each response, but under that representation the stopping rule is based on their difference rather than on the value of each process taken separately (as in the LBA).

Specifications for the diffusion model are taken from Ratcliff and Smith (2004), who treat the OU model as a generalization of the Wiener model. The response process in any

[5] The nonlinear ballistic accumulator model proposed by Brown and Heathcote (2005) has response processes with finite limits, but it cannot be made universal through start-point variability because starting points there do not act additively and do not affect the processes' asymptotes. However, universality through random starting points would hold under a minor modification to the model's decay term, replacing $-\beta R_r^{s,c}(t)$ with $-\beta\left(R_r^{s,c}(t) - z_r^{s,c}\right)$ so that decay is toward the starting point instead of zero (see Brown & Heathcote's Equation 1). We consider this change an implementational detail with no substantive bearing on the model's theoretical content. Moreover, Ratcliff and Smith (2004) advocate exactly this modification to the decay term in the OU model (which we adopt here; see Equation 2.19), because it enables the effects of start-point variability to persist rather than be lost to decay.

experimental condition, $R^{s,c}(t)$, is a stochastic process defined by a random starting point, z^c, a random growth rate, k^s (also referred to as drift rate), and a diffusion rate, σ^2. The OU model includes a decay term with parameter β, which is set to 0 for the Wiener model. The dynamics on each trial can be defined by the following stochastic differential equation:

$$\mathrm{d}R^{s,c}(t) = \left(k^s - \beta\left(R^{s,c}(t) - z^c\right)\right)\mathrm{d}t + \sigma \mathrm{d}B(t) \tag{2.19}$$

Here, $B(t)$ represents a Brownian motion process. The diffusion parameter, σ, is generally treated as fixed, because changing its value can always be nullified by rescaling other model parameters.[6] On each trial, an accumulation process starts, at the time of stimulus onset, at $R^{s,c}(0) = z^c$. The process terminates when it crosses either a lower threshold set at 0, or an upper threshold, a^c, respectively corresponding to responses $r = 1$ and $r = 2$. Which threshold is crossed first and the time of the first passage respectively determine the response and RT.

Additional model parameters determine random variability across trials, following the same assumptions as adopted in the LBA model (Brown & Heathcote, 2008). The starting point of the diffusion process, z^c, is drawn from a uniform distribution with mean $\bar{z}^c$, which is a free parameter when modeling response bias and is otherwise fixed at $a^c/2$. The range of this distribution, which we denote δ_z, is constrained to be less than $\min\left\{2\bar{z}^c, 2\left(a^c - \bar{z}^c\right)\right\}$, so that z^c always lies between 0 and a^c. The growth rate, k^s, is drawn from a normal distribution with mean v^s and variance η^2. In addition, the nondecision time, t_0, is drawn from a uniform distribution with mean T_er and range δ_t (where $\delta_t \leq 2T_\mathrm{er}$). The nondecision time on each trial is added to the first-passage time of the diffusion process to predict total RT. Table 2 summarizes the parameters of the diffusion model. As the notation indicates, the diffusion model incorporates the selective influence assumptions that trial-level variation in stimuli affects only v^s, whereas block-level manipulations affect only a^c and $\bar{z}^c$ (and δ_z in some implementations; e.g., Ratcliff & Rouder, 1998).

The diffusion model, and discrete-time (random walk) variants, were originally motivated by the sequential likelihood-ratio test, which performs optimal Bayesian inference over a stream of noisy data (Stone, 1960). Under this interpretation, the input stream represents

[6] However, fixing this parameter while fitting multiple experimental conditions, with constraints linking other parameters across conditions, imposes unintended limitations on the model (Donkin, Brown, & Heathcote, 2009). With the diffusion model, this constraint leads to an additional conceptual problem. Although it is often assumed that the only parameter that can vary as a function of stimulus within a block of trials is the mean growth rate, the standard process interpretation of the model as sequential sampling implies that the diffusion rate should also depend on the stimulus. Indeed, there are variants of the diffusion model that predict such a dependence (Busemeyer & Townsend, 1993). Fixing the diffusion rate across stimuli thus removes flexibility that theoretically should be present.

Table 2: Diffusion model parameters

Parameter	Meaning	Associated Distribution
β	Decay rate; set to 0 for Wiener model	
a^c	Threshold separation	
σ	Diffusion rate	
v^s	Mean growth rate	Gaussian for growth rate k^s
η	Standard deviation of growth-rate distribution	
$\bar{z}^c$	Mean starting point	Uniform for starting point z^c
δ_z	Range of start-point distribution	
T_{er}	Mean nondecision time	Uniform for nondecision time t_0
δ_t	Range of nondecision distribution	

the momentary likelihood ratio of sensory evidence, assumed to be white noise with a non-zero mean. The goal is to infer whether the mean is positive or negative, corresponding to one or the other correct response. The diffusion process is the integral of the input stream over time (or sum, in the discrete-time case), and responding once this process reaches a threshold is equivalent to waiting until the Bayesian posterior for one response reaches some predetermined threshold probability. This pure normative model is a special case (with no decay and no variability in growth rates or starting points) of the full diffusion model defined here. However, it makes empirically incorrect predictions, such as equality of mean RTs for correct and error responses (Laming, 1968; Ratcliff, 1978). Including random variation in starting points and growth rates has been found to solve these problems and enables the diffusion model to provide excellent fits to data (Ratcliff et al., 1999; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004). However, an important question is whether these fits are at least to some extent due to the model's architecture, or they are solely consequences of the flexibility afforded by the inclusion of this random variability.

It turns out the answer is the latter: the only predictive constraints of the diffusion model — with random variability in growth-rate, start-point, and nondecision time included — come from the technical assumptions about the shapes of these distributions, together with the selective influence assumption that growth rate does not depend on conditions fixed within each block of trials. As with the LBA results presented in the previous section, if these assumptions are removed, the diffusion model becomes universally applicable. Once again, the model is universal from growth-rate variability alone; the other sources of variability are unnecessary.

We refer to the diffusion model without restrictions on the growth-rate distribution as the *general diffusion model* (gDM). As with the gLBA, we allow the growth-rate distribution to

depend on block-level manipulations in addition to stimuli (see Section 3 for more discussion of this assumption). All other model parameters are independent of the stimulus.

To see that the gDM is universal, consider the special case with no diffusion ($\sigma = 0$), no decay ($\beta = 0$), no start-point variability ($\delta_z = 0$), no non-decision variability ($\delta_t = 0$), fixed $a^c = a$, and no response bias ($\bar{z} = a/2$). Under these assumptions, the accumulation process always follows a linear path starting at $a/2$. The only variability in the model comes from variation across trials in the growth rate, $k^{s,c}$. Under this simplified model, the response and RT are determined according to

$$r = \begin{cases} 1 & \text{if } k^{s,c} < 0 \\ 2 & \text{if } k^{s,c} > 0 \end{cases} \tag{2.20}$$

and

$$\mathrm{RT} = T_{\mathrm{er}} + \frac{a}{2\left|k^{s,c}\right|}. \tag{2.21}$$

It is easy to see that appropriate specification of the growth-rate distribution can generate any desired R&T distribution. Equations 2.20 and 2.21 show there is a one-to-one mapping between $k^{s,c}$ and the pair (r, RT). Therefore, any joint distribution on r and RT can be directly translated to a distribution on $k^{s,c}$.[7] This correspondence leads to a universality result for the general Wiener diffusion model (see Appendix A for proof).

Theorem 10 (Universality of general Wiener diffusion model)**.** *The gDM can reproduce any family of R&T distributions over two responses. Moreover, this is possible without diffusion, decay, random variability in starting points or nondecision time, or response bias ($\sigma = \beta = \delta_z = \delta_t = 0$, $\bar{z} = a/2$), for any fixed $a^c = a$, and for any T_{er} with $G^{s,c}(\cdot, T_{\mathrm{er}}) = 0$ for all s and c.*

Similar to the result of Theorem 8 above, linearity of the response process is not essential for universality of the diffusion model. The same conclusion holds when the response process takes the form $k^{s,c}L(t) + z^c$, where $L(t)$ is any continuous, strictly increasing function with $L(0) = 0$. In particular, the OU model follows this form (again, in the limiting case of no diffusion), for any value of the decay rate, β. Therefore, the OU model is universal for any predetermined value of β (see Appendix A for proof).

Theorem 11 (Universality of general Ornstein-Uhlenbeck model)**.** *The gDM can reproduce any family of R&T distributions over two responses, for any predetermined choice of β.*

[7] The case $k^{s,c} = 0$ corresponds to $t = \infty$, so if global incompleteness is excluded (i.e., a response is given on every trial), then $k^{s,c} = 0$ will be assigned zero probability.

Moreover, this is possible without diffusion, random variability in starting points or nondecision time, or response bias ($\sigma = \delta_z = \delta_t = 0$, $\bar{z} = a/2$), for any fixed $a^c = a$, and for any T_{er} with $G^{s,c}(\cdot, T_{er}) = 0$ for all s and c.

Although these universality results are based on simplified versions of the diffusion model, reintroducing the model's other free parameters (σ, δ_z, δ_t, a^c, and $\bar{z}$) only confers the model additional flexibility — or would, if it were not already universal. One might ask instead how the universality property fares if some of the parameters set to zero in Theorems 10 and 11 are fixed at other values. In particular, if diffusion is taken as a theoretical commitment of the model, then one could require the diffusion rate to be strictly positive ($\sigma > 0$). Such an assumption might be expected to constrain the model's predictions, because of the smoothing effect diffusion has on the predicted RT distributions. In fact, Ratcliff (2013) suggests that stochasticity in the diffusion process "washes out" any effects of growth-rate and starting-point distributions during the course of the trial, so that the choice of these distributions has little effect on model predictions. If this suggestion were correct, it would limit the practical significance of the present universality arguments, in that flexibility in growth-rate distribution might add little flexibility to the model's predictions when $\sigma > 0$.

One response to Ratcliff's (2013) claim is that a model with a fixed positive diffusion rate can be brought arbitrarily close to one without diffusion by rescaling other variables. Specifically, the arbitrary internal scale of the response process can be freely changed, by multiplying σ, k^s, z^c, and a^c by any number $x > 0$ (substitute $x \cdot R^{s,c}(t)$ for $R^{s,c}(t)$ in Equation 2.19, and the equation holds under these rescaled parameters). Consider an arbitrary R&T distribution $G(r,t)$ and a gDM model with $\sigma = 0$ reproducing that distribution, as provided by Theorem 10 or 11. By a continuity argument, $G(r,t)$ can also be approximated arbitrarily well by taking σ to be positive but sufficiently close to zero. We can then scale up the parameters of the model so that σ is equal to any desired value. Therefore, absent any constraints on these other variables, the diffusion model can approximate any R&T distribution to arbitrary precision, for any *a priori* choice of σ. This conclusion is summarized in the following theorem (see Appendix A for proof).

Theorem 12 (Universality with strictly positive diffusion)**.** *Let $G(r,t)$ be any R&T distribution admitting a well-defined density $g(r,t)$. Choose any values of $\sigma > 0$ and β, and let ϵ be any positive constant. Then there exists a gDM model with those values of σ and β for which the predicted R&T distribution $\tilde{G}(r,t)$ satisfies*

$$\max_{r,t} \left| G(r,t) - \tilde{G}(r,t) \right| < \epsilon$$

for all r and t.

One might argue that Theorem 12 allows the model's parameter values to be "unreasonably large." However, there is no *a priori* basis for what parameter values are or are not reasonable. There are values that have typically been found in fits of the Gaussian-drift model, but if the model had found dramatically different values, then those would be the values considered typical and reasonable. Moreover, the fact that the values found have been somewhat consistent across applications is not evidence that those values are correct. If the model is misspecified, then its parameters will be systematically and consistently biased.

Considerations of parameter values depend on assumptions about both the scale of the model's internal evidence dimension and the scaling of time. This dependence can be made explicit by defining a time constant $\theta = a^2/\sigma^2$ (the time it takes for the standard deviation $\sigma\sqrt{t}$ of the diffusion process to equal the threshold separation a) and then reparameterizing the model with a new response process $\breve{R} = R/a$ and time scale $\tau = t/\theta$. Under this reparameterization, the starting point, growth rate, and decay rate become $\breve{z} = z/a$, $\breve{k} = \theta k/a$, and $\breve{\beta} = \theta\beta$. These transformations remove the scaling of the internal evidence dimension and the scaling of time. The model's dynamics are now defined by

$$\mathrm{d}\breve{R}(\tau) \;=\; \left(\breve{k} - \breve{\beta}\left(\breve{R}(\tau) - \breve{z}\right)\right)\mathrm{d}\tau + \mathrm{d}B(\tau) \tag{2.22}$$

with

$$t = \theta\tau. \tag{2.23}$$

Equation 2.22 matches the original dynamics of Equation 2.19 except that the diffusion rate and threshold separation are no longer free parameters (both are fixed at unity), and the variables $\breve{R}$, $\breve{z}$, $\breve{k}$, $\breve{\beta}$, and τ are all dimensionless. The model's only free parameters are the starting point and growth rate (both random variables) and the time constant θ that scales the model's predictions onto physical time.

The utility of this reparameterization is that it shows more directly how diffusion constrains the model's predictions. Specifically, diffusion is appreciably constraining only for larger values of τ, whereas for smaller values the model is arbitrarily flexible. The situation is illustrated in Figure 3. Figure 3A shows the possible predictions of the model in the case of $\breve{z} = 1/2$ and $\breve{\beta} = 0$ (no response bias, start-point variability, or decay). The two families of curves represent possible RT distributions for the two responses, each curve corresponding to a single value of the growth rate $\breve{k}$. Growth rates closer to zero produce the broader curves to the right, and they also produce significant rates of the less likely response (shown as dashed curves, with each dashed curve corresponding to a solid curve for the opposite response). Larger growth rates—positive or negative—produce the narrower curves to the left, which are more separated and have negligible rates of the alternate response.

Under a free growth-rate distribution, the predictions of the full model are an arbitrary mixture of the curves for individual values of $\breve{k}$. For larger values of τ, the form of this mix-

ture is constrained, both across times (i.e., smoothness in the RT distribution) and between responses. For values of τ approaching zero, the mixture becomes arbitrarily flexible. When confronted with empirical data, such as the distributions shown in Figure 3B (typical data, from Jones, Curran, Mozer, & Wilder, in press), one cannot know to what values of τ they correspond without knowing the time constant θ. If the data correspond to larger τ values (smaller θ), then diffusion appreciably limits the model's flexibility, but if the data correspond to smaller τ values (larger θ) it does not. The time constant θ depends on both the diffusion rate and the threshold separation, neither of which can be objectively determined based on behavioral data. It may be possible to use neural or other process-level data to measure these quantities, but as we caution in Section 5, any such endeavor depends on strong assumptions about how the cognitive model maps onto physical processes.

2.5. An Example

To illustrate the universality results, this section gives an example of the gLBA mimicking the diffusion model. Universality of the gLBA implies it can perfectly match the R&T distribution predicted by the diffusion model (for any values of that model's parameters), under appropriate specification of the gLBA's growth-rate distributions.

Figure 4A shows the predictions of the standard diffusion model for a single stimulus and condition, using typical values of its parameters ($\beta = 0$, $a = .1$, $\sigma = .1$, $v = .15$, $\eta = .15$, $\bar{z} = .05$, $\delta_z = .03$, $\delta_t = 0$; T_{er} arbitrary because it can be matched by t_0 in the gLBA).[8] Each track indicates a sample trajectory of the diffusion process for one hypothetical trial. The curves paralleling the two thresholds indicate the model's predicted R&T distribution. Each curve shows the conditional RT distribution for correct (upper) or incorrect (lower) responses, with total area under the curve indicating response probability. These curves were calculated analytically using methods described in Appendix B.

The diffusion model's predicted R&T distribution was converted to a joint hazard function and then translated into first-passage distributions for an independent race model using Equation 2.11. According to Theorem 3, any independent race model generating these first-passage distributions will perfectly match the diffusion model's predictions. Figure 4B shows an instance of the gLBA derived to satisfy this condition. The curve paralleling the threshold for each response is the first-passage distribution obtained from Equation 2.11. The gLBA can trivially reproduce these (or any other) first-passage distributions through flexibility of its growth-rate distributions. Figure 4C shows the growth-rate distributions

[8] We use seconds as the unit of time whenever referencing numerical values of model parameters. The internal scale of the response process is in arbitrary units.

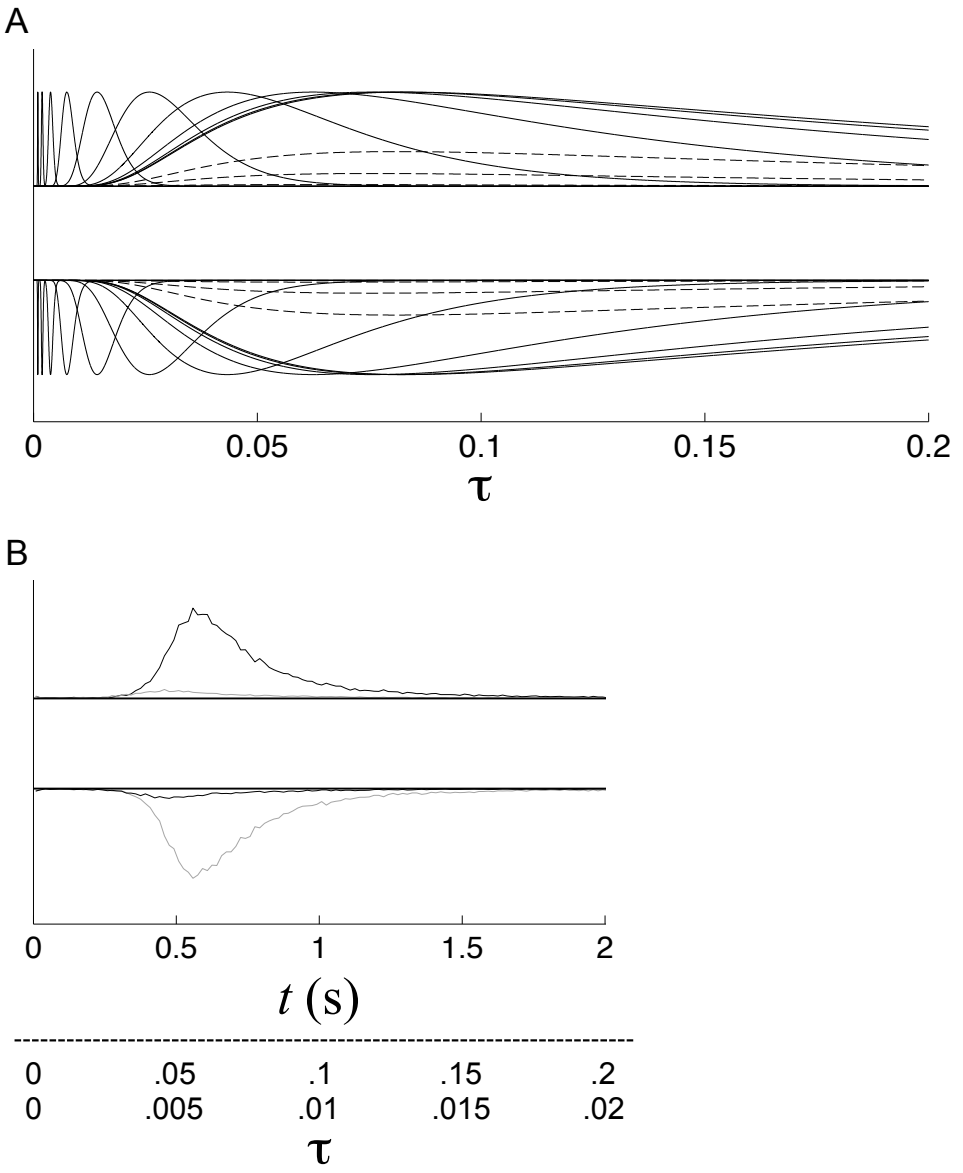

Figure 3: Illustration of the flexibility from temporal scaling of the general diffusion model (gDM) when the model is expressed in dimensionless parameters (Equation 2.22). A: The possible predictions of the model in terms of the dimensionless time parameter τ. For simplicity, the dimensionless starting point $\breve{z}$ is constant at 1/2, and the decay rate is zero. Each curve represents a theoretical RT distribution predicted under a particular value of the dimensionless growth rate $\breve{k}$. Upper curves represent RTs for one response, and lower curves for the other response. Values of $\breve{k}$ used were 0, ±1, ±2, ±4, ±8, ±16, ±32, ±64, ±128, ±256, ±512. Dashed curves represent the less likely response for each growth rate (lower response for $\breve{k} > 0$ and upper response for $\breve{k} < 0$) and are only visible for $\breve{k} = \pm1$, ±2, and ±4. The pair of curves for each value of $\breve{k}$ is normalized by its maximum for ease of viewing. Predictions of the full model (with random growth rate) for each stimulus and condition are an arbitrary linear combination of curves like the ones shown (subject to the constraint that total probability across responses and times equals unity). This graph shows

how the model is constrained in its predictions for larger values of τ, in terms of links across different times and across the two responses, but that these constraints become vanishingly weak as τ approaches zero. B: Example RT data, taken from Jones et al. (in press, Experiment 2, low coherence condition). The upper and lower pairs of curves correspond to the two responses, with black and grey corresponding to two different stimuli. The data are defined in real time (seconds), and there is flexibility in how this time scale maps onto the abstract scale of the model. The two τ scales below the graph represent two possible mappings, corresponding to values for the timescale parameter, θ, of 10 s (upper τ scale) or 100 s (lower τ scale). Each mapping implies a different placement of the empirical data within the space of possible model predictions shown in Figure 3A. The gDM may or may not provide a good fit with a small value of θ, but it is guaranteed to provide a good fit with a larger value.

used in Figure 4B. In summary, by adopting these growth-rate distributions, the gLBA perfectly mimics the diffusion model. Moreover, it achieves this without the start-point variability assumed by the standard LBA.

The growth-rate distributions in Figure 4C do not conform to the Gaussian shape assumed by the standard LBA. However, there is no evident reason that these distributions should be considered any less psychologically plausible than Gaussian ones. Therefore, the theoretical principles underlying the LBA are empirically indistinguishable from the diffusion model, despite the significant structural differences between the two modeling frameworks. Any differences in predictive success between the two models are informative only in the context of the technical assumption of Gaussian growth-rate distributions.

2.6. Summary of Universality Results

The preceding sections demonstrate that several important psychological models of choice RT become unfalsifiable when certain parametric and selective influence assumptions are removed. The parametric assumptions include the shapes of response processes and the probability distributions of their growth rates and starting points. We refer to the models without these assumptions as universal, because they can match any joint distributions of response and RT.

Three families of models were analyzed: the Grice framework (Grice, 1968), ballistic accumulators (Brown & Heathcote, 2005, 2008), and diffusion models (Busemeyer & Townsend, 1993; Ratcliff, 1978). A principal difference among these families concerns which component of the model is free to vary when fitting data for different stimuli and experimental conditions. In the Grice family, the free part of the model is deterministic, and the stochastic component of the model follows a predetermined distribution. In the original Grice framework, the (deterministic) response processes are free and condition-dependent, and the thresholds follow a predetermined (though arbitrary) probability distribution (Theorems 1 and 2). In the two equivalent variants considered in Section 2.2, the stochastic component is moved to the growth rates of the response processes (Theorem 5) or to their starting points (Theorem 6). In all three cases, allowing full (nonparametric) flexibility in the models' deterministic components makes them universal.

In the results for the generalized LBA and diffusion models, universality comes from freedom in stochastic components of the models (i.e., freedom in their probability distributions). Allowing full flexibility in the distribution of growth rates across trials was shown to make the LBA (Theorems 7 and 8), Wiener diffusion (Theorem 10), and OU models (Theorem 11) all universal. In addition, ballistic accumulators that assume bounded nonlinear response processes can be made universal through flexibility in their distributions of starting points

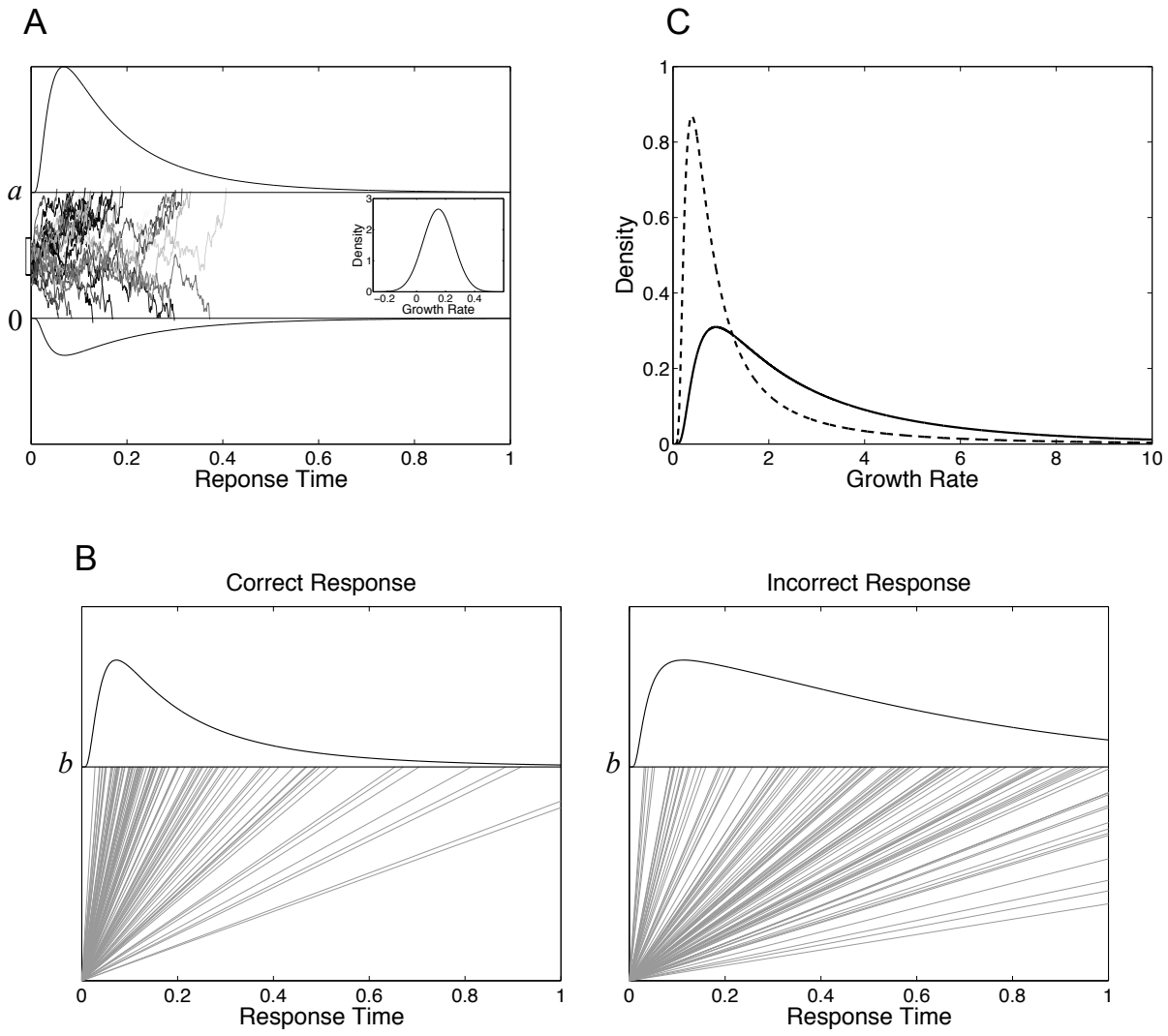

Figure 4: Example of the general linear ballistic accumulator (gLBA) mimicking the diffusion model. A: Diffusion model predictions under typical parameter values (given in main text). Response processes on sample trials are shown in random greyscale (to facilitate discrimination). Growth-rate distribution is shown in inset, and start-point distribution is shown at left edge. Curves paralleling thresholds show the theoretical distribution of response times (RT) for each response, scaled by response probability. B: Illustration of gLBA. Curves paralleling thresholds are first-passage distributions, derived from Equation 2.11, under which any independent race model will reproduce the diffusion model's predictions. The gLBA matches these first-passage distributions through flexibility in its growth-rate distributions. Grey lines show response processes on sample trials. Start-point variability is set to zero. C: Growth-rate distributions used in B. Under these distributions, the gLBA perfectly matches the diffusion model's predictions. Solid: correct response; dashed: incorrect response.

(Theorem 9).

The universality results for ballistic accumulator models can be viewed as stemming from the universality of the first-passage distributions of their individual response processes. Allowing full flexibility in growth-rate distributions or (in some cases) start-point distributions enables the first-passage distributions to take on any desired form. From Theorem 3, this flexibility of first-passage distributions allows the model to match any desired R&T distributions. The same reasoning applies to the Grice framework when the thresholds are chosen to be mutually independent—freedom in the shapes of the response processes again provides full flexibility in the first-passage distributions. Theorem 3 does not apply when the thresholds are stochastically dependent, but the proof for this case follows a similar argument (see Dzhafarov, 1993). In the diffusion model, the relationship between the process and the outcome (i.e., response and RT) is simpler, because there is only one process.[9] In the limit case with no diffusion, the relationship is one-to-one, and it is straightforward to translate any R&T distribution directly into a distribution of growth rates under which the model reproduces that R&T distribution.

An advantage frequently cited for stochastic-accumulation models of speeded choice is that they jointly capture the overall response probabilities and the RT distribution associated with each response. The implied suggestion is that there is some coupling among these measures inherent in the models, so that fitting all of them simultaneously is a more stringent test. The universality results imply there is no such coupling, other than that arising from the parametric and selective influence assumptions. Complete freedom in the joint distribution of response choice and RT is equivalent to freedom in the marginal response probabilities and in the conditional distribution of RT under each separate response. In other words, a universal model can fit all of these measures simultaneously and independently.

The gLBA and gDM models defined here have not been a focus of previous research, and it might be argued that their unfalsifiability is, taken alone, not directly relevant to the extant literature. The primary theoretical import of the universality results lies in their implications for the standard, parameteric versions of the models that have been central to research on speeded choice. Although the standard models are falsifiable, we have shown that their predictive constraints derive entirely from parametric and selective influence assumptions, and it is thus these assumptions that constitute the models' explanatory content.

[9] Diffusion processes have also been proposed within the framework of race models, with one diffusion process per response (Bogacz & Gurney, 2007; Jones, Mozer, & Kinoshita, 2009; see also Usher & McClelland, 2001). Universality of such a model (with free growth-rate distributions) can be shown by the same strategy used here with the LBA model: in the limit of no diffusion, the model becomes a ballistic accumulator, and Theorem 8 applies.

Therefore, it is critical to understand the impact of these assumptions on the models' predictions, if we are to understand how the models explain empirical phenomena. The parametric assumptions have received little attention in the previous literature, being "merely" implementation details and not part of the underlying theory. The justification for and impact of the selective influence assumptions, when separated from the parametric assumptions, remain undetermined too. We focus on these issues in the remainder of this article.

3. SELECTIVE INFLUENCE

An important consideration in the context of the universality results presented in the preceding sections is the concept of selective influence. In general, selective influence refers to the notion that changes in certain aspects of the stimulus, observation conditions, or instructions to a subject might be limited in which parameters of a model they can affect. In the domain of speeded choice, interest in selective influence has focused primarily on two factors: stimulus information that is not available to the subject before the start of each trial (e.g., stimulus intensity), and the speed-accuracy bias of the subject as often manipulated by instructions to emphasize one or the other (other factors not considered here include stimulus probability; e.g., Thomas, 2006). In the context of stochastic accumulation models, it is often assumed that stimulus information can only affect the mean growth rates of the response processes, and not their starting points or the decision thresholds. It is also often assumed that speed-accuracy bias can affect only the decision thresholds (and in some cases the variability of the starting point), but not the response processes. We refer to these as the first and second selective influence assumptions, respectively.

It is easy to see from the statements of the universality theorems that the first selective influence assumption does not in itself impart predictive constraints. That is, the gDM and gLBA frameworks remain unfalsifiable with this assumption included. This follows from the fact that Theorems 7, 10, and 11 all hold with the thresholds assumed to be constant across all stimuli s and conditions c (in fact, their conclusions also hold if the thresholds vary arbitrarily across stimuli and conditions). Therefore, the first selective influence assumption has no impact on the universality results or on their relevance to extant models.

The second selective influence assumption does impart predictive contraints, and the universality theorems do not hold if this assumption is taken as an essential part of a model's architecture rather than an implementational assumption. This fact is evident in the statements of the theorems and in their proofs, in that the growth rate must be specified as a random variable depending on stimulus and on observation conditions ($k^{s,c}$), rather than depending only on the stimulus (k^s). Moreover, it is not possible to extend the proofs to cover the models with the second selective influence assumption maintained. As a simple example,

consider adding the second selective influence assumption to the gLBA without start-point variability. Under this model, the only allowable change across observation conditions is in the threshold, which acts to change the first-passage distributions of the response processes by a scalar factor. Thus, the first-passage distribution for any response can vary across conditions only by multiplicative scaling (i.e., it cannot change shape). Because of the essentially one-to-one relationship, implied by Theorem 3, between the R&T distribution and the set of first-passage distributions for all responses, this constraint implies the model is falsifiable. That is, it is straightforward to construct a family of R&T distributions that, via Equation 2.11, yields first-passage distributions violating the constraint of multiplicative scaling across conditions, and by Theorem 3 such a family cannot be represented by the model in question.

Of course in more complex models, for example with start-point variability or diffusion included, the predictive constraints from the second selective influence assumption will be subtler and weaker than in this example. One important goal for future research might be to determine mathematically what those constraints are for different architectures (e.g., the full gLBA or gDM). Results of this type might be useful in empirically testing selective influence without the confound of the models' parametric distributional assumptions (e.g., Gaussian growth-rate distributions). Furthermore, they might be valuable in assessing the relative contribution of selective influence vis-à-vis parametric assumptions in the past empirical successes of the models. Even though the second selective influence assumption makes the models in principle falsifiable, it is possible that in practice they are not, and that the parametric assumptions have done most of the explanatory work in fitting data.

In general, the value of selective influence assumptions is that they link a model's predictions across different conditions or stimuli, thus offering more demanding empirical tests than fitting one condition at a time. Nevertheless, for such an approach to be viable, there must be strong justification for the invariance. In the case of speeded choice, the traditional justification for the first selective influence assumption is strong: The stimulus or stimulus category is unknown to the subject in advance of each trial, so this information cannot affect any "preparatory" processes such as starting points or threshold settings. The standard justification for the second selective influence assumption is that information available before the start of a trial (such as the weighting of speed vs. accuracy) can influence preparatory processes but should not affect evidence accumulation (e.g., Ratcliff & Smith, 2004). We consider the argument for this latter assumption to be significantly weaker. Nothing seems to exclude *a priori* the dependence of evidence accumulation not only on stimulus values but also on instructions and other factors that act before stimulus onset. Because these factors have already been determined at the start of any trial, they could very well affect the subject's cognitive state in a way that alters stimulus processing.

Indeed, models that attempt to explain the mechanisms underlying stochastic evidence accumulation predict that the accumulation process should be affected by cognitive variables. Consider the exemplar-based random walk model of Nosofsky and Palmeri (1997), which explains choice RT in perceptual categorization. A critical variable in this model is the subject's attentional weighting of the dimensions of stimulus variation. This weighting affects similarity of the current stimulus to stored exemplars, which affects exemplar retrieval probabilities, which in turn determine the statistics of the random walk. Thus the model predicts different accumulation processes (e.g., different mean growth rates) for different attentional states. Attention switches among stimulus dimensions also influence the diffusion processes in multiattribute decision field theory (MDFT; Diederich, 1997; Roe et al., 2001). Although MDFT does not address the determinants of attention, attention should be expected to depend on cognitive variables relevant to the decision task. In particular, it seems reasonable that in many decision tasks there exist different stimulus dimensions with different speed-accuracy profiles, in that some are processed more rapidly but others yield greater asymptotic performance. As a simple model of this sort of tradeoff, we could assume two attentional states, α_1 and α_2 (each a possible attended stimulus dimension or a continuous-valued weighting of dimensions), with $v^{s,\alpha_1} > v^{s,\alpha_2}$ for all s, and $\eta^{\alpha_1} > \eta^{\alpha_2}$. For the right values of v and η, we might expect the subject to strategically choose α_1 under speed instructions and α_2 under accuracy instructions. Starns, Ratcliff, and McKoon (2012) make a similar suggestion in the context of recognition memory, based on the proposal that subjects use poorer quality memory probes under speed instructions (because better probes take longer to develop; Diller, Nobel, & Shiffrin, 2001; Malmberg, 2008). They found that a model obeying selective influence gave a poor fit to the data, and a model allowing growth rates to depend on speed-accuracy instructions yielded estimates of faster mean growth rates in the accuracy condition.

Another way of arguing that the second selective influence assumption is not *a priori* compelling is to observe that in all the models considered the change of a threshold (or threshold separation) is mathematically equivalent to changes in other model components (growth rate, starting point, and diffusion rate) with thresholds left unchanged. Thus, to consider a simple case, if all processes in the gLBA are assumed to start at zero, then a change of a threshold by a factor x is equivalent to changes of the growth-rate distributions by a factor of $1/x$. If stated in this form, the second selective influence assumption does not seem to be structural. For instance, if a model with this assumption does not fit data, one would not consider it a dramatic change to replace multiplicative scaling of the growth rates with some other monotone transformation.

In summary, of the two selective influence assumptions made by the standard diffusion and LBA models, the first has no impact on universality, and the second is logically suspect

and perhaps even psychologically unlikely. Therefore, to support the claim that the models are theoretically informative in light of our results, one would need (1) to offer some sort of argument or empirical evidence for the second selective influence assumption, and (2) to demonstrate that it constrains the models enough to make useful psychological conclusions in the absence of parametric distributional assumptions.

There is also a completely different question about selective influence assumptions, related to translating falsifiable models between universal modeling languages. Let L and L' be two universal (hence interchangeable) modeling languages, and let some falsifiable model M be formulated in language L. By universality of L', one can always translate model M from language L to language L' to obtain a model M', equivalent to M. The question is: could M' lose some of the selective influence assumptions made in the original model? If the answer to this question were affirmative, one might have a basis to prefer M to M' even though they are equivalent. But the answer in fact is that model M' will necessarily retain all the selective influence assumptions of model M. The parameters of M will carry over to parameters in M', and the latter will automatically obey the same selective influence rules as do the former. That is, if a parameter θ in M is invariant with respect to changes in some experimental factor f, then M' will include a corresponding parameter θ' that is invariant with respect to f as well.

As an example, consider translating from the gLBA to the Grice framework, with some pre-specified threshold distribution for the latter. Universality of the Grice framework implies there is a canonical translation by which any model expressed in the gLBA framework can be re-expressed as an equivalent Grice model. In particular, such a Grice model exists for the standard LBA, with its parametric and selective influence assumptions included. Under this translation, the LBA's A^c and v^s parameters become parameters determining the Grice response processes: $R^{s,c}(t) = R(t; A^c, v^s, \ldots)$, where "..." designates other parameters of the LBA model. Because $A^c, v^s, \ldots$ uniquely determine the RT distributions in the LBA model, the process $R^{s,c}(t)$ may be described in terms of these and only these parameters. All selective influence assumptions are automatically retained, only A^c and v^s are now interpreted as aspects of the process $R^{s,c}(t)$ rather than of starting-point and growth-rate distributions. The particular *form* of selective influence has arguably changed, but only if one assumes a naive correspondence between modeling languages: response processes mapping to response processes and thresholds mapping to thresholds.

In the Grice language, the thresholds have fixed distributions and only the response processes are allowed to vary. Nevertheless, the Grice model retains selective influence unless one seriously maintains that there is a difference between separate mathematical "entities" and properties of a single mathematical "entity." As a simple example to show that such a position would be untenable, let $R^{s,c}(t)$ be Taylor-expanded as $R(0) + v^s t + \frac{1}{2} w^c t^2 +$

$u(t)$, where v^s and w^c are two selectively influenced parameters. These parameters can be interpreted as properties of the process $R^{s,c}(t)$, namely its initial velocity and acceleration, so that $R^{s,c}(t)$ depends on both s and c. However, they can also be represented as aspects of separate processes $R_1(t) = v^s t$ and $R_2(t) = \frac{1}{2} w^c t^2$, additively combined with $R_3(t) = R(0) + u(t)$. Clearly, one's interpretation of the selectiveness of influences in this model cannot depend on which of these (or many other) ways of thinking about the two parameters one adopts.

Therefore, selective influence can only aid in deciding among modeling languages if one has a basis to claim that the form of selective influence manifesting in one language is more psychologically plausible in one language than the form manifesting in another. With regard to the modeling languages considered in the present article, the next section translates the LBA and diffusion models into the Grice framework and shows that the model parameters and their selective influences remain just as natural and interpretable as in the original models.

4. TRANSLATING THE LBA AND DIFFUSION MODELS INTO THE GRICE FRAMEWORK

The core message of this article is that, when a modeling framework is universal, the predictive content of any model expressed in that framework lies in whatever falsifiable assumptions that model makes. For the standard LBA and diffusion models, these assumptions are the forms of the probability distributions for growth rate (Gaussian), starting points (uniform), and nondecision time (uniform, for the diffusion model), together with the selective influence assumptions. A useful way to evaluate these assumptions is to translate them across different modeling languages. That is, for any falsifiable model expressed in one modeling framework, an equivalent model can be derived in a different framework. This translation can offer new insight into the implications of a particular model's assumptions, and it can enable comparison of the explanatory utility of different modeling languages.

In this section, we translate the standard LBA and diffusion models into the Grice framework. Universality of the Grice framework implies that it can exactly mimic any other model. Because this universality holds under essentially any threshold distribution (Theorem 1), we choose the thresholds to be mutually independent, each following a unit-exponential distribution: $\Pr[\theta_r \leq x] = 1 - e^{-x}$. With this choice, the response processes acquire a simple form (see Equations 2.10 and 2.11):

$$R_r^{s,c}(t) = \int_0^t h^{s,c}(r,\tau)\mathrm{d}\tau, \tag{4.1}$$

where $h^{s,c}(r,\tau)$ is the joint hazard function of the model being translated into the Grice framework. We refer to the resulting Grice model as a *Grice representation* of the original model.

The Grice representation of a model can be viewed as an alternative way of representing the predictions of that model, equivalent to deriving its predicted R&T distribution (or joint hazard function, or R&T density) directly. Thus, the Grice representations of the LBA and diffusion models can offer new insights into the implications of those models' assumptions. Specifically, we systematically vary the parameters of these models, and by observing the effects on their Grice representations evaluate what types of flexibility the models' parameters do and do not provide. In addition, we derive Grice representations of the LBA and diffusion models with their parameters fit to empirical data, and compare the results to the Grice representation of the dataset itself (i.e., the Grice model that perfectly fits the data), to assess the different models' fits to data in one and the same theoretical language (the Grice framework).

4.1. Grice Representation for the LBA

Because the LBA and Grice frameworks both comprise independent race models, they will predict the same R&T distribution if they agree on the first-passage distribution for each response. Therefore, any LBA model can be translated to a Grice model by deriving the LBA's first-passage distribution for each response and then deriving response processes for the Grice model that reproduce those distributions. The calculations are carried out in Appendix C, with Equations C.4 and C.5 defining a Grice model that matches the LBA's predictions for any parameter settings.

To assess the effects of the LBA's parameters in a two-choice task, b, A, v_2, and η were systematically varied to produce a series of Grice representations for each.[10] Nondecision time, t_0, was held to zero in this analysis because its effect is only to shift the Grice processes later in time. Following Brown and Heathcote (2008), the mean growth rates for the two responses were constrained to sum to 1, by setting $v_1 = 1 - v_2$ (where 1 and 2 represent the nominally incorrect and correct responses, respectively).[11] The parameter values used, shown in Table 3, were chosen to be similar to values from fits to empirical data (see Brown

[10] As the sources of systematic variation in the parameter values are immaterial in the present analysis, we suppress superscripts for stimuli and conditions in this and the following section.

[11] Brown & Heathcote's motivation for this constraint is that it eliminates the degenerate degree of freedom in the model's parameters that arises from the arbitrary scaling of the response processes and threshold. As noted above for the diffusion model, this strategy mistakenly limits the generality of the LBA when parameters are yoked across multiple experimental conditions (Donkin et al., 2009).

Table 3: Values of LBA model parameters used in investigation of individual parameter effects on Grice representation

Parameter	Values				
b	.4	.6	**.8**	1.0	1.2
A	0	.2	**.4**	.6	.8
v_2	.6	.7	**.8**	.9	1.0
η	.1	.2	**.3**	.4	.5

Note: For each parameter, model predictions were derived for all five values, while holding the other three parameters to the values shown in boldface (e.g., each of the five values of b was combined with $A = .4$, $v_2 = .8$, and $\eta = .3$). Mean growth rates were constrained by $v_1 + v_2 = 1$. Nondecision time was held to zero throughout.

& Heathcote, 2008), with ranges selected to illustrate the type of variability introduced by each parameter. For each parameter, five Grice representations were derived, corresponding to the five chosen values of that parameter, while the other parameters were held to their central values (i.e., to the third entry in each line of Table 3, shown in boldface).

Figure 5 displays the resulting Grice representations. Each panel shows the results of varying a different parameter, with curves colored light grey through black corresponding to the first through fifth entries in each row of Table 3. Solid curves are for the nominally correct response ($r = 2$), and dashed curves are for the incorrect response ($r = 1$). The third-darkest curves in all plots are the same, corresponding to the third value for all four parameters.

These results reveal a great deal of flexibility in the LBA model, even with its parametric assumptions retained (see Table 5 for summary). Variation in mean growth rates produces opposite effects on the two Grice response processes, speeding one while slowing the other (Figure 5C). Variation in b or A has the effect of speeding or slowing both Grice processes together (Figures 5A & 5B). Thus, variation in v_2 together with variation in b or A provides the degrees of freedom to independently manipulate the growth rates of both Grice processes (essentially by manipulating their difference and their mean).

Moreover, b and A have different effects on the shapes of the Grice processes. The effect of A is approximately temporally uniform, shifting the Grice processes earlier in time but preserving their shape. The effect of b is temporally nonuniform, being greater at larger values of t. Combining variation in b and A thus allows selective control of only the early or late portions of the Grice processes, as shown in Figures 6A and 6B, respectively. Table 4 shows the parameter values used in generating Figure 6. In both panels, b and A were jointly varied through the five pairs of values in Table 4, while v_2 and η were fixed at their third

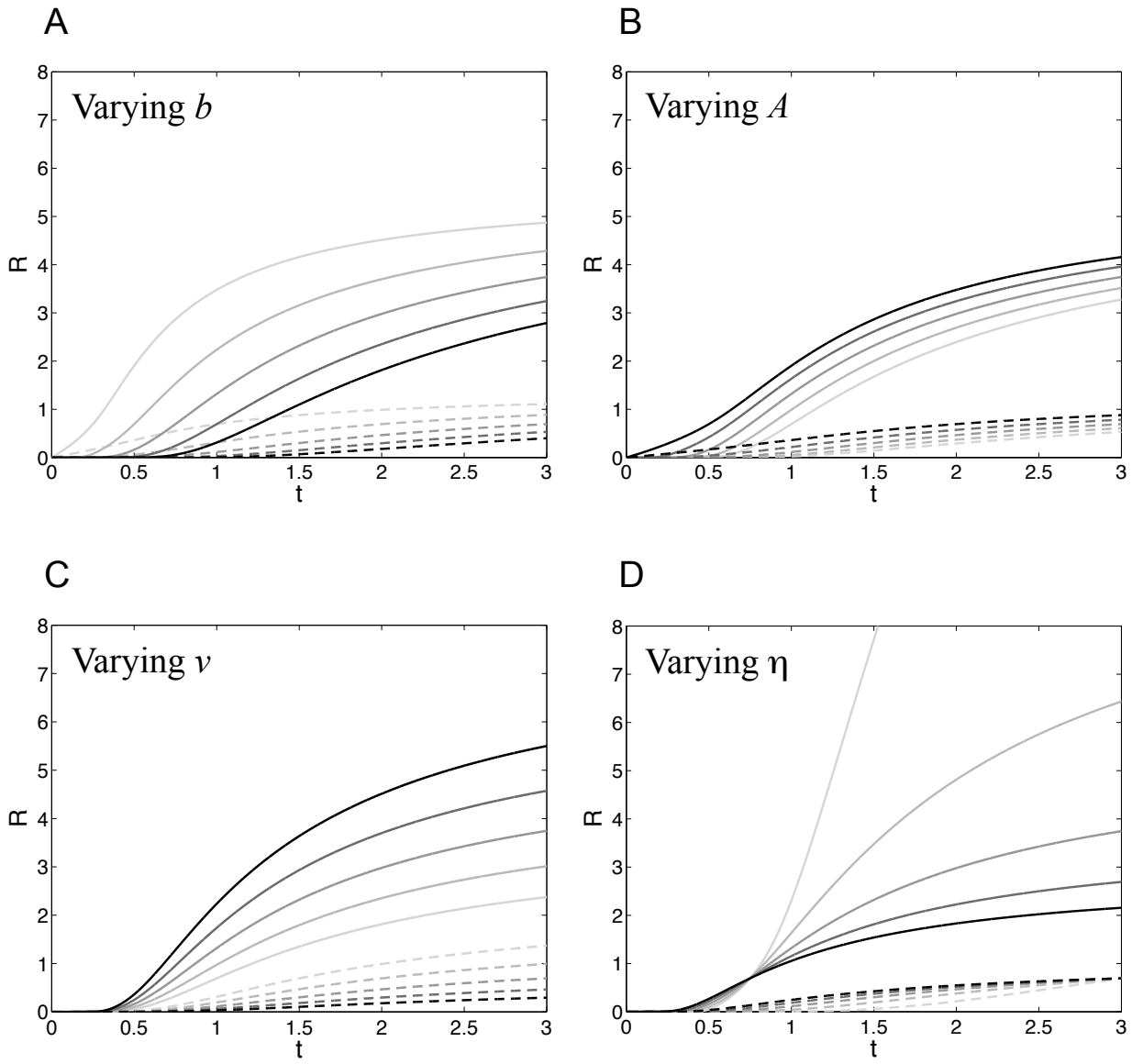

Figure 5: Grice representations of LBA model under variation of individual parameters. Light through dark curves correspond to first through fifth entries of each row of Table 3. Each curve shows a Grice response process (R) as a function of time (t). Solid: correct or modal response ($r = 2$); dashed: incorrect or nonmodal response ($r = 1$).

(boldface) values in Table 3.

Variation in η produces somewhat more complex effects in the Grice response processes (Figure 5D). Smaller values of η delay their initial rise from zero but make them steeper thereafter. This result is sensible, because decreased variance in slopes of the LBA's response processes narrows the resultant RT distribution, reducing the proportions of both fast and slow responses.

Table 4: Values of LBA model parameters used in joint variation of threshold and start-point variability

Parameter	Values				
Figure 6A					
b	.48	.56	.64	.72	.8
A	0	.2	.4	.6	.8
Figure 6B					
b	.4	.6	.8	1.0	1.2
A	.4	.6	.8	1.0	1.2

Note: Both parameters were varied simultaneously through the five pairs of values (i.e., columns) shown for each figure. Other model parameters were fixed at their boldface values in Table 3.

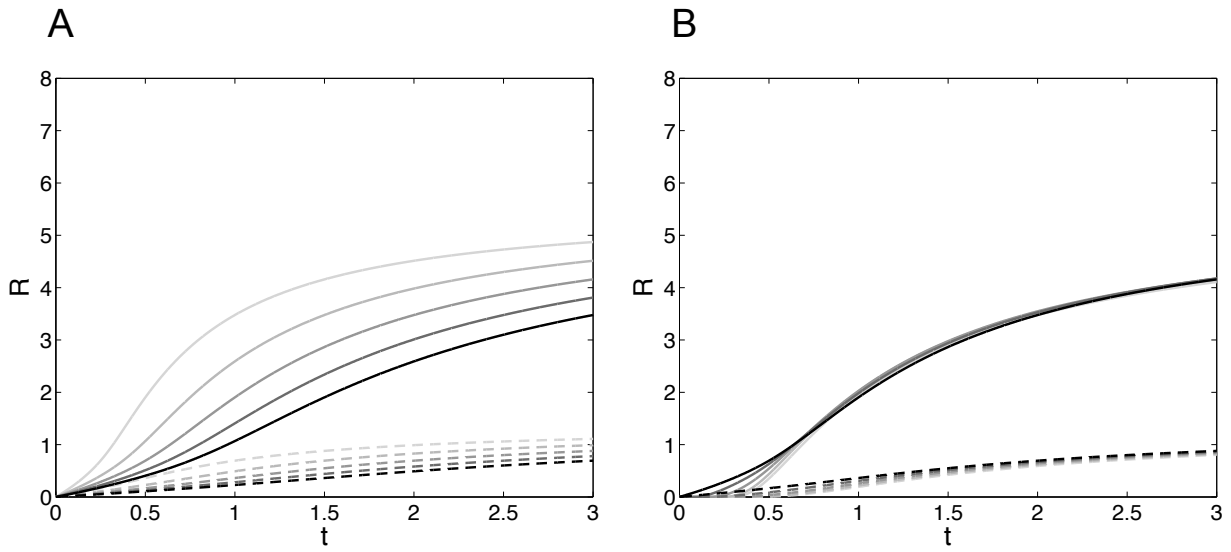

Figure 6: Grice representations of LBA model obtained from two different forms of simultaneous variation of b and A. Light through dark curves correspond to first through fifth columns of Table 4. Solid: correct or modal response ($r = 2$); dashed: incorrect or nonmodal response ($r = 1$).

In conclusion, variation of the LBA's parameters reveals the constraints on its Grice representation. These constraints arise from its distributional assumptions regarding growth rates and starting points (without which it would have no constraints). Specifically, the response processes of the standard LBA's Grice representation (using independent unit-exponentially distributed thresholds) are always sigmoid in shape, consisting of an initial period of negligible growth, an effective onset point where they rise from zero, and a final phase of concavity (deceleration). The LBA's parameters all have straightforward qualitative interpretations within this framework, allowing near-complete flexibility of the sigmoid

Table 5: Interpretations of LBA and diffusion model parameters within Grice framework

LBA Parameter	Diffusion Parameter	Grice Interpretation
b	a	Growth rate of both processes
A		Growth rate of both processes
	δ_z	Curvature of response processes
v_2	v	Differential growth rate between processes
η	η	Curvature of response processes
t_0	T_{er}	Onset time of response processes
	δ_t	Abruptness of onset
	β	Growth rate of both processes
	σ	Growth rate of both processes

pattern. The effective time of onset and initial growth rate are controlled by A and b, the differential growth rate between the two responses is controlled by v, and the change in growth rate over time (concavity) is controlled by η.

4.2. Grice Representation for the Diffusion Model

To obtain Grice representations for the diffusion model, its predicted R&T distribution and joint hazard function were computed using numerical integration (see Appendix B) and then translated to Grice response processes using Equation 4.1. As with the LBA, each of the diffusion model's parameters was systematically varied to obtain a series of Grice representations. Although the diffusion rate is traditionally held fixed as a scaling parameter, it was included in this analysis to understand its unique contribution (equivalent to simultaneously varying a, $\bar{z}$, δ_z, v, and η). Table 6 shows the parameter values used. Each parameter was varied through three values, with the other parameters fixed at their middle values. The results are shown in Figure 7.

As with the LBA, the effects of the diffusion model's parameters on its Grice representation reveal a great deal of flexibility, even with its parametric assumptions retained (see Table 5). The mean growth rate (v) provides flexibility in the difference in slopes of the two Grice response processes, whereas the boundary separation (a) enables the slopes of both processes to change together. Therefore, as with the LBA, joint variation in v and a enables independent variation of the slopes of both Grice processes. Furthermore, the diffusion model can effect these changes in multiple ways. The standard deviation of growth rates (η) has primarily the same effect as does v, although at longer timescales η can be seen to affect the processes' curvature as it does in the LBA. The decay parameter of the

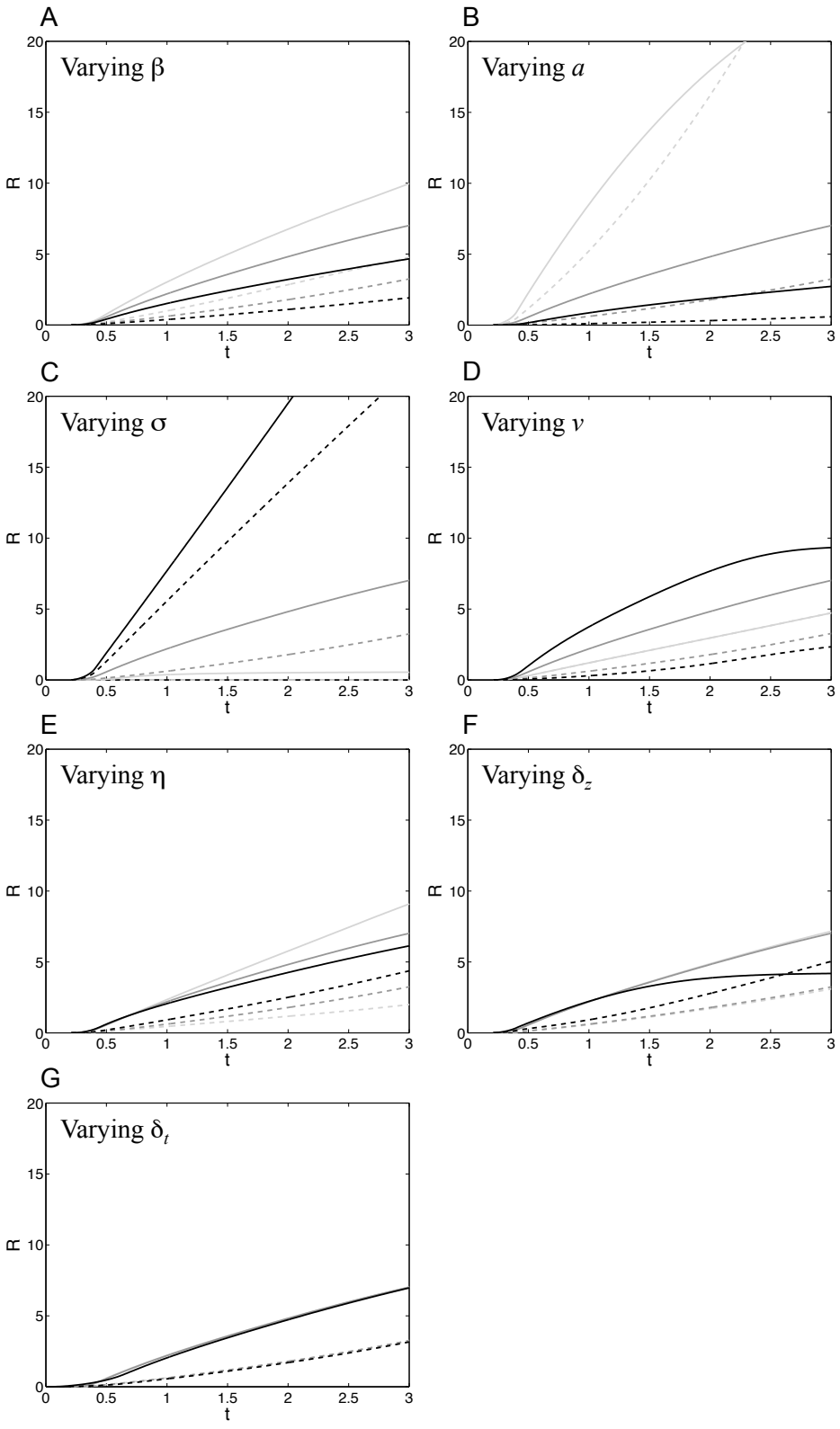

Figure 7: Grice representations of diffusion model under variation of individual parameters. Light through dark curves correspond to the first through third entries in Table 5. Solid curves: correct or modal response; dashed curves: incorrect or nonmodal response.

Table 6: Values of diffusion model parameters used in investigation of individual parameter effects on Grice representation

Parameter	Values		
β	0	**4**	8
a	.05	**.1**	.15
σ	.01	**.1**	.2
v	0	**.15**	.3
η	0	**.1**	.2
δ_{z}	0	**.03**	.09
δ_{t}	0	**.2**	.6

Note: For each parameter, model predictions were derived for all three values, while holding the other six parameters to the values shown in boldface. In all cases, the mean starting point, $\bar{z}$, was held to $a/2$, and mean nondecision time, T_{er}, was set to .3 s.

OU model (β) and the diffusion rate (σ) both have primarily the same effect as does a.

The variability in nondecision time (δ_{t}) primarily affects only the initial part of the Grice processes, enabling them to depart from zero suddenly or more gradually. This finding is consistent with previous conclusions that nondecision variability can be necessary for fitting the leading edge of empirical distributions (Ratcliff, Gomez, McKoon, 2004) but that it has little effect on later portions of those distributions (Ratcliff & Tuerlinckx, 2002). The variability in starting points of the diffusion process (δ_{z}) has little effect in the range of values found in fitting empirical data (compare Table 6 with Table 7 below), but at higher values it affects the curvature of the Grice processes, in opposite directions for the two responses.

In conclusion, the diffusion model's parameters all have relatively simple interpretations within the Grice framework. Furthermore, the parameters appear to provide near-complete flexibility in determining the slope of each Grice process, their (common) time of departure from zero, the smoothness of that transition, and their differential curvature.

4.3. Fits to Empirical Data

The Grice representation for an empirical dataset can be derived similarly as for model predictions, to obtain a family of response processes under which the Grice model perfectly reproduces the R&T distribution of the data. The derivation involves a minor modification to account for the discreteness of the RT distribution from a finite set of trials, as explained in Appendix D. The Grice representation of the data can be compared to Grice representations of models used to fit those data, to yield insight into what aspects of the data the models

Table 7: Diffusion model parameters estimated from Experiment 2 of Ratcliff et al. (2001)

Model	a^{speed}	a^{acc}	v^1	v^2	v^3	v^4	η	δ_{z}	T_{er}	δ_{t}
Wiener	.0821	.1440	.0391	.1320	.1944	.3208	.1485	.0324	.3109	.10
OU ($\beta = 4$)	.0700	.1128	.0399	.1349	.1983	.3312	.1417	.0111	.3302	.15
OU ($\beta = 8$)	.0687	.0975	.0342	.1150	.1702	.2772	.0845	.0100	.3185	.10

Notes: Fits taken from Ratcliff and Smith (2004). Stimulus discriminability levels range from 1 (difficult) to 4 (easy). Superscripts "speed" and "acc" refer to speed instructions and accuracy instructions, respectively.

capture, and how.

This approach is applied to Experiment 2 of Ratcliff et al. (2001), in which subjects viewed a pair of dots on each trial and classified their distance as "small" or "large" by one of two keypresses. Dot separation took on 32 different values. Feedback was probabilistic, such that larger separations were more likely to be followed by reinforcement of the "large" response. Following Ratcliff et al., data were collapsed among stimulus levels with similar response proportions and mean RTs, as well as over the symmetry between responses (e.g., "large" responses to the largest dot separations were pooled with "small" responses to the smallest separations), yielding four stimulus difficulty levels for modeling. Although feedback was probabilistic, we refer to the modal and nonmodal responses for each stimulus level as correct and incorrect, respectively. Finally, there were two instruction conditions, emphasizing speed and accuracy, which alternated between blocks.

Ratcliff and Smith (2004) fit three diffusion models to the data: the Wiener model ($\beta = 0$) and two OU models defined by $\beta = 4$ and $\beta = 8$. The models were fit to the data pooled across subjects, using the following procedure. For each condition and stimulus level, the response probabilities and the .1, .3, .5., 7., and .9 quantiles of the conditional RT distribution for each response were calculated for each subject and then averaged across subjects. The mean quantiles were used as cutoffs to divide the range of possible RTs into 6 bins per response, for a total of 12 possible outcomes on each trial. The model's predicted probabilities for these 12 outcomes were compared to the empirical frequencies to define a modified chi-square statistic reflecting goodness of fit, which was minimized in parameter estimation. Table 7 presents the best-fitting parameters for all three diffusion models, as obtained by Ratcliff and Smith. We applied the same fitting procedure to the LBA model, with best-fitting parameters shown in Table 8.

To derive Grice representations for the data, data were pooled across subjects by averaging individual subjects' response probabilities and conditional RT quantiles, as was done in fitting the models. Because derivation of Grice representations requires full (quasi-

Table 8: LBA model parameters estimated from Experiment 2 of Ratcliff et al. (2001)

b^{speed}	A^{speed}	b^{acc}	A^{acc}	v_2^1	v_2^2	v_2^3	v_2^4	η	t_0
.341	.228	.561	.395	.560	.645	.748	.895	.347	.121

Notes: v_2^s indicates mean growth rate for correct response, with $v_1^s = 1 - v_2^s$. Stimulus discriminability levels range from $s = 1$ (difficult) to $s = 4$ (easy). Superscripts "speed" and "acc" refer to speed instructions and accuracy instructions, respectively.

continuous) RT distributions, we used a richer set of quantile probabilities, ranging from .0001 to 1 in steps of .0001. Each of the corresponding 10000 quantiles was computed separately for each subject and response, and then averaged across subjects. The resulting mean quantiles were then converted back to a pooled conditional RT distribution for each response. The conditional distribution for each response was scaled by the corresponding response rate (averaged across subjects) to produce a pooled joint R&T distribution. This joint R&T distribution has the properties that (1) the marginal probability of each response equals the average response rate across subjects, and (2) every quantile of the conditional RT distribution for each response equals the average quantile across subjects. This procedure was performed separately for each combination of instruction condition and stimulus level.

Figure 8 shows the Grice representations of the data and all four models, in the speed condition. Figure 9 shows the corresponding results from the accuracy condition. The sparsity in the tails of the empirical RT distributions makes the response processes exceedingly noisy beyond 1 s (speed condition) or 2 s (accuracy condition). Therefore, all graphs are terminated at these times. The models differ in their predictions at longer RTs, for example with the OU model exhibiting an eventual crossover between correct and incorrect response processes (not shown), but these differences are not considered further because the data are insufficient to distinguish among them. Note that due to the universality of Grice representations the curves shown in Figures 8A and 9A should be viewed as *data* rather than as theoretical fits to data, on a par with representing the data in the form of empirical distribution functions, quantile functions, hazard functions, etc.

In the speed condition, the Grice response processes for the diffusion models are all nearly linear in the range of the data, starting from the time they depart from zero. The only aspects of the Grice-represented data the diffusion models capture are the departure points and the slopes. The mean-nondecision parameter (T_{er}) allows the model to match any departure point, provided it is the same for all responses in all stimulus conditions (a constraint the data satisfy). As observed in the previous section, the diffusion model's parameters allow it to produce Grice processes of any slopes. The first selective influence assumption, that only the mean growth rate can vary across stimulus levels, implies that

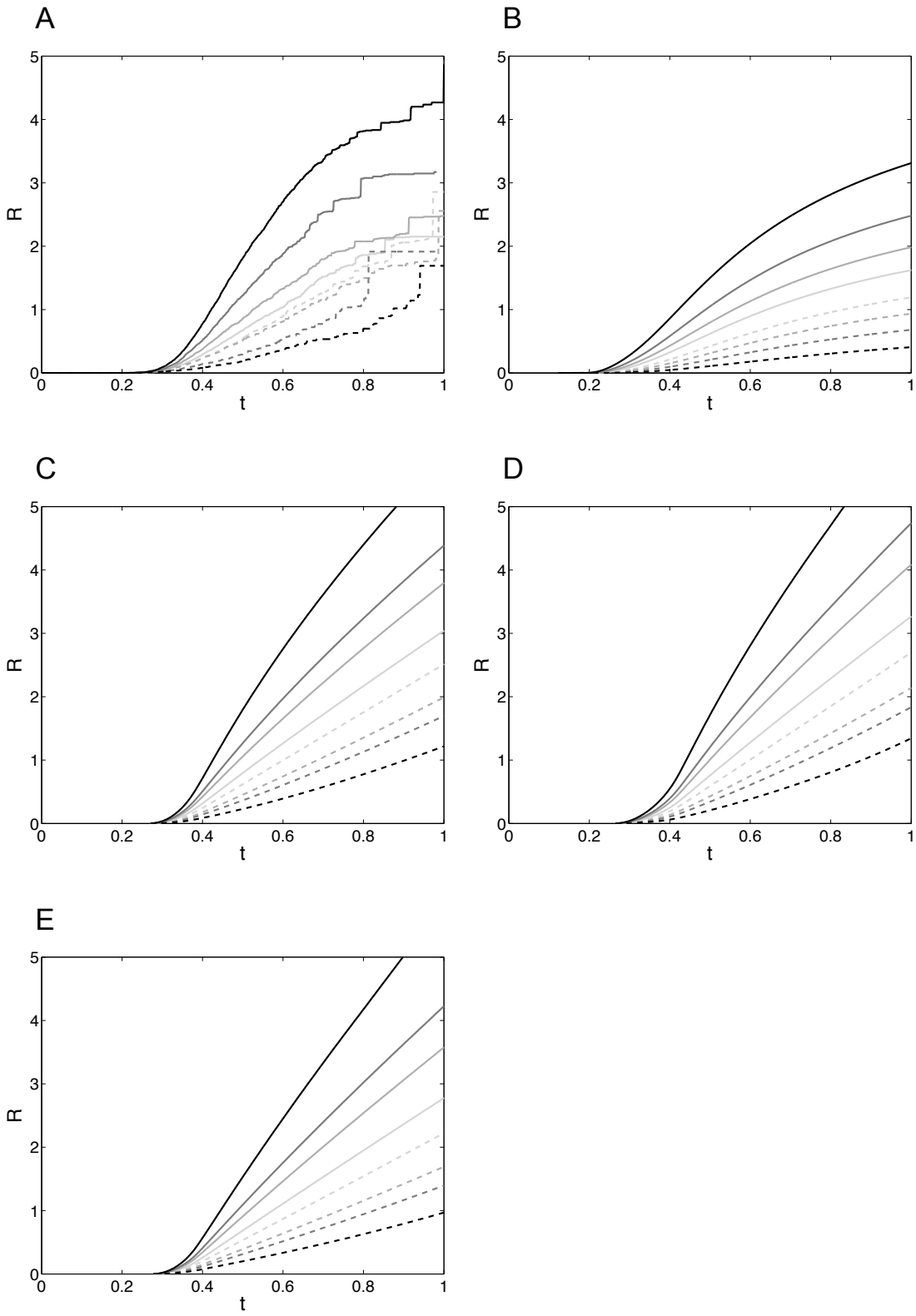

Figure 8: Grice representations derived from data and model fits for speed condition of Experiment 2 of Ratcliff et al. (2001). A: Empirical data. B: LBA model. C: Wiener model ($\beta = 0$). D: OU model, $\beta = 4$. E: OU model, $\beta = 8$. Light through dark curves indicate hardest through easiest stimulus levels. Solid curves: modal response; dashed curves: nonmodal response.

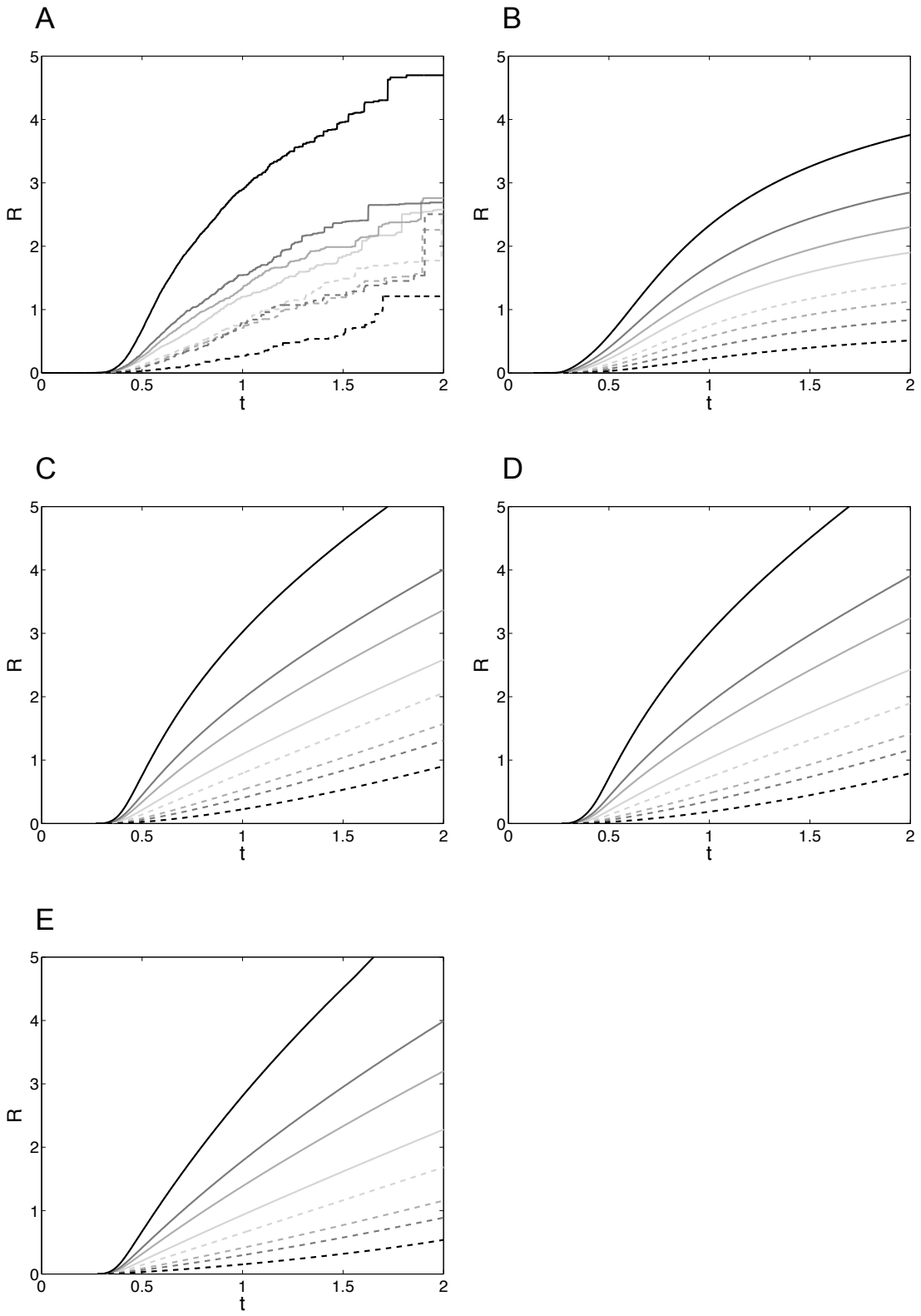

Figure 9: Grice representations derived from data and model fits for accuracy condition of Experiment 2 of Ratcliff et al. (2001). A: Empirical data. B: LBA model. C: Wiener model ($\beta = 0$). D: OU model, $\beta = 4$. E: OU model, $\beta = 8$. Light through dark curves indicate hardest through easiest stimulus levels. Solid curves: modal response; dashed curves: nonmodal response. Note the difference in the timescale when comparing to Figure 8.

the sets of Grice processes for the two responses must be approximately symmetric about some intermediate trajectory for each condition. This prediction is supported by the data. Therefore, the symmetric-trajectory and common-departure properties seem to be the only aspects of the data that substantively support the diffusion model.

The fits of the LBA model to the speed condition show less-linear Grice processes, with an overall negative acceleration. This prediction only partially holds in the data, where the Grice processes are concave for the correct response but somewhat convex (positively accelerated) for the incorrect response. The excessive concavity of the LBA's predictions results in its significantly overestimating the .9 RT quantiles for both responses (not shown). The universality property implies the LBA could perfectly match the data if allowed non-Gaussian growth-rate distributions, but these results show how the Gaussian assumption constrains the model's predictions. Other than this constraint, the LBA matches the data in the same manner as does the diffusion model. The Grice processes' departure point, overall average slope, and difference across stimulus values in relative slopes for the two responses are all trivially reproduced. The only necessary predictions of the LBA model that are supported by the data seem to be the symmetric-trajectory and common-departure properties.

In the accuracy condition, the empirical Grice processes are more nonlinear, concave for the correct response and somewhat convex for the incorrect response. The diffusion model reproduces this curvature, especially with lower values of β, as does the LBA. In all other respects, the same analysis holds as for the speed condition.

Finally, the second selective influence assumption of the diffusion and LBA models, that only the threshold and (in the LBA) start-point variability can vary across instruction conditions, manifests as a constraint linking their Grice representations across conditions. Specifically, the overall average growth rates of the Grice processes differ between conditions, but the differences in growth rates between responses and across stimulus levels do not. This prediction is consistent with the data (note that the timescales differ between Figures 8 & 9).

In conclusion, the Grice representations offer a new perspective on the predictive constraints in the diffusion and LBA models arising from their parametric and selective influence assumptions. They show that the flexibility identified in the previous two subsections enables the diffusion and LBA models to match most aspects of the data in a post hoc manner. That is, had these features taken on different values, the models could have matched them as well. There are four aspects of the data, when expressed as Grice representations, that are necessary predictions of the models: (1) The shapes of the Grice response processes are characterized by initially negligible growth, an effective onset point, and a final phase of roughly constant concavity or convexity. (2) The effective onset points are constant across responses and stimulus levels. (3) Variation in stimulus difficulty has opposite and approximately

equal effects on the growth rates of the Grice processes for the two responses. (4) Manipulation of subjects' speed-accuracy emphasis affects the growth rates of the Grice processes approximately uniformly across responses and stimulus levels. The first of these predictions is the only one that appears related to the models' parametric distributional assumptions. The second prediction is primarily due to the assumption that the non-decision component of the models is independent of the response and of stimulus and instruction manipulations. The last two predictions arise from the selective influence assumptions. Critically, none of these predictions is closely tied to the fundamental architectures of the models. They are just as naturally interpretable when expressed as assumptions in the Grice framework as when expressed in the LBA and diffusion frameworks.

5. DISCUSSION

Whenever a cognitive model provides a good account of empirical data, it is critical to understand which of its assumptions are responsible for its predictive success. Such understanding is important for theoretical progress and for generalizing to other paradigms or domains. Moreover, the assumptions of most formal models can be roughly divided into ones corresponding to theoretical principles the model is meant to embody, and technical details that are necessary to generate quantitative predictions but are chosen without theoretical consideration and can be modified or dispensed with as need arises. Issues of model flexibility and mimicry are a challenge in any domain, and often the most powerful solution is a nonparametric approach that can be applied to whole model families, to determine the implications of their theoretical assumptions taken alone (e.g., Townsend & Wenger, 2004). For an empirical test of a model to constitute a test of its theoretical principles, it is important to know that those principles provide the main constraint on the model's predictions, and that the choices of technical assumptions are largely inconsequential.

The present work shows that for currently influential models of choice RT, the situation is reversed, in that the predictions are driven entirely by technical assumptions and (to an as yet unknown extent) by the second selective influence assumption. Aside from the second selective influence assumption (which we have argued is ill-motivated), the theoretical principles of both the diffusion and ballistic accumulation frameworks are unfalsifiable, because the inclusion of random growth-rate variability allows any conceivable pattern of RT distributions to be matched. This result is similar to Dzhafarov's (1993) result for the Grice modeling framework (Grice, 1968), that the theoretical assumption of deterministic processes racing to stochastic thresholds is unfalsifiable, and that predictive constraints come only from technical assumptions about the forms of the response processes under specific choices of threshold distributions.

The universality results thus impose a significant limitation on what can be inferred from existing research about psychological processes of speeded decisions. Although the diffusion and LBA models have been highly successful in fitting data from a variety of task domains (e.g., Brown & Heathcote, 2008; Ratcliff & Smith, 2004), this success does not imply any support for the theoretical or structural assumptions of these models. For example, the fact that the universality results hold without diffusion, decay, start-point variability, or nondecision variability implies that these components are superfluous if the growth-rate distributions are unconstrained. All of these mechanisms have been postulated as necessary theoretical principles, yet mathematically none of them is needed to fit data. Any version of the model with these other components included is equivalent to a model without them, with appropriately altered growth-rate distributions.

Setting aside the techical assumption of Gaussian growth-rate distributions, the sole link remaining between theory and prediction is the second selective influence assumption. With this assumption, the models are falsifiable, and without it they are not. This assumption has certainly been posed in the literature as a theoretical principle, not an arbitrary technical detail, and thus one might claim that our results are irrelevant to models that hold this principle. There are several problems with such a position. First, as we argued in Section 3, the second selective influence assumption is not as compelling as the first selective influence assumption (which does not impart predictive constraints). There is no a priori reason that stimulus processing cannot vary across cognitive states, and indeed there are prominent models that assume it does (Nosofsky & Palmeri, 1997; Roe et al., 2001). Second, selective influence has not heretofore been attributed anything close to the lynchpin role our results imply would be needed to save the models. It should be useful to researchers relying on these models to learn that their explanatory value hinges fully on this one assumption.

Third, although selective influence imparts predictive constraints, the strength and nature of those constraints are open questions. It seems likely that the models' choices of parametric distributions (i.e., the Gaussian and uniform assumptions) are responsible for much if not the bulk of the explanatory work in past fits to empirical data. To truly defend selective influence as a standalone theory of choice RT, one would have to derive the logical consequences of the second selective influence assumption in the absence of any parametric assumptions (consequences that presumably would differ across architectures, such as the gDM, gLBA, or Grice framework). Unfortunately, most modern mathematical modeling does not take this type of analytic approach. Instead, models are constructed as conjunctions of many assumptions, making it difficult to separate their logical consequences. This practice puts the field in a position where we do not always know why a model does or does not succeed empirically, and consequently we often have little guidance on what psychological conclusions can be drawn from comparisons of models to data.

A related argument against the present results might be that the diffusion and LBA models both assume that growth-rate distributions are of the same parametric form across stimuli, responses, and conditions. Mathematically, this statement is meaningless. Any finite set of distribution functions can be characterized as instances of some parametric family, that is, made to be of the same "parametric form." In fact, any finite set of parametric families of functions can be united into a single parametric family. Normal distributions and Poisson distributions can be viewed as two parametric classes, but they can also be viewed as belonging to the exponential family of distributions. Linear and cubic functions are members of the polynomial family. To take an unconventional example, the families of functions $A \sin ax$ and Bx^{β} are subsets of the family $Cx^{\beta} (\sin ax)^{\delta}$, provided $\beta = 0$ and $\delta = 0$ are not excluded. Therefore the relevant question is not why a single parametric form is possible, but why the particular parametric family that has dominated past work (viz., the Gaussian) has been so empirically successful.

Empirical success by a falsifiable model implies that it is capturing some regularity in the data, and hence in cognitive processing, but it is not clear in the current case what that regularity is. One possibility is that the Gaussian distribution is psychologically correct, but this is doubtful for three reasons. First, there is no clear reason to expect a Gaussian. Whereas the Gaussian distribution of within-trial variability emerging from a diffusion process can be explained by the summation of many independent neural events (via the central limit theorem), there is no obvious candidate for a between-trial analog—that is, a large number of independent and identically distributed between-trial variables that sum to determine the drift rate. Second, Ratcliff (2013) and Donkin and Little (2013) have recently shown through simulation that the diffusion and LBA models yield similar predictions if the Gaussian drift distribution is replaced by a uniform or beta distribution, suggesting that any of a range of distributions might be adequate. Further work along these lines might be informative in determining what aspects of a growth distribution are needed for good fits to data (e.g., symmetry or thin tails). However, a third caveat is that the empirical success of a distribution will generally depend on the model architecture. For example, the standard LBA can be translated to an equivalent gLBA model without start-point variability, and this new model will have quite different growth distributions. In initial derivations not reported here, we have found these distributions to be positively skewed with heavier tails. Heathcote and Love (2012) have recently explored a similar model and found that a lognormal growth distribution gives good fits to data. Therefore one needs to consider both distributional and structural assumptions jointly.

When a modeling framework is universal, it is best thought of as a language for expressing models. Empirical tests of a model are tests of falsifiable assumptions stated in that language, not of the framework itself. Nevertheless, a modeling language can be regarded as more useful

or less useful depending on the interpretability of the assumptions needed to give good fits of data. Constraining assumptions will take different forms when translated between different languages, some perhaps more transparent or suggestive than others. Thus, a relevant question is which framework admits models that can explain the data in a way that is both mathematically convenient (e.g., using few free parameters and allowing for analytic solutions) and also easily interpretable (e.g., in terms of which parameter controls which theoretical construct and reflects which observable property of RT distributions).

Although the aim of this article is not to advocate for the Grice framework as a language for modeling speeded choice, the results of Section 4 suggest it fares well on both of the above criteria. First, the Grice representation of the empirical data from Ratcliff et al. (2001) suggests that human behavior can be well matched using fairly simple Grice response processes. In terms of free parameters, the formal translation between modeling languages implies the Grice processes require no more than the number of free parameters in the model (i.e., diffusion or LBA) from which the translation was derived. However, consideration of the qualitative form of the Grice representations derived here suggests a simpler parameterization might be possible: One might need only to specify an onset (i.e., time of departure from zero), slope for each response, and possibly quadratic terms for curvature. These assumptions are mathematically simpler than the conjunction of the assumptions of a Gaussian distribution, two uniform distributions, Brownian motion, and decay. Moreover, one might argue that deterministic processes paired with random thresholds are inherently simpler than stochastic processes, because a stochastic process always involves some dynamics in time and one or more random variables at each time point, but in addition one has to specify (under certain regularity assumptions) a joint distribution of these random variables across every finite set of time points. Second, the unique effects of stimulus level and of speed-accuracy manipulations have natural interpretations in the Grice framework. Increasing stimulus intensity can be assumed to increase the growth rate of the correct response process and to decrease the growth rate of the incorrect process. Shifting the subject's emphasis toward speed can be assumed to increase the growth rates of all processes. Both of these assumptions seem no less theoretically motivated than assumptions regarding growth rates and decision thresholds in the other frameworks.

Finally, it is important to note that the present results apply to models that make predictions only for choice and RT. The postulated mechanisms within these models (viz., response processes and decision thresholds) are treated only as mathematical entities that determine predictions for these observable variables. Other measures have been used to investigate speeded decision-making, such as confidence ratings (Ratcliff & Starns, 2009), motor trajectories (Spivey, Grosjean, & Knoblich, 2005), eye movements (Krajbich, Armel, & Rangel, 2010), and neural recordings (Roitman & Shadlen, 2002), and the identifiability

problems raised here might be reduced by requiring a model to predict these variables in addition to choice and RT. Such an extension would require additional assumptions about how the internal mathematical constructs of the model map onto physical observations. For example, a large body of research on monkeys' saccade responses to visual motion has led to the proposal that eye movements are triggered by threshold-crossing of activity in the lateral intraparietal area, which follows a diffusion process by temporally integrating upstream activity in the middle temporal visual area (Mazurek, Roitman, Ditterich, & Shadlen, 2003). Although there have been empirical challenges to this theory (e.g., Filimon, Philiastides, Nelson, Kloosterman, & Heekeren, 2013), it exemplifies how, by committing to physical interpretations of otherwise abstract model constructs, one can potentially leverage neural and behavioral data to make better progress than is possible from either alone.

However, in taking an implementational stance on otherwise algorithmic-level models (see Marr, 1982), one should be cautioned against automatically adopting naturalistic analogies prompted by terminological conventions such as calling some theoretical constructs "processes" and others "thresholds." For example, the finding that RT correlates better with the growth rate of neural activity than with its final level (Hanes & Schall, 1996) might suggest that neural interpretations are more compatible with variability residing in accumulation processes than in thresholds, but Grice response processes need not be directly identified with physical activations of neurons. Because the Grice framework is situated at an algorithmic level of description, it can be mapped onto neural processing in many alternative ways, some of which might be found to provide a natural correspondence. Recall, as a simple example, the Grice-equivalent frameworks of Theorems 5 and 6. The "deterministic response processes" and "random thresholds" in the Grice model both correspond to aspects of the response processes in these alternative models. Therefore, if the response processes in either alternative model are identified with neural activation, then the Grice framework can be seen as compatible with stochastic neural activity developing toward a fixed neural threshold.

5.1. Conclusions

A long history of experimental research in speeded choice has produced a rich body of empirical regularities regarding choice probability, RT distributions, and their dependence on various factors. Mathematical modeling has produced models that often yield impressive fits to these data with relatively few free parameters. Nevertheless, the theoretical implications of these modeling results are far less certain than they have been made out to be. As we have shown here, the models' predictions derive not from their structural assumptions but from technical aspects that have been considered irrelevant details. Understanding

the predictive constraints and theoretical implications of these technical assumptions, together with those of the second selective influence assumption, is thus an important goal. The methods introduced here for translating falsifiable models between universal modeling languages may be useful toward that end.

References

[1] Bogacz, R. & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation, 19*, 442-477.

[2] Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review, 112*, 117–128.

[3] Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*, 153-178.

[4] Buonocore, A., Nobile, A. G., & Ricciardi, L. M. (1987). A new integral equation for the evaluation of first-passage-time probability densities. *Advances in Applied Probability, 19*, 784-800.

[5] Busemeyer, J., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432-459.

[6] Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology, 41*, 260– 274.

[7] Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 414–435.

[8] Donkin, C., Averell, L., Brown, S. D., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the Linear Ballistic Accumulator model. *Behavior Research Methods, 41*, 1095-1110.

[9] Donkin, C., Brown, S.D, & Heathcote, A. (2009). The over-constraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review, 16*, 1129-1135.

[10] Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models for response time, same conclusions about psychological mechanisms? *Psychonomic Bulletin & Review, 55*, 140-151.

[11] Donkin, C., & Little, D. R. (2013). Response time models without within-trial variability are robust to changes in their distributional assumptions. *Manuscript submitted for publication.*

[12] Draglia, V. P., Tartakovsky, A. G., & Veeravalli, V. V. (1999). Multihypothesis sequential probability ratio tests – part I: asymptotic optimality. *IEEE Transactions on Information Theory, 45*, 2448-2461.

[13] Durbin, J. (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. *Journal of Applied Probability, 8*, 431-453.

[14] Dzhafarov, E.N. (1993). Grice-representability of response time distribution families. *Psychometrika, 58*, 281-314.

[15] Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Vol. 1.* New

York: Wiley.

[16] Filimon, F., Philiastides, M. G., Nelson, J. D., Kloosterman, N. A., Heekeren, H. R. (2013). How embodied is perceptual decision making? Evidence for separate processing of perceptual and motor decisions. *Journal of Neuroscience, 33*, 2121-2136.

[17] Fortet, R. (1943). Les fonctions aléatoires du type de Markoff associées á certaines équations linéaires aux derivées partielles du type parabolique. *Journal de Mathématiques Pures et Appliquées.*

[18] Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review, 75*, 359-373.

[19] 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.

[20] 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.

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

[22] Hanes D. P., & Schall J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*, 427–430.

[23] Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology, 3*, 292.

[24] Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (in press). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review.*

[25] Jones, M., Mozer, M. C., & Kinoshita, S. (2009). Optimal response initiation: Why recent experience matters. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems, 21*, 785-792.

[26] Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*, 1292–1298.

[27] Laming, D. R. J. (1968). *Information Theory of Choice Reaction Time.* New York: Wiley.

[28] Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40*, 77–105.

[29] Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization.* New York: Oxford University Press.

[30] Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*, 335–384.

[31] 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.

[32] Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman.

[33] Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex, 13*, 1257-1269.

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

[35] Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*, 376-404.

[36] Pólya, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Mathematische Zeitschrift, 8*, 171-181.

[37] Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108.

[38] Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review, 120*, 281-292.

[39] Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 142–165.

[40] Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347-356.

[41] Ratcliff, R. & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333-367.

[42] Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116*, 59-83.

[43] Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16*, 323-341.

[44] Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review, 9*, 438-481.

[45] Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261-300.

[46] Ricciardi, L. M., & Sato, S. (1988). First-passage-time density and moments of the Ornstein-Uhlenbeck process. *Journal of Applied Probability, 25*, 43-57.

[47] Roe, R. M., Busemeyer, J., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic artificial neural network model of decision making. *Psychological Review, 108*, 370-392.

[48] Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal

area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22*, 9475-9489.

[49] Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44*, 408–463.

[50] Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*, 10393–10398.

[51] Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variability and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology, 64*, 1-34.

[52] Stone, M. (1960). Models for choice reaction time. *Psychometrika, 25*, 251–260.

[53] Thomas, R.D. (2006) Processing time predictions of current models of perception in the classic additive factors paradigm. *Journal of Mathematical Psychology, 50*, 441-455.

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

[55] Townsend, J. T., & Altieri, N. (2012). An accuracy-response time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review, 119*, 500-516.

[56] Townsend, J. T., Houpt, J. W., & Silbert, N. H. (2012). General recognition theory extended to include response times: Predictions for a class of parallel systems. *Journal of Mathematical Psychology, 56*, 476-494.

[57] Townsend, J. T. & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*, 1003-1035.

[58] Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*, 550–592.

Appendix A: Proofs of Theorems

Recall the definition of joint hazard function in Equation 2.5. Analogously, given a random variable T with density function $f(t)$ and (cumulative) distribution function $F(t)$, the hazard function is defined as

$$h(t) = \begin{cases} \frac{f(t)}{1-F(t)} & \text{if } F(t) < 1 \\ \\ 0 & \text{if } F(t) = 1. \end{cases} \tag{A.1}$$

Theorem 3 (Universality of independent race models)

We begin by observing that the marginal hazard function for RT is related to the joint hazard function by

$$h^{s,c}(\cdot,t) = \frac{g^{s,c}(\cdot,t)}{1-G^{s,c}(\cdot,t)} = \frac{\sum_r g^{s,c}(r,t)}{1-G^{s,c}(\cdot,t)} = \sum_r h^{s,c}(r,t) \tag{A.2}$$

with $h^{s,c}(\cdot,t) = h^{s,c}(r,t) = 0$ if $t > t^{s,c}_{\max}$. It follows that, for all t,

$$G^{s,c}(\cdot,t) = 1 - \exp\left(-\int_0^t h^{s,c}(\cdot,\tau)\,\mathrm{d}\tau\right) = 1 - \prod_r \exp\left(-\int_0^t h^{s,c}(r,\tau)\,\mathrm{d}\tau\right), \tag{A.3}$$

and

$$g^{s,c}(r,t) = h^{s,c}(r,t)\left(1-G^{s,c}(\cdot,t)\right) = h^{s,c}(r,t)\prod_i \exp\left(-\int_0^t h^{s,c}(i,\tau)\,\mathrm{d}\tau\right). \tag{A.4}$$

Let $F^{s,c}_r(t)$ be the (cumulative) distribution functions for $T^{s,c}_r$ according to the model M, and let $f^{s,c}_r(t)$ and $\nu^{s,c}_r(t)$ be the corresponding probability density and hazard functions, respectively. The probability density of process r winning the race at moment t is given by

$$\gamma^{s,c}(r,t) = f^{s,c}_r(t)\prod_{i\neq r}\left(1-F^{s,c}_i(t)\right) = \nu^{s,c}_r(t)\prod_i\left(1-F^{s,c}_i(t)\right). \tag{A.5}$$

We have to prove that $\gamma^{s,c}(t) = g^{s,c}(t)$ if and only if

$$F^{s,c}_r(t) = 1 - \exp\left(-\int_0^t h^{s,c}(r,\tau)\,\mathrm{d}\tau\right) \tag{A.6}$$

for $t < t^{s,c}_{\max}$.

First, assume Equation A.6 holds for $t < t^{s,c}_{\max}$. From Equation A.4, if $\int_0^t h^{s,c}(i,\tau)\,\mathrm{d}\tau = \infty$ for any i and t, then $g^{s,c}(r,t') = 0$ for all r and all $t' > t$. Therefore $\int_0^t h^{s,c}(i,\tau)\,\mathrm{d}\tau$ must be finite for all $t < t^{s,c}_{\max}$. Equation A.6 then implies that for $t < t^{s,c}_{\max}$, $F^{s,c}_r(t) < 1$ and

$$\nu_r^{s,c}(t) = \frac{\mathrm{d}F_r^{s,c}(t)/\mathrm{d}t}{1 - F_r^{s,c}(t)} = h^{s,c}(r,t). \tag{A.7}$$

Using Equation A.6 in Equation A.5 and substituting $h^{s,c}(r,t)$ for $\nu_r^{s,c}(t)$,

$$\gamma^{s,c}(r,t) = h^{s,c}(r,t) \prod_i \exp\left(-\int_0^t h^{s,c}(i,\tau)\,\mathrm{d}\tau\right). \tag{A.8}$$

By Equation A.4, this means

$$\gamma^{s,c}(r,t) = g^{s,c}(r,t). \tag{A.9}$$

To prove the converse, assume $\gamma^{s,c}(r,t) = g^{s,c}(r,t)$, and substitute $g^{s,c}$ for $\gamma^{s,c}$ in Equation A.5 to obtain

$$g^{s,c}(r,t) = \nu_r^{s,c}(t) \prod_i \left(1 - F_i^{s,c}(t)\right). \tag{A.10}$$

Since by assumption the model M is an independent race representation for $G^{s,c}$, $F_i^{s,c}(t) < 1$ for $t < t_{\max}^{s,c}$ and

$$\prod_i \left(1 - F_i^{s,c}(t)\right) = 1 - G^{s,c}(\cdot,t). \tag{A.11}$$

Therefore for $t < t_{\max}^{s,c}$,

$$\nu_r^{s,c}(t) = \frac{g^{s,c}(r,t)}{1 - G^{s,c}(\cdot,t)} = h^{s,c}(r,t), \tag{A.12}$$

and Equation A.6 follows.

Theorem 4 (Incomplete termination times)

In this theorem s and c are fixed, so they are omitted from the notation. We begin by constructing an independent race model

$$M = \left((R_1(t),\theta_1),\ldots,(R_n(t),\theta_n)\right)$$

whose first-passage times T_r are incomplete for $r \in \{i_1,\ldots,i_m\}$ and complete otherwise $(0 \le m < n)$. Let all θ_r be independent and uniformly distributed between 0 and 1, and let

$$R_r(t) = p_r \cdot \left(1 - e^{-t}\right), \tag{A.13}$$

where $p_r < 1$ if $r \in \{i_1,\ldots,i_m\}$ and $p_r = 1$ otherwise. The model's first-passage distributions are then given by

$$\Pr[T_r \le t] = R_r(t). \tag{A.14}$$

For $r \in \{i_1,\ldots,i_m\}$, $\Pr[T_r \le t]$ increases to $\lim_{t\to\infty} R_r(t) = p_r < 1$; that is, the first-passage times are incomplete. For $r \notin \{i_1,\ldots,i_m\}$, $\Pr[T_r \le t]$ increases to 1, reaching it

asymptotically at $t = \infty$. This completes the construction of M. Denote by $G(r,t)$ the R&T distribution generated by M, and by $h(r,t)$ the associated hazard function. Clearly, $t_{\max} = \infty$ here.

Assume M' is any other independent race model generating $G(r,t)$, and let T'_r be the random variables for its first-passage times. From Theorem 3, we have for all r and t

$$\Pr\left[T'_r \le t\right] = 1 - \exp\left(-\int_0^t h(r,\tau)\mathrm{d}\tau\right) = \Pr\left[T_r \le t\right]. \tag{A.15}$$

Therefore the first-passage times for M' are incomplete for $r \in \{i_1,\ldots,i_m\}$ and complete for $r \notin \{i_1,\ldots,i_m\}$, as claimed.

Theorem 7 (Universality of general LBA)

Let $F_r^{s,c}(t)$ be the (cumulative) distribution functions for first-passage times defined by Equation 2.11. From Theorem 3, the gLBA will reproduce $G^{s,c}(r,t)$ if it has first-passage distributions equal to $F_r^{s,c}(t)$. Under the assumptions of Theorem 7, the first-passage times for the gLBA are given by Equation 2.18. Define the gLBA's growth-rate distribution functions as

$$\Pr\left[k_r^{s,c} \le k\right] = \begin{cases} 1 - F_r^{s,c}\left(\left\lfloor \frac{b}{k} + t_0 \right\rfloor\right) & \text{if } k > 0 \\ 1 - \lim_{t\to\infty} F_r^{s,c}(t) & \text{if } k = 0, \end{cases} \tag{A.16}$$

extending them below $k = 0$ arbitrarily. (The $\lfloor\cdot\rfloor$ notation indicates the left limit: $F(\lfloor t\rfloor) = \sup_{\tau<t} F(\tau)$.) Then the first-passage distributions for any $t > t_0$ are equal to

$$\begin{aligned} \Pr\left[T_r^{s,c} \le t\right] &= \Pr\left[\frac{b}{k_r^{s,c}} + t_0 \le t \text{ and } k_r^{s,c} > 0\right] \\ &= \Pr\left[k_r^{s,c} \ge \frac{b}{t - t_0}\right] = F_r^{s,c}(t). \end{aligned} \tag{A.17}$$

Theorem 8 (Universal ballistic models with random growth rates)

As with Theorem 7, we need only to show that the model can generate any given set of first-passage distributions, $F_r^{s,c}(t)$. The first-passage time for response r here is equal to the value of t for which $k_r^{s,c} L_r(t) = b$. Therefore

$$T_r^{s,c} = \begin{cases} L_r^{-1}\left(\frac{b}{k_r^{s,c}}\right) & \text{if } k_r^{s,c} > \frac{b}{e_r} \\ \infty & \text{if } k_r^{s,c} \le \frac{b}{e_r} \end{cases}, \tag{A.18}$$

where $e_r = \lim_{t\to\infty} L_r(t)$. If $L_r(t)$ is unbounded we put $e_r = \infty$ and $b/e_r = 0$. Let the distribution functions for growth rates $k_r^{s,c}$ be

$$\Pr\left[k_r^{s,c} \le k\right] = \begin{cases} 1 - F_r^{s,c}\left(\left\lfloor L_r^{-1}\left(\frac{b}{k}\right)\right\rfloor\right) & \text{if } k > \frac{b}{e_r} \\ 1 - \lim_{t\to\infty} F_r^{s,c}(t) & \text{if } k = \frac{b}{e_r} \end{cases}, \tag{A.19}$$

arbitrarily extended below $k \le \frac{b}{e_r}$. The first-passage distributions are then given by

$$\begin{aligned} \Pr\left[T_r^{s,c} \le t\right] &= \Pr\left[L_r^{-1}\left(\frac{b}{k_r^{s,c}}\right) \le t \text{ and } k_r^{s,c} > \frac{b}{e_r}\right] \\ &= \Pr\left[k_r^{s,c} \ge \frac{b}{L_r(t)}\right] = F_r^{s,c}(t)\,. \end{aligned} \tag{A.20}$$

Theorem 9 (Universal ballistic models with random starting points)

As with the previous two proofs, we need only to show that the model can generate any given set of first-passage distributions, $F_r^{s,c}(t)$. The first-passage time here is

$$T_r^{s,c} = \begin{cases} L_r^{-1}\left(b - z_r^{s,c}\right) & \text{if } z_r^{s,c} > b - e_r \\ \infty & \text{if } z_r^{s,c} \le b - e_r \end{cases}, \tag{A.21}$$

where $e_r = \lim_{t\to\infty} L_r(t)$. If $L_r(t)$ is unbounded, $b - e_r$ is replaced with $-\infty$. Let the start-point distribution functions be

$$\Pr\left[z_r^{s,c} \le z\right] = \begin{cases} 1 - F_r^{s,c}\left(\left\lfloor L_r^{-1}\left(b - z\right)\right\rfloor\right) & \text{if } z > b - e_r \\ 1 - \lim_{t\to\infty} F_r^{s,c}(t) & \text{if } z = b - e_r \end{cases}, \tag{A.22}$$

extended below $z = b_r - e_r$ arbitrarily. Then the first-passage time distributions are given by

$$\begin{aligned} \Pr\left[T_r^{s,c} \le t\right] &= \Pr\left[L_r^{-1}\left(b - z_r^{s,c}\right) \le t \text{ and } z_r^{s,c} > b - e_r\right] \\ &= \Pr\left[z_r^{s,c} \ge b - L_r(t)\right] = F_r^{s,c}(t)\,. \end{aligned} \tag{A.23}$$

Remark. Theorems 8 and 9 can be easily generalized in several respects: The continuity and nonnegativity constraints on $L_r(t)$ can be relaxed. One can allow the thresholds to have different values for different responses. All quantities being fixed at arbitrarily chosen values, by the very arbitrariness of their choice, can be made arbitrarily dependent on c. For instance, $b_1, \ldots, b_n$ in Theorem 9 can be made $b_1^c, \ldots, b_n^c$ with the stipulation $\lim_{t\to\infty} L_r(t) \le \sup_c b_r^c$.

Theorem 10 (Universality of general Wiener diffusion model)

Given any family of R&T distributions, $G^{s,c}(r,t)$ for $r = 1, 2$, define the growth-rate distribution for each condition by

$$\Pr\left[k^{s,c} \leq k\right] = \begin{cases} G^{s,c}\left(1, T_{\text{er}} - \frac{a}{2k}\right) & \text{if } k < 0 \\ \lim_{t\to\infty} G^{s,c}\left(1,t\right) & \text{if } k = 0\,. \\ 1 - G^{s,c}\left(2, \left\lfloor T_{\text{er}} + \frac{a}{2k}\right\rfloor\right) & \text{if } k > 0 \end{cases} \tag{A.24}$$

Using Equations 2.20 and 2.21, we have for any $t > T_{\text{er}}$:

$$\begin{aligned} \Pr\left[r = 1 \text{ and RT} \leq t\right] &= \Pr\left[k^{s,c} < 0 \text{ and } T_{\text{er}} - \frac{a}{2k^{s,c}} \leq t\right] \\ &= \Pr\left[k^{s,c} \leq -\frac{a}{2\left(t - T_{\text{er}}\right)}\right] = G^{s,c}(1,t), \end{aligned} \tag{A.25}$$

and

$$\begin{aligned} \Pr\left[r = 2 \text{ and RT} \leq t\right] &= \Pr\left[k^{s,c} > 0 \text{ and } T_{\text{er}} + \frac{a}{2k^{s,c}} \leq t\right] \\ &= \Pr\left[k^{s,c} \geq \frac{a}{2\left(t - T_{\text{er}}\right)}\right] = G^{s,c}(2,t). \end{aligned} \tag{A.26}$$

Remark. The theorem is proved under the assumption that $G^{s,c}$ is globally complete (i.e., a response is given on every trial with probability 1). Then $\Pr\left[k^{s,c} = 0\right] = 1 - \lim_{t\to\infty} G^{s,c}\left(\cdot, t\right) = 0$. If global incompleteness is allowed, then the expression for $\Pr\left[k^{s,c} \leq 0\right]$ in Equation A.24 should be changed to $1 - \lim_{t\to\infty} G^{s,c}\left(2,t\right)$, with no further consequences.

Theorem 11 (Universality of general Ornstein-Uhlenbeck model)

Under the simplifying assumptions of the theorem, the response process for the OU model is given by

$$R^{s,c}(t) = \frac{a}{2} + \frac{k^{s,c}}{\beta}\left(1 - e^{-\beta\left(t - T_{\text{er}}\right)}\right). \tag{A.27}$$

The response and RT are given by

$$r = \begin{cases} 1 & \text{if } k^{s,c} < -\frac{\beta a}{2} \\ 2 & \text{if } k^{s,c} > \frac{\beta a}{2} \\ \text{undefined} & \text{if } -\frac{\beta a}{2} \leq k^{s,c} \leq \frac{\beta a}{2} \end{cases} \tag{A.28}$$

and

$$\text{RT} = \begin{cases} T_{\text{er}} - \frac{1}{\beta}\log\left(1 - \frac{\beta a}{2\left|k^{s,c}\right|}\right) & \text{if } \left|k^{s,c}\right| > \frac{\beta a}{2} \\ \infty & \text{if } \left|k^{s,c}\right| \leq \frac{\beta a}{2}. \end{cases} \tag{A.29}$$

Given any family of R&T distributions $G^{s,c}(r,t)$, define the distribution functions for growth rates as

$$\Pr\left[k^{s,c} \leq k\right] = \begin{cases} G^{s,c}\left(1, T_{\text{er}} - \frac{1}{\beta}\log\left(1+\frac{\beta a}{2k}\right)\right) & \text{if } k < -\frac{\beta a}{2} \\ 1 - G^{s,c}\left(2, \left\lfloor T_{\text{er}} - \frac{1}{\beta}\log\left(1-\frac{\beta a}{2k}\right)\right\rfloor\right) & \text{if } k > \frac{\beta a}{2}, \end{cases} \tag{A.30}$$

arbitrarily interpolated on the interval $|k| \leq \frac{\beta a^c}{2}$. (Under the global completeness assumption, the interpolated portion is a constant equal to $1 - \lim_{t\to\infty} G^{s,c}(\cdot,t)$.) Then the general OU model (with $\sigma = \delta_{\text{z}} = \delta_{\text{t}} = 0$) can be verified to reproduce $G^{s,c}(r,t)$. For any $t > T_{\text{er}}$:

$$\begin{aligned} \Pr\left[r = 1 \text{ and } \text{RT} < t\right] &= \Pr\left[k^{s,c} < -\frac{\beta a}{2} \text{ and } T_{\text{er}} - \frac{1}{\beta}\log\left(1+\frac{\beta a}{2k^{s,c}}\right) \leq t\right] \\ &= \Pr\left[k^{s,c} \leq -\frac{\beta a}{2\left(1-e^{-\beta(t-T_{\text{er}})}\right)}\right] = G^{s,c}(1,t), \end{aligned} \tag{A.31}$$

and

$$\begin{aligned} \Pr\left[r = 2 \text{ and } \text{RT} < t\right] &= \Pr\left[k^{s,c} > \frac{\beta a}{2} \text{ and } T_{\text{er}} - \frac{1}{\beta}\log\left(1-\frac{\beta a}{2k^{s,c}}\right) \leq t\right] \\ &= \Pr\left[k^{s,c} \geq \frac{\beta a}{2\left(1-e^{-\beta(t-T_{\text{er}})}\right)}\right] = G^{s,c}(2,t). \end{aligned} \tag{A.32}$$

Remark. Note that this result holds for $\beta < 0$ (positive feedback) as well as for $\beta > 0$ (decay), even though the OU model is generally taken as restricted to the latter case.

Theorem 12 (Universality with nonzero diffusion)

Let $G(r,t)$ be any R&T distribution that is continuous in time, and let β be any real number. By Theorem 11, there exists a gDM model M_0 with $\sigma = 0$, $a = 1$, $z = \frac{1}{2}$, and decay rate β such that M_0 generates $G(r,t)$. Let k be the random growth rate assumed by M_0, and let $R_0(t) = \frac{1}{2} + kt$ be the associated stochastic response process.

We put $r = 1$, with $r = 2$ being considered analogously. By the definition of M_0, the following relation holds:

$$G(1,t) = \Pr\left[\text{response}_{M_0} = 1 \text{ and } RT_{M_0} \leq t\right] = \Pr\left[\min\left\{x : R_0(x) \leq 0\right\} \leq \min\left(t, \min\left\{x : R_0(x) \geq a\right\}\right)\right]. \tag{A.33}$$

Let $W(t)$ be an OU process with zero drift, diffusion rate equal to 1, and decay rate β. Choose a sequence of OU processes $\sigma_n W(t)$ with $\sigma_n \to 0$, and note these processes all have decay rate β. Consider

$$G_n(1,t) = \Pr\left[\min\left\{x : R_n(x) \leq 0\right\} \leq \min\left(t, \min\left\{x : R_n(x) \geq a\right\}\right)\right], \tag{A.34}$$

where

$$R_n(t) = R_0(t) + \sigma_n W(t). \tag{A.35}$$

It can easily be shown that

$$G_n(1,t) \to G(1,t) \tag{A.36}$$

for any $t \geq 0$, and that

$$G_n(1,\infty) \to G(1,\infty). \tag{A.37}$$

By an extension of Pólya's theorem on pointwise convergence of distribution functions (Pólya, 1920), the convergence is uniform; that is,

$$\sup_{t\geq 0} |G_n(1,t) - G(1,t)| \to 0. \tag{A.38}$$

Combining now over both responses, the uniform convergence implies that for any $\epsilon > 0$ one can find a σ_n such that

$$\sup_{r,t\geq 0} |G_n(r,t) - G(r,t)| < \epsilon. \tag{A.39}$$

Now define a gDM model M_n with $\sigma = 1$, $a = 1/\sigma_n$, $z = a/2$, decay rate β, and random growth rate k/σ_n. The stochastic response process in this model is equal to $R_n(t)/\sigma_n = aR_n(t)$, and thus M_n produces the R&T distribution $G_n(r,t)$. The model M_n therefore satisfies the theorem.

Appendix B: Predictions and Grice Representation of Diffusion Model

The predicted R&T distribution and density of the diffusion model were calculated following the method of Smith (2000). To simplify notation, the response process on each trial is redefined by $\widetilde{R}(t) = R(t) - z$, so that the starting point is $\widetilde{R}(0) = 0$ and the new thresholds are $b_1 = -z$ and $b_2 = a - z$. We also suppress superscripts s and c in this section. With this notation, the unconstrained transition function of the response process (i.e., ignoring whether it has or has not crossed either decision threshold) can be shown to equal

$$\begin{aligned} f(x,t|y,\tau) &= \frac{\mathrm{d}}{\mathrm{d}x}\Pr\left[\widetilde{R}(t) \leq x | \widetilde{R}(\tau) = y\right] \\ &= \sqrt{\frac{\beta}{\pi\sigma^2\left(1 - e^{-2\beta(t-\tau)}\right)}} \cdot \exp\left(\frac{-\beta\left[x - ye^{-\beta(t-\tau)} - \frac{k}{\beta}\left(1 - e^{-\beta(t-\tau)}\right)\right]^2}{\sigma^2\left(1 - e^{-2\beta(t-\tau)}\right)}\right) \end{aligned} \tag{B.1}$$

or, when $\beta = 0$,

$$f(x,t|y,\tau) = \sqrt{\frac{1}{2\pi\sigma^2(t-\tau)}} \cdot \exp\left(\frac{-\left(x - y - k(t-\tau)\right)^2}{2\sigma^2(t-\tau)}\right). \tag{B.2}$$

The next step is to relate the predicted R&T density, $g(r,t)$, to the transition function using a renewal equation (Durbin, 1971; Fortet, 1943). This equation expresses the total probability of transitioning from $(0,0)$ to (b_r,t), for $r=1$ or 2, in terms of the first-passage distribution multiplied by the transition probability from the first-passage point to (b_r,t):

$$f\left(b_r,t|0,0\right)=\int_0^t g(1,\tau)f\left(b_r,t|b_1,\tau\right)\mathrm{d}\tau+\int_0^t g(2,\tau)f\left(b_r,t|b_2,\tau\right)\mathrm{d}\tau. \tag{B.3}$$

Because of instabilities that arise in numerically solving Equation B.3, it is transformed into the following expression (Buonocore, Nobile, & Ricciardi, 1987):

$$\begin{aligned} g\left(1,t\right) &= -2\Psi\left(b_1,t|0,0\right)+2\int_0^t g\left(1,\tau\right)\Psi\left(b_1,t|b_1,\tau\right)\mathrm{d}\tau+2\int_0^t g\left(2,\tau\right)\Psi\left(b_1,t|b_2,\tau\right)\mathrm{d}\tau \\ g\left(2,t\right) &= 2\Psi\left(b_2,t|0,0\right)-2\int_0^t g\left(1,\tau\right)\Psi\left(b_2,t|b_1,\tau\right)\mathrm{d}\tau-2\int_0^t g\left(2,\tau\right)\Psi\left(b_2,t|b_2,\tau\right)\mathrm{d}\tau \end{aligned} \tag{B.4}$$

with the new integration kernel, Ψ, given by

$$\Psi\left(x,t|y,\tau\right)=\frac{1}{2}f\left(x,t|y,\tau\right)\left[\beta x-k-\frac{2e^{-\beta(t-\tau)}}{1-e^{-2\beta(t-\tau)}}\left(e^{\beta(t-\tau)}\left(\beta x-k\right)-\beta y+k\right)\right], \tag{B.5}$$

or, when $\beta=0$,

$$\Psi\left(x,t|y,\tau\right)=-\frac{f\left(x,t|y,\tau\right)\left(x-y\right)}{2(t-\tau)}. \tag{B.6}$$

Following Ratcliff and Smith (2004), Equation B.4 was integrated numerically using a step size of $\mathrm{d}\tau=.01$ (i.e., 10 ms). The resulting R&T density was then integrated over the distributions of z, k, and t_0 to obtain the predicted R&T density of the full model.

The Grice representation for the diffusion model was obtained using Equation 4.1, with the diffusion model's joint hazard function $h(r,t)$ obtained from $g(r,t)$ by numerical intega-tion. Because of the nested integration in Equations 4.1 and B.4, accumulation of rounding errors produced artificial cusps and nonmonotonicities in $R_r(t)$ for extreme values of t. These minor numerical errors are negligible in standard analyses of diffusion models, which focus directly on RT distributions and quantiles, but they cause problems in the present analysis. Therefore, to eliminate rounding error, the tail of $g(r,t)$ for each response r within each stimulus and condition (or parameter setting) was replaced by an exact exponential func-tion. Plots of $\log\left(g(r,t)\right)$ showed extended regions of near-perfect linearity in the tails of the RT distributions, indicating the predicted RT distributions have asymptotically expo-nential tails. This property can be proven analytically when k, z, and t_0 are constant and there is only one threshold (Ricciardi & Sato, 1988), and this graphical technique showed exponentiality to be an excellent approximation for the full model. For all of the analyses reported, the linearity of $\log\left(g(r,t)\right)$ only broke down for extreme values of t, at the point where rounding error from limits on machine precision came into play. To eliminate these

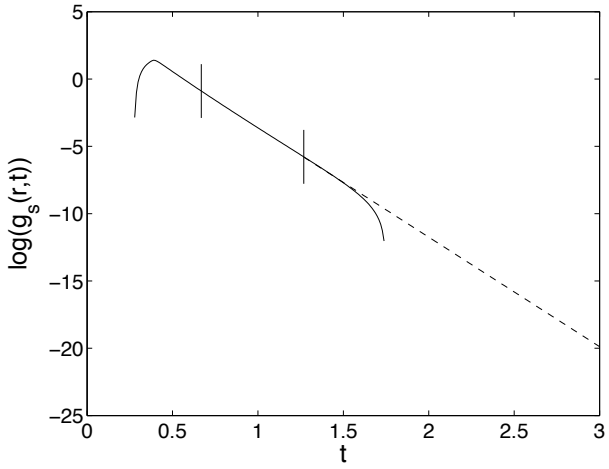

Figure 10: Illustration of correction to tail of reaction time (RT) distributions derived from diffusion models. Solid curve indicates log-transformed joint R&T density for correct response, $g(2,t)$, as approximated by numerical integration. After 1.74 s, the approximated density becomes negative due to rounding error. Vertical black lines indicate a conservative region within which $\log\left(g(2,t)\right)$ is very nearly linear ($R^2 = .99998$). Dashed line indicates extrapolation of this linear range, which was used to replace the tail of $g(2,t)$ with an exact exponential function, in defining the corrected distribution. Model parameters for this example are taken from Ratcliff and Smith's (2004) fit of the Wiener model to Ratcliff et al.'s (2001) Experiment 2, speed condition, stimulus level 3.

effects of rounding error, the linear portion of $\log\left(g(r,t)\right)$ was extrapolated to cover the remaining tail of the RT distribution (Figure 10).

Appendix C: Deriving Grice Representation of LBA Model

Brown & Heathcote (2008) derive the cumulative first-passage distribution of each response process in the standard LBA by integrating over the growth rate (k_r^s) and the starting point (z_r^c):

$$\Pr\left[T_r^{s,c} \leq t + t_0\right] = 1 - \frac{b^c - tv_r^s}{A^c}\Phi\left(\frac{b^c - tv_r^s}{t\eta}\right) + \frac{b^c - A^c - tv_r^s}{A^c}\Phi\left(\frac{b^c - A^c - tv_r^s}{t\eta}\right) - \frac{t\eta}{A^c}\phi\left(\frac{b^c - tv_r^s}{t\eta}\right) + \frac{t\eta}{A^c}\phi\left(\frac{b^c - A^c - tv_r^s}{t\eta}\right) \quad \text{(C.1)}$$

When $A^c = 0$, the distribution reduces to

$$\Pr\left[T_r^{s,c} \leq t + t_0\right] = 1 - \Phi\left(\frac{b^c - tv_r^s}{t\eta}\right). \quad \text{(C.2)}$$

In a Grice model with independent unit-exponential threshold distributions, the first-

passage distributions are given by

$$\Pr\left[T_r^{s,c} \leq t\right] = 1 - e^{-R_r^{s,c}(t)}. \tag{C.3}$$

This expression assumes the response processes are nondecreasing, which can always be made true by substituting $\max_{\tau \leq t}\left\{R_r^{s,c}(\tau)\right\}$ for $R_r^{s,c}(t)$ (Dzhafarov, 1993). Combining Equations C.2 and C.3 yields the response processes with which the Grice model (with independent unit-exponentially distributed thresholds) mimics the LBA. For $A^c > 0$,

$$\begin{aligned} R_r^{s,c}(t+t_0) = -\log\Bigg[&\tfrac{b^c - tv_r^s}{A^c}\Phi\left(\tfrac{b^c - tv_r^s}{t\eta}\right) - \tfrac{b^c - A^c - tv_r^s}{A}\Phi\left(\tfrac{b^c - A^c - tv_r^s}{t\eta}\right) \\ &+ \tfrac{t\eta}{A^c}\Phi\left(\tfrac{b^c - tv_r^s}{t\eta}\right) - \tfrac{t\eta}{A^c}\Phi\left(\tfrac{b^c - A^c - tv_r^s}{t\eta}\right)\Bigg], \end{aligned} \tag{C.4}$$

and for $A^c = 0$,

$$R_r^{s,c}(t+t_0) = -\log\left[\Phi\left(\frac{b^c - tv_r^s}{t\eta}\right)\right] \tag{C.5}$$

with $R_r(t) = 0$ for $t \leq t_0$.

Appendix D: Grice Representation of Discrete Data

Assume we are given a finite number of response-RT pairs, with all RTs distinct. Let $t_{r1} < \ldots < t_{rn_r}$ be the ordered sample of RTs paired with response r. We add to this sequence a dummy element $0 = t_{r0} < t_{r1}$.

Let T_r be the first-passage times for the Grice representation of the data. It is clear that T_r is discretely distributed, with nonzero probability at each t_{ri} $(i > 0)$ and zero probability elsewhere (except perhaps for $t > t_{\max}$). The probability that no response has occurred before time t is given by

$$\Pr\left[RT \geq t\right] = \prod_r \Pr\left[T_r \geq t\right] \tag{D.1}$$

and the probability of response r occurring at time t is given by

$$\Pr\left[\text{response} = r \text{ and } RT = t\right] = \Pr\left[T_r = t\right] \cdot \prod_{r' \neq r} \Pr\left[T_{r'} > t\right] = \Pr\left[T_r = t\right] \cdot \prod_{r' \neq r} \Pr\left[T_{r'} \geq t\right] \tag{D.2}$$

(since all response times are distinct, if $\Pr\left[T_r = t\right] > 0$ then $\Pr\left[T_{r'} = t\right] = 0$ for all $r' \neq r$). Combining Equations D.1 and D.2 yields a relationship between the empirical hazard functions for the first-passage distributions and what we refer to as the empirical joint hazard function for R&T, $\hat{h}\left(r,t\right)$. For all $t \leq t_{\max}$:

$$\frac{\Pr\left[T_r = t\right]}{\Pr\left[T_r \geq t\right]} = \frac{\Pr\left[\text{response} = r \text{ and } RT = t\right]}{\Pr\left[RT \geq t\right]} = \hat{h}\left(r,t\right). \tag{D.3}$$

Let $F_r(t)$ be the (cumulative) distribution function for T_r, a staircase function with jumps at t_{ri} for $i \geq 1$. With a unit-exponential threshold distribution, the response process generating $F_r(t)$ is

$$\begin{aligned} R_r(t) &= -\log\left(1 - F_r(t)\right) \\ &= -\sum_{t_{ri} \leq t} \log \frac{1 - F_r(t_{ri})}{1 - F_r\left(t_{r(i-1)}\right)} \\ &= -\sum_{t_{ri} \leq t} \log\left(1 - \frac{\Pr\left[T_r = t_{ri}\right]}{\Pr\left[T_r \geq t_{ri}\right]}\right) \\ &= -\sum_{t_{ri} \leq t} \log\left(1 - \hat{h}\left(r, t_{ri}\right)\right). \end{aligned} \tag{D.4}$$

This last formula is what we used in computing the Grice representation of the empirical data, separately for each stimulus and instruction condition.