Combining speed and accuracy to assess error-free cognitive processes

Mark E. Glickman[1]
Department of Health Services
Boston University

Jeremy R. Gray
Department of Psychology
Yale University

Carlos J. Morales
Department of Mathematics and Statistics
Worcester Polytechnic Institute

[1]Address for correspondence: Center for Health Quality, Outcomes & Economics Research , Edith Nourse Rogers Memorial Hospital (152), Bldg 70, 200 Springs Road, Bedford, MA 01730, USA. E-mail address: mg@bu.edu. Phone: (781) 687-2875. Fax: (781) 687-3106. The authors wish to thank Roger Ratcliff and three anonymous referees for their helpful comments, and Aureliu Lavric for providing the data analyzed in this paper.

Abstract

Both the speed and accuracy of responding are important measures of performance. A well-known interpretive difficulty is that participants may differ in their strategy, trading speed for accuracy, with no change in underlying competence. Another difficulty arises when participants respond slowly and inaccurately (rather than quickly but inaccurately), e.g., due to a lapse of attention. We introduce an approach that combines response time and accuracy information and addresses both situations. The modeling framework assumes two latent competing processes. The first, the error-free process, always produces correct responses. The second, the guessing process, results in all observed errors and some of the correct responses (but does so via non-specific processes, e.g., guessing in compliance with instructions to respond on each trial). Inferential summaries of the speed of the error-free process provide a principled assessment of cognitive performance reducing the influences of both fast and slow guesses. Likelihood analysis is discussed for the basic model and extensions. The approach is applied to a data set on response times in a working memory test.

Keywords: Competing risks, reaction times, speed/accuracy tradeoff, time-to-event model, working memory tasks.

1 Introduction

Objective measures of human performance have proven to be widely useful, e.g., in human resource management, competitive sports, and basic psychological research. The speed and the accuracy of responding are two important metrics of performance (Luce, 1986; Townsend and Ashby, 1983). For example, on the basis of these two measures and additional assumptions, it has been possible to draw strong inferences about cognition, the structure of human information processing (e.g., Sternberg, 1969). In the current article, we develop a latent-variable approach to combining speed and accuracy information to make these measures even more useful.

In a typical cognitive psychology experiment, a particular cognitive process is isolated by having participants perform both a task designed to require the process of interest and a closely matched control task that does not require it. Under different experimental conditions, both the speed and the accuracy of performance are then measured (often scored by a computer), usually on multiple trials within each condition to increase the stability of the estimates. For example, the effect of pleasant versus unpleasant emotional states (experimental conditions) on verbal versus nonverbal working memory (cognitive processes) can be assessed in this way (Gray, 2001). Participants are typically instructed to respond as quickly and accurately as possible, which is intended to keep participants from unduly emphasizing accuracy over speed or vice versa. Within each condition, accuracy depends on the number of errors, and speed is taken to be the mean of response times on each trial (sometimes assessed on correct trials only). A common procedure is to trim response times that are 2.5 standard deviations longer than the mean (as being unlikely to result from the process of interest) and that are implausibly quick (e.g., likely to have resulted from an accidental key press).

To infer differences in competence based on measures of performance, both speed and accuracy must be examined. In general, it is hoped that performance will vary by condition in the same direction for both speed (faster = better performance) and accuracy (more accurate = better performance), or on one but not the other. We highlight two recurrent issues that the latent-variable approach will help address.

First, participants sometimes sacrifice accuracy to enhance their speed of responding, or vice versa, known as a speed-accuracy tradeoff (SAT). Fast but inaccurate responding is not easily comparable to slow but accurate responding, creating an interpretive dilemma. Differences between conditions or individuals could reflect a difference in strategy or motivation, rather than a difference in performance or ability. If there is no SAT, inferences are easy to draw because both speed and accuracy suggest the same interpretation (faster and more accurate performance implies enhanced performance, slower and less accurate performance implies impaired performance). If there is a SAT, it is difficult to conclude anything about the effect of the experimental manipulation on performance, as distinct from effects on strategy, motivation, and so on. Even if there is no SAT at a group level, individual participants can vary in their emphasis on speed versus accuracy (Dickman and Meyer, 1988), complicating the interpretation of individual differences.

The main way psychologists address SAT is by experimental design. The usual approach is simply to give instructions to respond as quickly and accurately as possible. This helps constrain participants to be more similar to each other in their relative emphasis on speed and accuracy, and is usually successful at avoiding speed-accuracy tradeoffs at a group level. Nonetheless, individuals can still differ in their relative emphasis on one or the other, despite the objectively identical instructions. Another method is to assess the SAT function at several points by explicitly manipulating the tradeoff through instructions, incentives, or other means, in order to assess a response surface (Wickelgren, 1977). This is the most conceptually satisfactory method for addressing SAT (Luce, 1986), but is labor-intensive and not widely practiced. An alternative method is to give a response signal at some point during a trial, intended to force a decision at that point in time in order to estimate the accumulation of partial information upon which a decision could be based (e.g., Smith, Kounios, and Osterhout, 1997).

A second, more subtle issue arises when participants respond more slowly on error trials than on correct trials. The issue here concerns the estimate of the speed of the process of interest. If mean response time is taken across all trials (i.e., including error trials), response times that are unlikely to reflect the process of interest are included in the estimate. Using response times only from correct trials is more of a pure measure of speed because it eliminates the error trials. Nonetheless, some correct responses could be produced for the wrong reason (e.g., guessing) and these would still be included, biasing the estimate of response speed of the process of interest. In the extreme case, a participant may be guessing on every trial, so that the mean response time has no relation to the cognitive process of interest (participants performing at or near chance are typically excluded from analysis). Even for situations not as extreme, the true error rate (i.e., the failure rate of the process of interest) could be substantially higher than the number of errors actually committed. Response times on spuriously correct trials cannot be assumed to reflect the process of interest. Including response times from such trials in the overall estimate of the speed of the process of interest without proper adjustment would bias the estimate. The estimate could be either too low (if there are spuriously correct fast-guesses, SAT situation) or too high (if there are spuriously correct slow-guesses). This issue could be particularly germane in studies of individual differences, in which accurate assessment of quantitative differences is especially important.

In sum, the typical procedure in psychological research is to instruct participants to respond as quickly and accurately as possible. Inferences about ability are then based on measures of performance as indexed by accuracy and response time, taken separately. While this is useful and widely practiced (e.g., Gray, 1999, 2001), it has several short-comings. It is potentially vulnerable to SAT interpretive dilemmas, although steps can be taken to reduce the likelihood of such a situation at a group level. Usual practice also neglects the possibility that some correct responses are produced spuriously by a process other than the one of interest, biasing the estimate of speed and accuracy. Using only mean response time ignores potentially interesting information in the distributions (e.g., Heathcote, Popiel and Mewhort, 1991; Ratcliff, 1979). Thus potentially valuable information in behavioral data is often unused (i.e., the shape of the distribution around the mean) or estimates are potentially biased by non-specific influences (e.g., fast or slow guessing).

In this article, we present a latent-variables approach to analyzing behavioral data that addresses these issues. Our approach assumes the existence of two competing latent stochastic processes. The first, the error-free process, reflects a cognitive process of interest and is assumed always to produce correct responses. The goal of statistical inference in this setting is to summarize features of the error-free process. The second competing latent process reflects the need to respond despite a failure to engage or sustain the error-free process, for whatever reason. This can be the result of overly quick and therefore inaccurate responding (fast guesses, SAT situation), or can lead to extended processing times without enhanced accuracy (e.g., slow guesses, due to lapses of attention). Observed responses under this second process are assumed to be produced at random, and to have a probability of being correct that reflects the experimental design (i.e., if there are 50% targets then the guessing process has a 0.5 probability of producing a correct response). Our approach requires that both speed and accuracy are available for each of multiple trials within a condition, and can be applied to studies involving single cases, individual differences, and group data.

The strength of our statistical model is best understood when a large number of errors occur. At the extreme, if a participant answers at random, which in a SAT situation might occur if responses are supplied too quickly, then clearly no level of competence is being demonstrated, and our model recognizes that the data contain no information about cognitive competence even though many correct responses may be observed. When a participant is able to produce all correct responses, our model reduces to fitting a standard model to all of the reaction times.

Some caveats about our approach are worth mentioning at the outset. Our framework is intended as a statistical model for making inferences about high-level cognitive processes, not as a general psycho-physical model of response times. In this regard, we do not claim the psycho-physical existence of error-free and guessing processes beyond their statistical meaning. The concept of an error-free process is an idealized one, but provides a statistical baseline of comparison across task conditions and experimental participants. General models for response times often account for perceptual judgments over very wide ranges of accuracy (e.g., Ratcliff and Rouder, 1998, 2000; Ratcliff, Van Zandt & McKoon, 1999). Also, our model is not intended to be a theory of human decision making, and its usefulness does not derive from being such a theory. Rather, our model is adequately faithful to aggregate processes underlying human response times in cognitive studies, sufficiently so to make it competitive to standard methods for analyzing speed and accuracy data. In cognitive (as opposed to perceptual) studies, performance is not usually experimentally manipulated to be at chance levels, but rather is typically desired to be reasonably good but not perfect.

The framework of our model is introduced formally in Section 2. This is followed in Section 3 by a proof of the identifiability of the distribution of response times under the latent processes. We then discuss in Section 4 parametric modeling via a likelihood analysis, and, in particular, modeling the latent response time processes as a three-parameter Weibull distribution. The parametric analysis is applied in Section 5 to group and individual data from an experiment on the effect of induced emotion on working memory performance. A discussion of the utility of the model, along with proposed extensions, is presented in Section 6.

2 A statistical model for response times

Consider an experiment where a participant performs multiple trials of a task. We assume that trials are independent (see Section 6 for a discussion of this assumption), and that the participant is asked to perform the trials as quickly and as accurately as possible. Let Y be an observed response time for a trial, and let

$$D = \begin{cases} 1 & \text{if response on the trial is correct} \\ 0 & \text{if response on the trial is incorrect} \end{cases} .$$

The basic assumption of our model is the existence of two competing latent stochastic processes for an observed response. One of the processes, the error-free process, results in a correct response with probability 1. The other process, the guessing process, results in a correct response with probability η, $0 < \eta < 1$, where η is specified by design. The guessing process is understood to produce responses that might result from random guesses. For example, if a trial involves the participant identifying one out of five items, then by design η would be set to 0.2. Such guesses are initiated by a latent process that reflects the instructions to make a response on each trial, even for trials on which the error-free process either has not yet run to completion (SAT situation) or has been disrupted (slow-guess situation). That is, compliant participants are under pressure to respond, even when they lack a fully-adequate basis for choosing a particular response. We model this pressure to respond non-specifically as a latent process, one that competes in a race with the error-free process for the control of behavior on a given trial.

Let

$$\begin{aligned} T &= \text{Error-free response time for the trial} \\ R &= \text{Guessing response time for the trial} \end{aligned}$$

Our response time model assumes that both error-free and guessing processes occur simultaneously within a participant. The observed response time for a trial in our modeling framework is assumed to be the minimum of the two latent response times,

$$Y = \min(T,\ R).$$

When $T < R$, then the response for the trial will be correct with probability 1, that is

$$\mathrm{P}(D = 1 \mid T < R) = 1.$$

If the converse, then we have

$$\mathrm{P}(D = 1 \mid T \geq R) = \eta.$$

Equality of the two process times is assumed to occur on a set of zero probability, so that it is inconsequential how to treat a tie in our model.

It is worth noting that T and R are not observed quantities, but can be identified in certain cases. If, for example, a trial results in an incorrect response, then under our modeling assumptions the observation must have been produced by the guessing process, so that the observed time was a guessing response time.

By treating the observed response time as the result of a "race" between an error-free process and a guessing process, incorrect responses are explicitly accounted for in a model-based manner. In fact, our modeling framework allows for the possibility that the error-free process occurs sooner than the guessing process as well as vice versa, even within the same participant. Figure 1 displays four plausible reaction time scenarios. In each plot, the solid lines represent the probability density for the error-free time, T, and the dotted lines represent the density for the guessing time, R. The hypothetical densities are all 3-parameter Weibull models with the same shape parameter, but different location and shift parameters. For easy tasks, that is, tasks for which error-free reaction times are generally lower than guessing times, most of the probability density of T is to the left of the density of R, examples of which are represented in the upper two density plots in Figure 1. When the easy task is characterized by error responses that are slower than correct responses, the

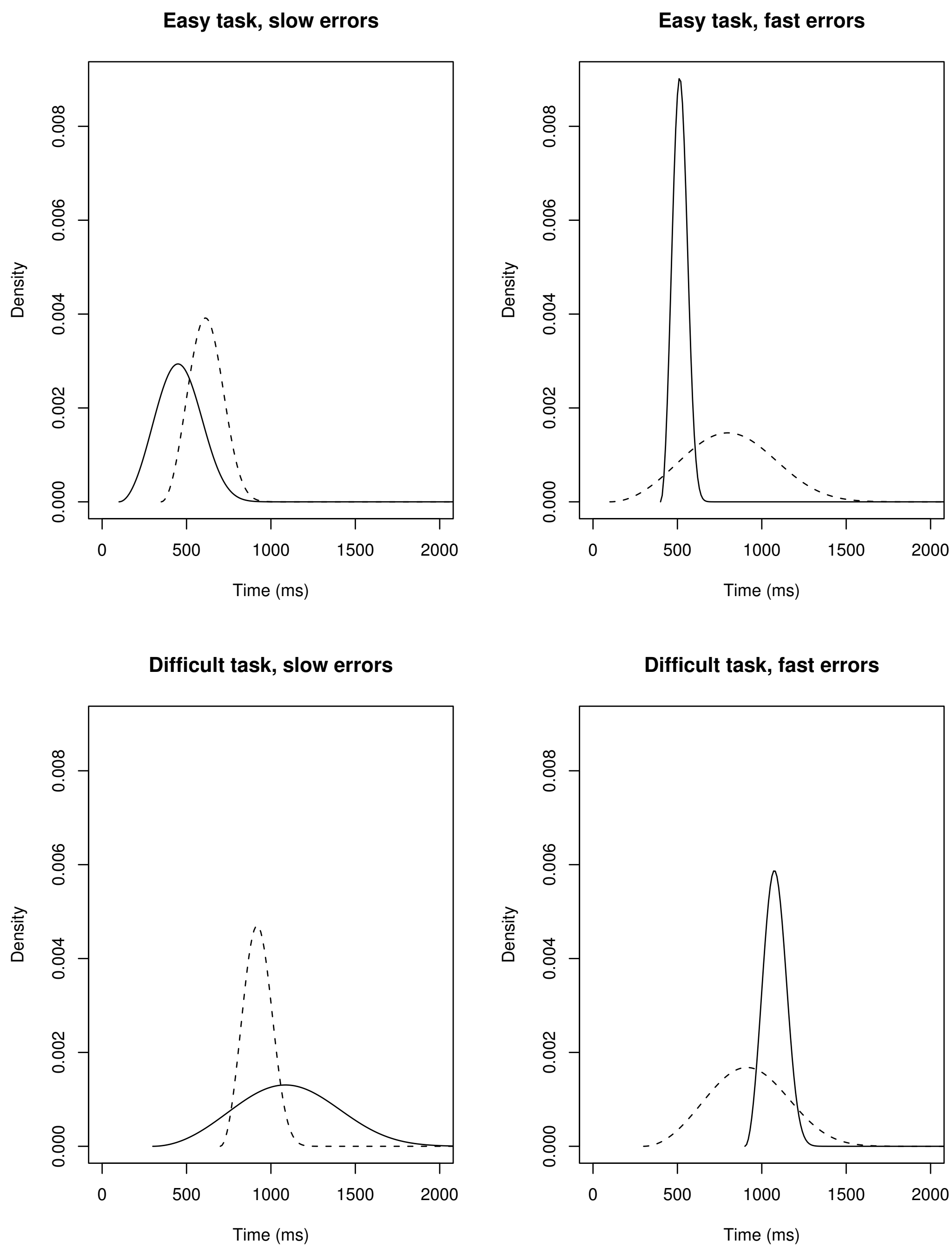

Figure 1: Underlying theoretical error-free and guessing reaction times for easy and difficult tasks, with slow and fast observed error reaction times. The solid lines correspond to the error-free response time densities, and the dotted lines correspond to the guessing response time densities. All distributions are 3-parameter Weibull models.

entire density of T is shifted to the left relative to that of R (upper left plot). However, when the easy task involves error responses that are quicker than correct responses, the left tail of the density of R extends further to the left than that of T, accounting for the quicker reaction times when the guessing process wins the "race."

Reaction times for difficult tasks can be understood analogously. When a task is difficult, the density of R has most of its mass to the left of the density of T (see the two bottom plots in Figure 1). For difficult tasks in which error reaction times are quicker than times for correct responses (bottom right plot), the density of T is shifted to the right relative to the density of R. Difficult tasks involving slower error reaction times than times for correct responses (bottom left plot), the left tail of the density of T is extended further to the left than that of R. Again, even though most of the mass of R is to the left of T (which means that observed reaction times will usually be produced by the guessing process), error-free times, when observed, will often come from the left tail of T, resulting in low mean observed correct reaction times relative to error reaction times. As noted by Ratcliff and colleagues (Ratcliff and Tuerlinckx, 2002; Ratcliff, Thapar and McKoon, 2001; Ratcliff and Rouder, 1998), easy tasks are associated with low error rates and often with errors that are quicker than correct responses. They have also observed an association between difficult tasks with slow errors. As demonstrated in Figure 1, these empirical findings can therefore be accounted for in our modeling framework.

Table 1 summarizes data that might result from the latent error-free and guessing distributions shown in Figure 1. Under the assumption of a two-choice experiment with equal frequency of responses (so that $\eta = 0.5$), values of T and R were simulated 50,000 times in pairs to determine the observed reaction time, followed by simulated correctness of responses (for cases where $R < T$). As shown in the table, easy tasks were associated with high correctness rates (above 90%), and difficult tasks were associated with low correctness rates (above 60%, but not much above randomness, i.e.,

Task Condition	Error Speed	Percent Correct	Mean RT (ms) Correct responses	Mean RT (ms) Incorrect responses
Easy	Slow	92	434	536
Easy	Fast	93	507	415
Difficult	Slow	65	850	912
Difficult	Fast	63	924	826

Table 1: Summaries of simulated data corresponding to the four scenarios in Figure 1. Calculations were based on 50,000 simulations from 3-parameter Weibull models.

50%). The observed reaction times for error responses are, on average, quicker or slower than the times for correct responses depending on the relationship between latent error-free and guessing distributions.

It should also be noted that, in some experiments, participants are exposed to easy tasks and difficult tasks in the same sequence of trials. This can be accommodated in our framework explicitly by recognizing the different task conditions, and stratifying the analysis (or, more generally, incorporating a treatment covariate in a model) by the different levels of the task conditions. In doing so, differing relationships between the error-free and guessing distributions can be inferred for different experimental conditions.

For some participants, the guessing process may be associated generally with longer response times. This may happen when a participant constantly experiences lapses in attention, letting the left tail of the guessing response time distribution be to the right of the error-free response time distribution. In other situations, a participant may feel pressure to respond quickly even if he/she is not ready to perform the cognitive task. This might result in quicker guessing response times than error-free times. Our framework is flexible enough to infer both types of situations, as well as the ones observed by Ratcliff and his colleagues. One restriction on our modeling framework, however, is that these two types of situations do not occur across trials within a single participant when there is no change in the experimental setting.

The goal of inference in our framework is to estimate features of the error-free response time distribution. In studies involving multiple participants, or experiments with multiple treatments, important summaries, such as means of the error-free response time distributions, can be compared meaningfully across participants and treatments to assess inter-participant variability, or treatment effects. Randomness in response times is relegated to the guessing process, so that the error-free process contains the information relevant to measuring treatment effects on response times.

Our framework has connections to previous work, including the foundational Simple Fast Guess (SFG) model developed by Ollman (1966) and Yellott (1971). Similar to our basic model, the SFG model assumes a mixture of two processes, and interest lies in inferring features of the mean response time for one of them. Unlike our approach, however, the SFG model assumes merely a mixture of two processes, but not a "race." Thus, in the SFG model, the occurrence of either type of response depends only on a random mechanism, and applies only for fast guesses, not slow guesses as well.

3 Identifiability of the latent reaction time distributions

We show in this section that the latent error-free response time and guessing distributions can be inferred directly from the data without imposing assumptions on either the error-free or guessing reaction time distributions. Let $f_T(t)$ and $f_R(t)$ denote the unknown probability densities of the error-free and guessing time distributions, respectively. Our goal, therefore, is to demonstrate in this section that $f_T(t)$ and $f_R(t)$ are identified from observed information.

Let Y and D generically denote the observed response time and correctness of a response to a trial. We first note that the following information is directly identifiable from a sample of response time data: $\mathrm{P}(D = d)$ for $d = 0, 1$; $\mathrm{S}_Y(t) = \mathrm{P}(Y > t)$; and $\mathrm{S}_{Yd}(t) = \mathrm{P}(Y > t \mid D = d)$ for $d = 0, 1$. The latter two functions are the marginal survival function for Y, and the survival function of Y conditional on $D = d$. Because the number of observations with $D = 0$ is usually small relative to the number of observations with $D = 1$, $\mathrm{S}_{Y0}(t)$ can be expressed in terms of quantities that are typically more precisely inferred:

$$\mathrm{S}_Y(t) = \mathrm{P}(D=1)\mathrm{S}_{Y1}(t) + \mathrm{P}(D=0)\mathrm{S}_{Y0}(t)$$

so that

$$\mathrm{S}_{Y0}(t) = \frac{\mathrm{S}_Y(t) - \mathrm{P}(D=1)\mathrm{S}_{Y1}(t)}{\mathrm{P}(D=0)}.$$

To identify $\mathrm{P}(T < R)$, we note that

$$\begin{aligned} \mathrm{P}(D=1) &= \mathrm{P}(T<R)\mathrm{P}(D=1 \mid T<R) + \mathrm{P}(T \geq R)\mathrm{P}(D=1 \mid T \geq R) \\ &= \mathrm{P}(T<R) \cdot 1 + (1 - \mathrm{P}(T<R))\eta \end{aligned}$$

so that

$$\mathrm{P}(T<R) = \frac{\mathrm{P}(D=1) - \eta}{1-\eta}.$$

This then leads to the identification of $\mathrm{P}(T < R \mid D = 1)$ by recognizing

$$\mathrm{P}(T<R \mid D=1) = \frac{\mathrm{P}(D=1 \mid T<R)\mathrm{P}(T<R)}{\mathrm{P}(D=1)} = \frac{\mathrm{P}(T<R)}{\mathrm{P}(D=1)}.$$

Let $S_{RT}(t) = \mathrm{P}(Y > t \mid T \geq R)$, and $S_{TR}(t) = \mathrm{P}(Y > t \mid T < R)$. We first note that $S_{RT}(t) = S_{Y0}(t)$. This can be seen as follows.

$$\begin{aligned} &S_{Y0}(t) = \mathrm{P}(Y>t \mid D=0) \\ &= \mathrm{P}(T<R \mid D=0)\mathrm{P}(Y>t \mid D=0, T<R) + \mathrm{P}(T \geq R \mid D=0)\mathrm{P}(Y>t \mid D=0, T \geq R) \\ &= \mathrm{P}(Y>t \mid D=0, T \geq R) = \mathrm{P}(Y>t \mid T \geq R). \end{aligned} \tag{1}$$

The last line in (1) follows because $\mathrm{P}(T < R \mid D = 0) = 0$, and that $Y > t$ is conditionally independent of $D = 0$ given $T \geq R$.

To obtain $S_{TR}(t)$, we note

$$\mathrm{P}(Y>t \mid D=1)$$

$$
\begin{aligned}
&= \mathrm{P}(T < R \mid D = 1)\mathrm{P}(Y > t \mid D = 1, T < R) + \mathrm{P}(T \geq R \mid D = 1)\mathrm{P}(Y > t \mid D = 1, T \geq R) \\
&= \frac{\mathrm{P}(T < R)}{\mathrm{P}(D = 1)}\mathrm{P}(Y > t \mid D = 1, T < R) + \frac{\mathrm{P}(D = 1) - \mathrm{P}(T < R)}{\mathrm{P}(D = 1)}\mathrm{P}(Y > t \mid D = 1, T \geq R) \\
&= \frac{\mathrm{P}(T < R)\mathrm{P}(Y > t \mid T < R) + [\mathrm{P}(D = 1) - \mathrm{P}(T < R)]\mathrm{P}(Y > t \mid D = 0)}{\mathrm{P}(D = 1)}
\end{aligned}
$$

The last equality recognizes that $Y > t$ is conditionally independent of $D = 1$ given $T < R$ or $T \geq R$ because the correctness of a response given which latent process occurred first provides no extra information on the distribution of the minimum of T and R, so that

$$\mathrm{P}(Y > t \mid D = 1, T \geq R) = \mathrm{P}(Y > t \mid T \geq R) = \mathrm{P}(Y > t \mid D = 0),$$

and

$$\mathrm{P}(Y > t \mid D = 1, T < R) = \mathrm{P}(Y > t \mid T < R).$$

Solving for $S_{TR}(t) = \mathrm{P}(Y > t \mid T < R)$ yields

$$S_{TR}(t) = \frac{\mathrm{P}(D = 1)\mathrm{P}(Y > t \mid D = 1) - [\mathrm{P}(D = 1) - \mathrm{P}(T < R)]\mathrm{P}(Y > t \mid D = 0)}{\mathrm{P}(T < R)}. \qquad (2)$$

Let $F_R(t)$ be the distribution function for the guessing response time, and let $F_T(t)$ be the distribution function for the error-free response time. Recall that $f_R(t)$ and $f_T(t)$ denote the corresponding density functions. Also let $S_R(t) = 1 - F_R(t)$ and $S_T(t) = 1 - F_T(t)$. Similar to the derivation in Nádas (1970), we now note that

$$\mathrm{P}(t < T < R) = \mathrm{P}(T < R)\mathrm{P}(Y > t \mid T < R) = \int_t^\infty S_R(x) f_T(x)\, dx. \qquad (3)$$

Taking derivatives with respect to t, we obtain

$$\mathrm{P}(T < R) f_Y(t \mid T < R) = S_R(t) f_T(t).$$

Noting that

$$S_Y(t) = S_R(t) S_T(t),$$

we have

$$\mathrm{P}(T < R) f_Y(t \mid T < R) = h_T(t) S_Y(t)$$

where

$$h_T(t) = \frac{f_T(t)}{S_T(t)},$$

the hazard function for T. We therefore have

$$h_T(t) = \frac{\mathrm{P}(T < R) f_Y(t \mid T < R)}{S_Y(t)}. \tag{4}$$

All terms on the right-hand side of (4) are identified; the term $f_Y(t \mid T < R)$ is the negated derivative of $S_{TR}(t)$, which has already been identified. The hazard function for T can be converted into the distribution function through

$$F_T(t) = 1 - \exp\left(-\int_0^t h_T(x)\, dx\right).$$

This proves that the distribution of T, along with all parameters of the distribution, are identifiable.

Similarly to (3), if we consider

$$\mathrm{P}(t < R < T) = \mathrm{P}(R < T)\mathrm{P}(Y > t \mid R < T) = \int_t^\infty S_T(x) f_R(x)\, dx,$$

then taking derivatives with respect to t yields

$$\mathrm{P}(R < T) f_Y(t \mid R < T) = h_R(t) S_Y(t),$$

implying

$$h_R(t) = \frac{\mathrm{P}(R < T) f_Y(t \mid R < T)}{S_Y(t)}.$$

Because all terms on the right are identified, the unconditional distribution of R is identified as well.

4 Likelihood analysis and Weibull modeling

By making reasonable assumptions on the response time distributions, a parametric likelihood-based analysis can be performed to obtain inferences about parameters of interest. An advantage of parametric or even semi-parametric models in this context is the ability to incorporate covariate information into the model, which we demonstrate in this section.

An important consequence of the identifiability of both latent variable distributions is that each can be parameterized separately without any parameters in common. A clear advantage for mutually exclusive sets of parameters is that the effects of covariates on the latent response times can be assumed completely distinct. The tradeoff is that inferences are not going to be precise. From simulation analyses, it is evident that a large number of trials is necessary to make reasonably precise inferences about features of the two latent distributions without assuming shared parameters. The approach we describe in the ensuing discussion assumes a model with some shared features.

Suppose a study involves n conditionally independent trials given model parameters. For $i = 1, \ldots, n$, let y_i be the observed response time, and let d_i be an indicator for correctness of trial i. The likelihood is therefore

$$L = \prod_{i=1}^{n} \left(p_i f_T(y_i) S_R(y_i) 0^{1-d_i} + (1-p_i) f_R(y_i) S_T(y_i) \eta^{d_i} (1-\eta)^{1-d_i} \right)$$

where $p_i = \mathrm{P}(T_i < R_i)$. The contribution of each term involves the weighted average of the density of (y_i, d_i) when $T_i < R_i$, and when $T_i \geq R_i$. The likelihood can be written in a more convenient form by indexing the product of values of the d_i:

$$\begin{aligned} L &= \prod_{\{i:d_i=0\}} (1-p_i) f_R(y_i) S_T(y_i) (1-\eta) \times \\ & \prod_{\{i:d_i=1\}} (p_i f_T(y_i) S_R(y_i) + (1-p_i) f_R(y_i) S_T(y_i) \eta) \end{aligned} \tag{5}$$

A choice of probability models for the response times that lends itself to straightforward likelihood analysis is the 3-parameter Weibull distribution. Other parametric models for response times are clearly possible, though the 3-parameter Weibull model lends itself to a tractable analysis, and is appropriate for reaction time data. Besides a shape and location parameter, the 3-parameter Weibull distribution assumes a shift parameter that sets the minimum (positive) value of a reaction time. With no shift parameter, and when only response times are observed, the competing risks Weibull model is the so-called "bi-Weibull" model (see, for example, Berger and Sun, 1993, 1996). In our situation, extra information (correctness of response) is available. It is interesting to note that response correctness only provides partial information about the identity of the process observed, whereas in typical competing risks problems either the identity is always unknown or is completely known. Assume the T_i and the R_i are modeled according to

$$\begin{aligned} S_T(y) &= \exp(-\lambda_T(y-\delta)^\alpha) \\ S_R(y) &= \exp(-\lambda_R(y-\delta)^\alpha) \end{aligned}$$

where λ_T and λ_R are location parameters for the separate Weibull models, α is the common shape parameter, and δ is the common shift parameter, so that

$$\begin{aligned} f_T(y) &= \alpha\lambda_T(y-\delta)^{\alpha-1}\exp(-\lambda_T(y-\delta)^\alpha) \\ f_R(y) &= \alpha\lambda_R(y-\delta)^{\alpha-1}\exp(-\lambda_R(y-\delta)^\alpha) \end{aligned}$$

for $y \geq \delta$. The assumption of a common shift parameter, δ, does preclude two of the four scenarios displayed in Figure 1, so that this assumption is a strong restriction. However, a crude examination of the experimental data to check whether the task is easy or difficult (via the rate of correctness), and whether errors are quick or slow, will provide some evidence of whether the assumption of a common shift parameter is plausible. Under the model with a single shift and shape parameter,

$$p_i = \mathrm{P}(T_i < R_i) = \int_\delta^\infty \mathrm{P}(y < R) f_T(y)\, dy$$

$$
\begin{aligned}
&= \int_{\delta}^{\infty} \exp(-\lambda_R(y-\delta)^{\alpha})\alpha\lambda_T(y-\delta)^{\alpha-1}\exp(-\lambda_T(y-\delta)^{\alpha})\, dy \\
&= \left(\frac{\lambda_T}{\lambda_T+\lambda_R}\right)\int_{\delta}^{\infty} \alpha(\lambda_T+\lambda_R)(y-\delta)^{\alpha-1}\exp(-(\lambda_T+\lambda_R)(y-\delta)^{\alpha}))\, dy \\
&= \frac{\lambda_T}{\lambda_T+\lambda_R}
\end{aligned}
$$

It is also worth noting that the distribution of Y conditional on the parameters λ_T, λ_R, α and δ is also Weibull, with

$$
S_Y(y) = \exp(-(\lambda_T+\lambda_R)(y-\delta)^{\alpha}) \tag{6}
$$

because Y is the minimum of two independent Weibull variables with the same shift and shape parameters. From (5), the likelihood for Weibull latent response times becomes

$$
\begin{aligned}
L &= \prod_{\{i:d_i=0\}} \left(\frac{\lambda_R}{\lambda_T+\lambda_R}\right)\alpha\lambda_R(y_i-\delta)^{\alpha-1}\exp(-\lambda_R(y_i-\delta)^{\alpha})\exp(-\lambda_T(y_i-\delta)^{\alpha})(1-\eta)\times \\
&\quad \prod_{\{i:d_i=1\}} \left[\left(\frac{\lambda_T}{\lambda_T+\lambda_R}\right)\alpha\lambda_T(y_i-\delta)^{\alpha-1}\exp(-\lambda_T(y_i-\delta)^{\alpha})\exp(-\lambda_R(y_i-\delta)^{\alpha})\right. \\
&\qquad \left. +\left(\frac{\lambda_R}{\lambda_T+\lambda_R}\right)\alpha\lambda_R(y_i-\delta)^{\alpha-1}\exp(-\lambda_R(y_i-\delta)^{\alpha})\exp(-\lambda_T(y_i-\delta)^{\alpha})\eta\right] \\
&= \left(\frac{\alpha}{\lambda_T+\lambda_R}\right)^{n}\lambda_R^{2n_0}(1-\eta)^{n_0}(\lambda_T^2+\lambda_R^2\eta)^{n_1}\prod_{i=1}^{n}(y_i-\delta)^{\alpha-1}\exp(-(\lambda_T+\lambda_R)(y_i-\delta)^{\alpha}) \qquad (7)
\end{aligned}
$$

where n_0 is the number of trials for which $d_i = 0$, and $n_1 = n - n_0$. This likelihood can be maximized with respect to the unknown parameters $(\lambda_T, \lambda_R, \alpha, \delta)$ using standard numerical procedures, such as the Newton-Raphson algorithm applied to the derivative of the log-likelihood.

The basic model can be extended to include covariate information that relates to response times, and to account for differences among participants. The ability to incorporate covariates in our framework is essential because this provides a mechanism to model the effects of experimental manipulations, such as measuring the effect of induced anxiety, or measuring the effect of making a task more difficult to perform. If speed-accuracy instructions are manipulated in an experimental setting, the response time distributions can reflect the manipulations by incorporating terms in each of the latent response time distributions. In fact, assuming separate effects in the error-free and guessing distributions by participant allows for a variety of possible outcomes, including error-free times that are shorter than guessing times for some levels of an experimental treatment, and vice versa for other levels. Response bias, where certain responses are rewarded more heavily than others, can be incorporated in a similar fashion.

Let x_{im} be a vector of p covariates for trial i of participant m, $m = 1, \ldots, M$, and let $\beta = (\beta_1, \ldots, \beta_p)$ be a p-vector of covariate effects. In the Weibull model, we let

$$\begin{aligned} \ln \lambda_{Tim} &= \beta_{0m} + x'_{im}\beta \\ \ln \lambda_{Rim} &= \beta_{0m} + x'_{im}\beta + \gamma_{im} \end{aligned} \tag{8}$$

be the location parameters for the error-free and guessing processes for trial i, participant m, where β_{0m} is the intercept for participant m, and γ_{im} is the parameter that distinguishes the error-free and guessing processes. Without any dependence on covariates, we can set $\gamma_{im} = \gamma_m$, that is, have the effect depend only on the participant. This assumption, however, restricts the types of data allowable by our model. For example, under this assumption, our model cannot be made consistent with data frequently observed in perceptual tasks where error-prone tasks have slow errors and nearly error-free tasks have fast errors. More generally, we can have

$$\gamma_{im} = \theta_{0m} + x'_{im}\theta_m$$

where θ_{0m} is a participant-specific intercept, and θ_m is (usually) a vector of the effects of a set of experimental conditions on trial i for participant m related through x_{im}. This parameterization, where the γ_{im} vary both by participant and experimental condition, addresses a wide variety of mechanisms that produce reaction-time data. Under this parameterization, inference for β, the effects of the experimental conditions, is of main interest. The γ_{im} can be treated as nuisance parameters that account for the randomness in the guessing process.

The likelihood in (7) can be maximized using the ECM algorithm (Meng and Rubin, 1993), an extension of the EM algorithm (Dempster, Laird and Rubin, 1977). The details of the implementation for our model appear in the Appendix.We also mention how the ECM algorithm can be adjusted to account for a model with different shift parameters for each latent reaction time per participant. The covariance matrix of the estimates can be found directly from the negative inverse of the Hessian of the observed-data log-likelihood in (7) evaluated at the maximum likelihood estimates. Alternatively, using both first and second derivative information from the complete-data log-likelihood evaluated at the maximum likelihood estimates, a by-product of the EM-algorithm, the covariance matrix of estimates can be obtained through methods described by Louis (1982).

5 Application to a working memory experiment

We applied the methods in Section 4 to data collected from an experiment on response times during moderately difficult cognitive tasks. The experiment investigated the influence of induced anxiety on working memory (Lavric, Rippon and Gray, 2003). Participants performed computerized verbal and spatial working memory tasks in the presence or absence of threat of mild electric shock. In addition to measuring response times and correctness of response, average heart-rate (beats per minute), an indicator of whether the task was verbal versus spatial, and an indicator of the threat of shock was recorded for each trial. Trials with missing response times (112 trials) were discarded from the analysis. The resulting number of trials per participant ranged from 102 to 124, with a total of 4353 trials across the 36 participants. The overall rate of correctness was 79%, which, with $\eta = 0.5$, corresponds to error-free times being quicker than guessing times ($\mathrm{P}(T < R) \approx 0.58$). The data indicate that correct responses are generally quicker than error responses, so that using the common shift parameter model in Section 4 is appropriate for this application. It is of interest to understand whether anxiety (threat of shock) affected working memory, and particularly whether it affected verbal and spatial working memory differentially.

Insight can be gained by examining the fit of our model to one participant. We fit our 3-

Stimulus	Fitted error-free time (msec)	Fitted guessing time (msec)	Fitted observed time (msec)	Observed mean time (msec)
Spatial task, no shock	1188	1346	1044	1094
Spatial task, shock	1932	1755	1501	1465
Verbal task, no shock	1786	1891	1502	1468
Verbal task, shock	2469	1822	1665	1664

Table 2: Fitted error free and guessing times, rounded to the nearest integer, for Participant 6 with a model that examined all four stimuli combinations.

Model	Number of Linear Parameters	Log-likelihood
`T ~ Participant`	72	−36373.73
`T ~ Participant + Shock`	109	−36269.33
`T ~ Participant + Task`	109	−36264.79
`T ~ Participant + Task + Shock`	146	−36157.25
`T ~ Participant + Task + Shock + Task:Shock`	177	−36105.09
`T ~ Participant + Task + Shock + Task:Shock + Heart`	178	−36100.04

Table 3: Summary of six models evaluated at the maximum likelihood estimates. Each model assumed that the guessing parameter was a function of the covariates.

parameter model to Participant 6, accounting for the four separate effects of each stimulus combination. We computed the estimated mean error-free and guessing times (in milliseconds) and summarized the results in Table 2. Because measures of competence in our model are contained in the error-free distribution, the estimated mean error-free times provide the comparison information we seek. In this case, it appears that without the threat of shock, this participant's error-free reaction times are quicker than with the threat of shock. This indicates that the participant may be more competent at the two tasks without the threat of shock. It is interesting to observe that the mean guessing times with the threat of shock are less than without the threat of shock, and vice versa for tasks without the threat of shock. Thus a greater frequency of errors are accompanied by prolonged reaction times, and this feature in the data is captured in our model.

We fit several models using different covariate combinations to these data. The fits of these models are summarized in Table 3. Linear terms for whether a trial involved the threat of shock

(`Shock`), whether the trial involved a verbal task (`Task`), the interaction of these two terms, and heart rate (`Heart`) were candidates for inclusion among the collection of covariates in the model. Because six of the participants had no incorrect responses for particular task-by-threat combinations, we assumed that the interaction effect in the model was the same for the error-free and guessing distributions for these six people. This simply means that we did not include an interaction term for λ_{im} for participants m where interactions could not be identified. The effect of heart rate, which was assumed to be the same across all participants, was also treated as having the same effect on the error-free and guessing distributions. Log-likelihoods were computed at the maximum likelihood estimates. Performing likelihood ratio tests sequentially on nested models via χ^2-tests on twice the difference of log-likelihoods, with degrees of freedom equal to the difference in the number of linear parameters, yields that the model involving the full interaction of `Task` and `Shock` along with a separate adjustment for heart rate is the best-fitting model among those we considered.

Table 4 summarizes the best fitting model by displaying the estimates and standard errors of the covariate effects, the Weibull exponent parameter, and the shift parameter. The negative estimates of β_{shock} and β_{task} indicate that the threat of shock corresponds to increased average response times, and that the verbal task has a higher average response time relative to the spatial task. The positive interaction parameter indicates that the average response time is decreased by the threat of shock combined with a verbal task. A similar effect was reported as the major finding in Lavric et al. (2003), although this was evident in accuracy and not response time. The positive estimate of heart-rate is associated with lower average reaction times for higher heart-rates. The estimated shift parameter of 215 msec is the time common to all participants that allows for the execution of the response. Other features of the model worth noting include the participant-specific intercepts varying between -19.4 to -15.1. With correspondingly small standard errors,

Model term	Estimate	Std Err
β_{shock}	–0.428	0.061
β_{task}	–0.247	0.056
$\beta_{shock:task}$	0.309	0.079
β_{heart}	0.013	0.0047
α	2.331	0.027
δ	214.90	1.93

Table 4: Estimates for linear parameters and the Weibull exponent parameter, along with standard errors.

this indicates noteworthy variation in baseline response times across participants, accounting for covariate adjustments.

It is also worth noting that summaries of participants and conditions can be provided in estimates of mean error-free response times. For our Weibull models, the mean error-free time is given by

$$\mathrm{E}(T_{ij}) = \delta + \Gamma\left(1 + \frac{1}{\alpha}\right)\exp(-\ln\lambda_{Tim}/\alpha). \tag{9}$$

Analogous expressions for the mean guessing time and mean observed time are obtained by replacing λ_{Tim} in (9) with λ_{Rim} and $\lambda_{Tim} + \lambda_{Rim}$, respectively. Substituting the parameters in (9) with estimates from the model fit yields estimated mean error-free response times. For example, for Participant 1, whose intercept estimate was -17.28, the estimated mean error-free response time (in milliseconds) with threat of shock, a heart rate of 90 beats per minute, which was observed for this treatment combination, and a spatial task is given by

$$\begin{aligned}\widehat{\mathrm{E}(T_{i1})} &= 214.90 + \Gamma\left(1 + \frac{1}{2.331}\right) \\ &\quad \times \exp(-(-17.28 + 1(-0.428) + 0(-0.247) + 1 \cdot 0(0.309) + 90(0.013))/2.331) \\ &= 1283.22\end{aligned}$$

By comparison, the mean observed response time for correct responses by this participant was 967.8 milliseconds. The higher latent error-free mean compared to the mean of correct responses reflects the assumption that longer error-free times were censored by the guessing process, so that the only observed error-free response times were those that were not censored.

Besides model selection using likelihood ratio tests, model checking can be performed graphically using residual plots. For ease of interpretation, we compute residuals as the difference between the logarithm of the observed reaction time and the logarithm of the fitted reaction time. Because fitted values can be obtained as described above, plots of residuals against covariates can reveal model inadequacies. As an example, Figure 2 shows a plot of heart rates against the residuals. The lack of pattern in the residuals as a function of heart rate indicates no clear lack of fit. Similar plots against the treatment variables (examining the distribution of residuals as boxplots for each treatment level) also indicate no obvious lack of fit.

6 Discussion

Research psychologists are often interested in characterizing the function of a particular mental process (e.g., spatial working memory) in different conditions and for different individuals. Psychometrically, it is of interest to estimate competence rather than simply summarize performance. That is, non-specific factors that influence performance, such as fast and slow guesses, are typically of less interest than differences in underlying ability or differences between experimental conditions. In our modeling framework, we capture this emphasis in terms of two latent processes, an error-free process and a guessing process. By using error information to adjust response time, the approach allows a more accurate assessment of the functioning of the process of interest despite non-specific effects attributable to fast or slow guessing. By summarizing inferences about the speed of a latent, error-free process we can compare participants against a more similar baseline. Furthermore, treatment effects can also be inferred based on the time required to produce a correct response.

Among the strengths of the approach we have described, perhaps foremost is that it yields a more accurate assessment of the specific process of interest, lessening the influence of non-specific

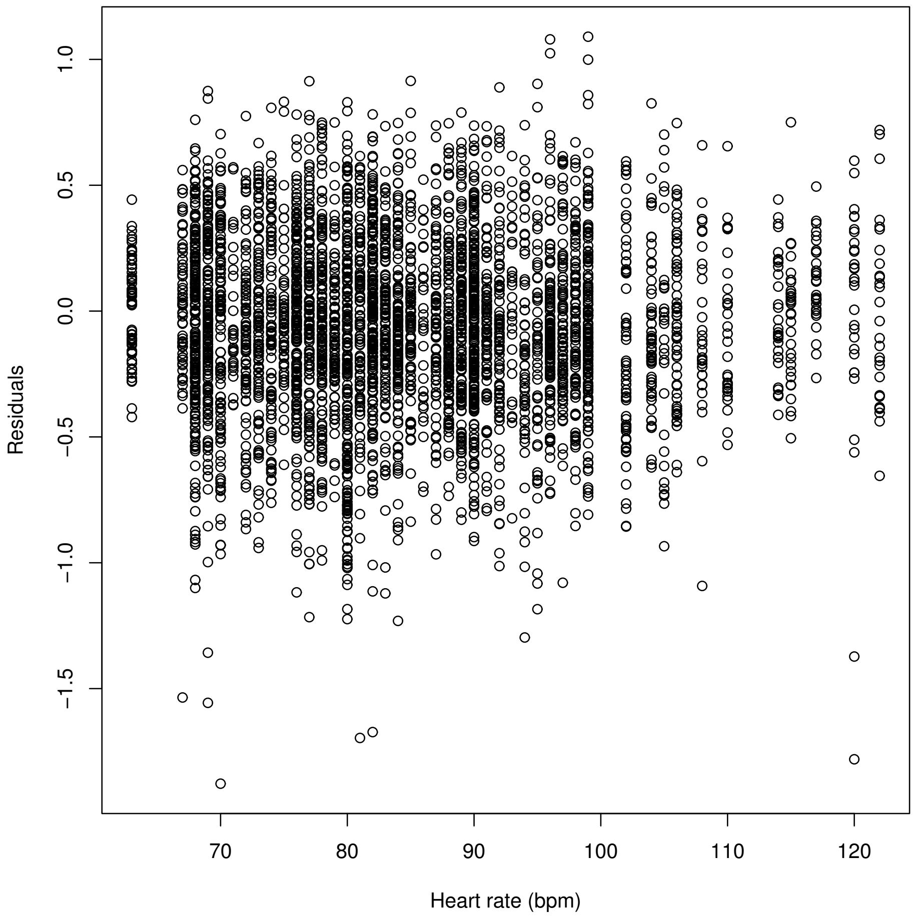

Figure 2: Plot of heart rates (beats per minute) against the difference between the logarithm of the observed reaction times and the logarithm of the fitted reaction times for the final model.

effects that can influence speed and accuracy. In particular, the approach addresses both speed-accuracy tradeoffs (fast guessing) as well as slow but inaccurate responding (e.g., due to momentary lapses of attention). Further, the flip slide of correcting for such effects at a group level is that such effects can be explicitly assessed in studies of individual differences. The approach can be used to derive a single parameter for each participant that indexes the degree and direction of the adjustment made to mean response time. Finally, the use of our modeling framework does not impose constraints on experimental design.

Our framework requires that the probability of a correct response in the guessing process, η, is specified by design. The parameter η can be treated as unknown, but then extra assumptions are required to identify the error-free distribution. This can be demonstrated by taking the derivative of $S_{TR}(t)$ with respect to t in (2) and substituting the expression in (4). This yields

$$h_T(t) = \frac{\mathrm{P}(D=1) f_Y(t|D=1) - [\mathrm{P}(D=1) - \mathrm{P}(T<R)] f_Y(t|D=0)}{S_Y(t)}. \tag{10}$$

Because $h_T(t)$ determines the error-free distribution, it suffices to show that the right-hand side of (10) depends on η. All of the terms in the expression depend only on data, and not on model parameters, except for $\mathrm{P}(T < R)$ which depends on η (see Section 3). For different values of η, the hazard function changes independently of the data (as represented by the other terms in the expression). A similar expression results for the hazard function of R, indicating that the value of η determines the distribution of R independently of data. Therefore, without any extra distributional assumptions, the specification of η is a necessary and sufficient condition for the identifiability of the latent response distributions. An alternative to specifying η is to impose restrictions on the latent response time distributions. One way to add restrictions is to assume a parametric model for the two processes and leave η as an unknown parameter. Unless η is not known by design (e.g., if the probability of a random correct guess is not known in advance in an experiment), we do not recommend this approach, as η will be partially aliased with other model parameters (typically with a function of the difference in location parameters between the error-free and guessing distributions), so that, in particular, the data will not provide unconfounded information about η.

For some experiments measuring response times, the probability of correctly guessing a response via the guessing process may be argued to vary across trials. For example, suppose an experiment requires a participant to type "f" with the left forefinger on a keyboard for one answer, and "j" with the right forefinger for a second answer. Further, suppose that 70% of the target answers correspond to typing "f." During the initial trials, a participant may equally likely type "f" or "j" when the guessing process produces a response. But in the later trials, after the participant has begun to learn that "f" is more likely the correct response, may begin to type "f" even when experiencing a lapse of attention. In this setting, our modeling framework is not appropriate for the analysis of response time data.

Our modeling approach is not appropriate for all behavioral data sets. For example, if accuracy is at near-chance levels for a given participant, then inferences about features of an error-free process cannot be made reliably. In the context of our model, such inferences would have to be made via inferred similarities to other participants, such as through the relationship of the response times with covariate information. Conversely, if the cognitive task is such that most participants demonstrate perfect accuracy, then little can be inferred about the guessing process. In such a situation, it may be advisable simply to analyze the response times for the correct responses, as this analysis would coincide with our framework because essentially all the latent guessing process responses are assumed to be censored by the error-free responses. The strength of our framework is best understood when participants exhibit a small to moderate proportion of errors.

Our modeling framework makes two independence assumptions. First, we assume that the error-free and guessing processes are independent within a trial. Similar assumptions have been made in the class of "speed-accuracy decomposition" models (Meyer, Irwin, Osman and Kounios, 1988; Smith et al. 1997). While the criticisms raised about speed-accuracy decomposition may apply to our model, any dependence between the two processes is lessened by the incorporation of covariate information. The ability of our model to condition on covariate effects removes a source of dependence between the error-free and guessing processes. Our framework, in its basic form, does not account for the possibility of temporal dependence of trials within a participant. In a typical experiment, participants' performances may change over time in a predictable manner (e.g., through learning). One way to extend our model to incorporate temporal dependence is by positing a model component that assumes certain linear parameters undergo a Markov process. For example, the linear intercept parameter may be modeled both as a function of participant and by trial, and that a Markov structure may be placed on intercept. In the context of our Weibull model, we may assume a normal process

$$\beta_{0im} \sim \mathrm{N}(\beta_{0(i-1)m},\ \sigma^2),$$

where β_{0im}, the intercept for subject m on trial i results from a normal "shock" at trial $(i-1)$. Similar time series structures for generalized linear models can be found, for example, in Lindsey and Lambert (1995).

It is possible to consider semi-parametric models for the latent response time processes. One such candidate alternative to a parametric model would include Cox's proportional hazards models with participant-specific effects in the parametric component. While such models can allow greater flexibility, they can be difficult to implement without simplifying assumptions, and inferences are less precise than the parametric counterpart.

Extensions of our basic model that allow greater pooling of information across participants can be considered. Rather than include participant-specific intercepts and error-free effects as linear terms in the model, they may be modeled hierarchically. This approach is also well-suited to modeling covariate effects that are suspected of varying by participant. This and other forms of modeling can be adapted to our basic framework while still retaining an interpretation of a participant-specific error-free effect.

Appendix: EM algorithm for the Weibull model

Maximizing the likelihood in (7) with respect to λ_{Tim}, λ_{Rim}, δ and α is difficult to obtain directly. An attractive approach is to maximize the likelihood using the EM algorithm (Dempster, Laird and Rubin, 1977). To do so, we introduce the latent variable

$$I_{im} = \begin{cases} 1 & \text{if } T_{im} < R_{im} \\ 0 & \text{if } T_{im} \geq R_{im} \end{cases}$$

which is an indicator of whether the observed response time was produced by the error-free process. The complete-data likelihood, treating the I_{im} as observed, is given by

$$\begin{aligned} L_{com} = & \prod_{m=1}^{M} \left\{ \prod_{\{i:d_{im}=0\}} [f_R(y_{im}) S_T(y_{im})(1-\eta)]^{1-I_{im}} \times \right. \\ & \left. \prod_{\{i:d_{im}=1\}} [f_T(y_{im}) S_R(y_{im})]^{I_{im}} [f_R(y_{im}) S_T(y_{im})\eta]^{1-I_{im}} \right\} \end{aligned}$$

where y_{im} and d_{im} are, respectively, the observed response time and response correctness for trial i, participant m, so that the complete data log-likelihood, ℓ_{com}, is

$$\begin{aligned} \ell_{com} = & \sum_{m=1}^{M} \left\{ \sum_{\{i:d_{im}=0\}} (1-I_{im}) \ln(f_R(y_{im}) S_T(y_{im})(1-\eta)) + \right. \\ & \left. \sum_{\{i:d_{im}=1\}} (I_{im} \ln(f_T(y_{im}) S_R(y_{im})) + (1-I_{im}) \ln(f_R(y_{im}) S_T(y_{im})\eta)) \right\}. \end{aligned}$$

The EM algorithm proceeds iteratively by finding the expectation of the complete-data log-likelihood conditional on observed data and the previous iteration's parameter values, and then maximizing the resulting expression to obtain the next set of parameter values. This process is repeated until convergence.

Suppose $\lambda_{Tim}^{(k)}$, $\lambda_{Rim}^{(k)}$, $\delta^{(k)}$ and $\alpha^{(k)}$ are the values of λ_{Tim}, λ_{Rim}, δ and α at iteration k of the EM-algorithm. Because the complete-data log-likelihood is linear in the I_{im}, it suffices in developing the EM-algorithm to determine the conditional expectations of the I_{im}. For $d_{im} = 0$,

$$\mathrm{E}(I_{im} \mid y_{im}, d_{im} = 0, \lambda_{Tim}^{(k)}, \lambda_{Rim}^{(k)}, \delta^{(k)}, \alpha^{(k)}) = 0,$$

and for $d_{im} = 1$,

$$\mathrm{E}(I_{im} \mid y_{im}, d_{im} = 1, \lambda_{Tim}^{(k)}, \lambda_{Rim}^{(k)}, \delta^{(k)}, \alpha^{(k)}) = \frac{\lambda_{Tim}^{(k)}}{\lambda_{Tim}^{(k)} + \eta\lambda_{Rim}^{(k)}}.$$

Letting $\Lambda_{Tim}^{(k)} = \frac{\lambda_{Tim}^{(k)}}{\lambda_{Tim}^{(k)}+\eta\lambda_{Rim}^{(k)}}$, and $\Lambda_{Rim}^{(k)} = 1-\Lambda_{Tim}^{(k)}$, the conditional expectation of the log-likelihood is given by

$$\begin{aligned}
&Q(\lambda_{Tim}, \lambda_{Rim}, \delta, \alpha \mid \lambda_{Tim}^{(k)}, \lambda_{Rim}^{(k)}, \delta^{(k)}, \alpha^{(k)}) = \mathrm{E}(\ell_{com} \mid \lambda_{Tim}^{(k)}, \lambda_{Rim}^{(k)}, \delta^{(k)}, \alpha^{(k)}) \\
&= \sum_{m=1}^{M} \Bigg\{ \sum_{\{i:d_{im}=0\}} (\ln\alpha + \ln\lambda_{Rim} + (\alpha-1)\ln(y_{im}-\delta) - (\lambda_{Rim}+\lambda_{Tim})(y_{im}-\delta)^{\alpha} + \ln(1-\eta)) + \\
&\qquad \sum_{\{i:d_{im}=1\}} \Lambda_{Tim}^{(k)} (\ln\alpha + \ln\lambda_{Tim} + (\alpha-1)\ln(y_{im}-\delta) - (\lambda_{Rim}+\lambda_{Tim})(y_{im}-\delta)^{\alpha}) + \\
&\qquad \sum_{\{i:d_{im}=1\}} \Lambda_{Rim}^{(k)} (\ln\alpha + \ln\lambda_{Rim} + (\alpha-1)\ln(y_{im}-\delta) - (\lambda_{Rim}+\lambda_{Tim})(y_{im}-\delta)^{\alpha} + \ln\eta) \Bigg\} \\
&= \sum_{m=1}^{M} \Bigg\{ n_m \ln\alpha + \sum_{i:d_{im}=0} \ln\lambda_{Rim} + \sum_{i:d_{im}=1} \left[\Lambda_{Tim}^{(k)} \ln\lambda_{Tim} + \Lambda_{Rim}^{(k)} \ln\lambda_{Rim}\right] \\
&\qquad + (\alpha-1)\sum_{i=1}^{n_m} \ln(y_{im}-\delta) - \left(\sum_{i=1}^{n_m} (\lambda_{Tim}+\lambda_{Rim})(y_{im}-\delta)^{\alpha}\right)\Bigg\} + c \qquad (11)
\end{aligned}$$

where n_m is the number of trials for participant m, and c is a constant that only depends on known quantities.

It is useful to have the function Q in (11) made more explicitly dependent on the linear parameters in (8). To do so, we let β^* be the entire vector of all linear parameters (β_{0m}, β, θ_{0m} and θ_m, for all $m = 1, \dots, M$), and define x_{im}^* and x_{im}^{**} to be a reorganization of the covariates such that

$$\begin{aligned}
\ln\lambda_{Tim} &= \beta_{0m} + x_{im}'\beta \;=\; {x_{im}^*}'\beta^* \\
\ln\lambda_{Rim} &= \beta_{0m} + x_{im}'\beta + \gamma_{im} \;=\; {x_{im}^{**}}'\beta^*
\end{aligned}$$

Furthermore, letting w_{im} be the component of x_{im}^{**} that multiplies the terms in γ_{im}, and v_{im} be the remaining portion of x_{im}^{**}, define the concatenation

$$\hat{x}_{im}^{(k)} = (v_{im},\ (1 - d_{im}\Lambda_{Tim}^{(k)})w_{im}).$$

Then the function Q in (11) can be rewritten as

$$Q = \sum_{m=1}^{M} \left\{ n_m \ln \alpha + \sum_{i=1}^{n_m} \hat{x}_{im}^{(k)}{}'\beta^* + (\alpha - 1)\sum_{i=1}^{n_m} \ln(y_{im} - \delta) - \sum_{i=1}^{n_m} (y_{im} - \delta)^{\alpha}(\exp(x_{im}^{*}{}'\beta^*) + \exp(x_{im}^{**}{}'\beta^*)) \right\} \tag{12}$$

The function Q in (12) can be maximized using the Gauss-Seidel algorithm, fixing pairs of α, β^* and δ separately and maximizing over the remaining parameter. Conditionally maximizing each parameter in the M-step of the EM algorithm, often termed the ECM algorithm, also converges to the maximum likelihood estimate under general regularity conditions (Meng and Rubin, 1993). We outline the computation below.

Assume that k iterations of the ECM algorithm have been performed. To determine the value of $\beta^{*(k+1)}$, we perform a single Newton-Raphson step fixing the values of $\alpha = \alpha^{(k)}$ and $\delta = \delta^{(k)}$. The Newton-Raphson step is given by

$$\beta^{*(k+1)} = \beta^{*(k)} + (X^{*\prime}W^{*(k)}X^* + X^{**\prime}W^{**(k)}X^{**})(\hat{X}^{(k)}{}'\mathbf{1} - X^{*\prime}z^{*(k)} - X^{**\prime}z^{**(k)})$$

where $W^{*(k)}$ and $W^{**(k)}$ are diagonal matrices with elements

$$(y_{im} - \delta^{(k)})^{\alpha^{(k)}} \exp(x_{im}^{*}{}'\beta^{*(k)})$$

and

$$(y_{im} - \delta^{(k)})^{\alpha^{(k)}} \exp(x_{im}^{**}{}'\beta^{*(k)}),$$

respectively, and the elements of the vectors $z^{*(k)}$ and $z^{**(k)}$ contain the diagonal elements of $W^{*(k)}$ and $W^{**(k)}$. The matrices X^* and X^{**} contain the $\sum_{m=1}^{M} n_m$ rows of the x_{im}^* and x_{im}^{**}, and the matrix $\hat{X}^{(k)}$ contains the $\sum_{m=1}^{M} n_m$ rows of the $\hat{x}_{im}^{(k)}$.

To determine the value of $\alpha^{(k+1)}$, we also perform a single Newton-Raphson step fixing the values of $\beta = \beta^{(k+1)}$ and $\delta = \delta^{(k)}$. This is given by

$$\alpha^{(k+1)} = \alpha^{(k)} + \frac{\sum_{m=1}^{M} \left\{ n_m/\alpha^{(k)} + \sum_{i=1}^{n_m} \ln(y_{im} - \delta^{(k)}) - \sum_{i=1}^{n_m} (\lambda_{Tim}^{(k+1)} + \lambda_{Rim}^{(k+1)})(y_{im} - \delta^{(k)})^{\alpha^{(k)}} \ln(y_{im} - \delta^{(k)}) \right\}}{\sum_{m=1}^{M} \left\{ n_m/{\alpha^{(k)}}^2 + \sum_{i=1}^{n_m} (\lambda_{Tim}^{(k+1)} + \lambda_{Rim}^{(k+1)})(y_{im} - \delta^{(k)})^{\alpha^{(k)}} (\ln(y_{im} - \delta^{(k)}))^2 \right\}}$$

where

$$\begin{aligned} \lambda_{Tim}^{(k+1)} &= \exp({x_{im}^{*}}' \beta^{*(k+1)}) \\ \lambda_{Rim}^{(k+1)} &= \exp({x_{im}^{**}}' \beta^{*(k+1)}). \end{aligned}$$

Finally, to determine $\delta^{(k+1)}$, we perform a bisection algorithm to maximize the function Q with fixed $\beta^* = \beta^{*(k+1)}$ and $\alpha = \alpha^{(k+1)}$. Bisection is a natural and appropriate approach because the range of δ is a closed interval, bounded by $0 \leq \delta \leq \min_{i,m} y_{im}$. It is worth noting that if the model were specified with possibly unequal shift parameters for T_{im} and R_{im}, the equation for Q in 11 would need to be adjusted accordingly. The bisection CM-step for each of the shift parameters in the ECM algorithm would be identical to the common parameter case, except that the sum would be over fewer terms. Thus, the model with multiple shift parameters poses no major difficulties for the implementation of the estimation algorithm.

These three steps are carried out until the change in parameter values is lower than pre-specified tolerance. Natural starting values for the algorithm are $\beta^{*(0)} = (0, \ldots, 0)$, $\alpha^{(0)} = 1$ and $\delta^{(0)} = \min_{i,m} y_{im}/2$. For δ, the suggested initial value is the midpoint of the interval of potential values.

References

Berger, J. O., & Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, **88**, 1412–1418.

Berger, J. O., & Sun, D. (1996). Bayesian inference for a class of poly-Weibull distributions. In D. A. Berry, K. M. Chaloner, and J. K. Geweke (Eds.), *Bayesian analysis in statistics and econometrics* (pp. 101–113). New York: Wiley.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Dickman, S. J. & Meyer, D. E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. *Journal of Personality and Social Psychology*, **54**, 274–290.

Gray, J. R. (1999). A bias toward short-term thinking in threat-related negative emotional states. *Personality and Social Psychology Bulletin*, **25**, 65–75.

Gray, J. R. (2001). Emotional modulation of cognitive control: Approach-withdrawal states double-dissociate spatial from verbal two-back task performance. *Journal of Experimental Psychology – General*, **130**, 436–452.

Heathcote, A., Popiel, S.J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, **109**, 340–347.

Lavric, A., Rippon, G., & Gray, J. R. (2003). Threat-evoked anxiety disrupts spatial working memory performance: An attentional account. *Cognitive Therapy & Research*, **27**, 489–504.

Lindsey, J. K. & Lambert, P. (1995). Dynamic generalized linear models and repeated measurements. *Journal of Statistical Planning and Inference*, **47**, 129–139.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226–233.

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

Meng, X. L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, **80**, 267–278.

Meyer, D. E., Irwin, D. E., Osman, A. M. & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Review*, **95**, 183–237.

Nádas, A. (1970). On estimating the distribution of a random vector when only the smallest coordinate is observable. *Technometrics*, **12**, 923–924.

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, **6**, 155–156.

Ratcliff, R. (1979). Group RT distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446–461.

Ratcliff, R. and Rouder, J. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347–356.

Ratcliff, R. and Rouder, J. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 127–140.

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

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

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

Smith, R. W., Kounios, J., & Osterhout, L. (1997). The robustness and applicability of speed-accuracy decomposition, a technique for measuring partial information. *Psychological Methods*, **2**, 95–120.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, **30**, 276–315.

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

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, **41**, 67–85.

Yellott, J. I. (1971). Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, **8**, 159–199.