Reaction Time

Intermediate article

James T Townsend, Indiana University, Bloomington, Indiana, USA

CONTENTS

Reaction time is, along with accuracy, the most important dependent variable employed in experimental cognitive psychology and perhaps in all of experimental psychology.

INTRODUCTION

Reaction time (RT) – also called response time – refers to the time taken by a person (called a 'subject' or 'participant') to perform some task in an experiment. The experimental psychologist must manipulate aspects of the experiment in order to discover important characteristics of behavior and psychological processes, and to test theories or models of performance. Variables in the experiment that are manipulated by the experimenter are called 'independent variables'. Examples include the various aspects of the stimulus (something presented to the participant), the details of the task requirements, and even environmental variables such as the illumination in the experimental chamber.

Reaction time is, in contrast, a dependent variable, because it (like the accuracy of the response) is recorded by the experimenter and used to draw conclusions of psychological interest. Reaction time is used, for instance, to study the characteristics of the psychological system within the participant that is performing various kinds of perceptual, mental and motor tasks. Professional RT analyses depend on an understanding of statistics, experimental design, and often mathematical modeling. This article will attempt only to give a feel for the topic. This entails some sacrifice of detail and precision.

STATISTICAL PROPERTIES OF RT

One of the very first problems that a psychologist must face is that human behavior is virtually always probabilistic, or statistical; that is, it varies according to the laws of chance. This chance, probabilistic, or equivalently, statistical, aspect is present in everything from the behavior of a single neuron all the way up to complex mental operations and actions. Hence, RTs are never the same from trial to trial. The scientist collects a series of them over the experimental session and plots the results in the form of a so-called *frequency function*. The underlying variable is called the 'variate'. In the present case, the variate is RT. The RTs are segregated into small bins, say of about 0.01 seconds each, and then the number of RTs within each time bin is plotted (usually standardized by dividing by the total number of RTs accumulated in the experiment). The result is the RT frequency function.

The *cumulative frequency function*, where the frequencies are summed from the smallest value of RT up to an arbitrary level, is also very useful. Thus, the value of the cumulative frequency function at, say, $RT = 1\,s$, measures how many RTs were at or below 1 second.

These functions show the likelihood of a particular RT for a given experimental condition. They provide the basis for all other analyses and conclusions regarding the experimental results (Wenger and Townsend, 2000).

The *mean*, giving the central tendency of a frequency function, and the *variance*, indicating the variability in the data, are the most important and often-employed statistics. They are used, along with standard statistical assumptions, to investigate hypotheses about mental processes, often employing statistical techniques such as the *t*-test, analysis of variance, and so on. For instance, an experimental group given a special kind of treatment in a memory experiment (e.g., a drug, or extra learning trials) may perform a memory task faster than the control group, as revealed by its mean RT being lower in a statistically significant fashion. Such statistical inference, involving only means and variances, provides the basis for much of modern cognitive psychology and other branches of psychology.

However, there are other statistics or statistical functions of the variate that are considerably more

powerful (Townsend, 1990; Townsend and Ashby, 1983). In fact, there exists a hierarchy of statistical functions organized according to their logical strength. The 'hazard function' reveals the chance that the subject will react in the next instant, given that he or she has not ret reacted. The 'likelihood function' is a ratio of two separate frequency functions evaluated at any particular value of RT. We will next illustrate this and the other statistics within a realistic example.

Suppose that a psychologist is interested in testing the effectiveness of a drug intended to facilitate memory search. Naturally, she wants to make sure the treatment is more effective than no treatment or a placebo. She administers the real drug and a placebo to two different groups, measures their RTs in a memory-search task, and plots the frequency function. Then, the various statistical functions noted – mean, cumulative frequency function, hazard function, and likelihood function – differ in their power to discriminate between the two groups. The mean is, of course, simply the arithmetic average of the sampled RTs. It can be written as $M = \Sigma_i P_i \times RT_i$ where P_i is the proportion of RTs that fall into the i^{th} time bin and RT_i is the reaction time at the i^{th} bin. The greek letter sigma (Σ) simply tells us to sum up the values of the multiplication just to the right ($P_i \times RT_i$). And, the subscript 'i' is the index over which we form the sum. Next, the cumulative frequency function, $F(RT_i)$, is the sum of the frequencies from the smallest time bin up to and including the j^{th}, $F(RT_i) = \Sigma_i^j P_i$, where i runs from 1 up to j, where j is less than or equal to the last index value (standing for the largest value of RT found in the current experiment). The hazard function at time RT_i is just $h(RT_i) = P_i / (1 - F(RT_i))$.

All of these statistics can be compared for the real drug vs. the placebo group. In fact we could write M(DRUG) and M(PLACEBO), $F(RT_i;DRUG)$ and $F(RT_i;PLACEBO)$, and $h(RT_i;DRUG)$ and $h(RT_i;PLACEBO)$, and then see which is bigger, the real drug statistics or the placebo statistics, or if they are basically the same. If M for the real drug group is smaller than that for the placebo group then that suggests that the drug group is performing faster than the placebo group. However, this inference of the real drug group being faster than the placebo group will be supported in an even stronger manner if F is always bigger for the real drug group and even stronger than either M or F, if h is larger for the real drug group than for the placebo group. Finally, the likelihood function is composed of a ratio of P_i for the real drug group over the P_i for the placebo group, which might be expressed as $P_i(DRUG)/P_i(PLACEBO)$. If that ratio, increases as RT grows (the same thing as the index i getting bigger) then that provides the strongest evidence of all for the true drug group being faster than the placebo group, indicating effectiveness of the pharmaceutical. Again, it is important to understand that the likelihood increasing implies the hazard function ordering, that is if $P_i(DRUG)/P_i(PLACEBO)$ increases as i increases, then it follows that $h(RT_i;DRUG) > h(RT_i;PLACEBO)$, where '>' is the 'greater than' relation, for every value of i. Similarly, if $h(RT_i;DRUG) > h(RT_i;PLACEBO)$, then the result that $F(RT_i;DRUG) > F(RT_i;PLACEBO)$ for every value of i is forced to occur. Any of these findings implies that $M(DRUG) < M(PLACEBO)$. Finally, as mentioned earlier, none of these implications works in reverse. For instance, the ordering in h does not force the likelihood function to be increasing and so on. That is what we mean by the indicators differing in the power of what they say about the data. Such results are not only important for ordinary inference about experimental conditions: internal cognitive mechanisms make predictions that permit their testing against other hypothetical mechanisms only if their variables are ordered in a relatively strong way (Townsend, 1990a).

MODEL TESTING

Model testing can be roughly divided into two categories, which we shall discuss. We must begin by defining the terms 'parameter' and 'free parameter'. A parameter is a variable in a quantitative model that must be given an exact value in order for the model to make numerical predictions. A simple example is the prediction that the dependent variable y is a linear function of the independent variable x: $y = ax + b$. Here, the slope a and the intercept b are the parameters. Before they are given exact values, they are considered free parameters. After being assigned values, they are no longer free.

Fitting the Model to the Data

The most common approach to model testing involves estimating free parameters so that the predictions of the model are as close as possible to the data. Then, the fit of the model – that is, how well the model predicts the data – is assessed.

There are almost no cases in psychology where parameters can be assigned numerical values before running the experiment. With some methods of fit, the assessment of how well a

model predicts the data can be done quantitatively, with the null hypothesis of the predictions not being significantly different from the data being explicitly tested. However, in a number of important cases, there is no way to statistically test the degree of fit. Another problem with fit strategies is that the more sparse the data (e.g., fewer trials or subjects), the more likely the model is to fit acceptably, even if the method of fit allows a statistical test. This makes it difficult to adequately test a model. The reasons are fundamentally the same as for the fact that a small sample size gives less power to reject the null hypothesis in ordinary statistical inference. One helpful strategy is to compare the fits of two or more competing models. This is especially useful where no statistical assessment of fit is available.

Within the model-fitting approach, one must distinguish between the simple positing of a particular probability or frequency function for the processing time of one or more hypothesized subtasks, and a processing model that is based on higher-order psychological hypotheses. Frequency functions that have been proposed for processing times, and sometimes for the complete RT probability law, include the 'ex-Gaussian' (the convolution of an exponential with a Gaussian frequency function), the exponential itself, and the gamma or general-gamma (identical with a convolution of exponentials with the same or different parameters, respectively) (Luce, 1986).

When constructing a process model for a psychological phenomenon, one may employ particular frequency functions in the model, but usually one also has to hypothesize an architecture of separate individual sub-processes, rules of interaction, and input and output. One of the most basic questions is whether processing is serial or parallel – that is, are any two sub-processes operating at the same time, or only one at a time? For instance, it has been proposed that when people search their memories for specific information, the search involves examination of each of the individual items in memory one at a time – in other words, serially (e.g., Sternberg, 1969). However, others have suggested that instead, this type of search might take place in a parallel fashion (see, e.g., the review in Townsend (1990)). In this type of search, all the items could be examined simultaneously, that is, in parallel.

Suppose two sub-processes are arranged in a serial fashion with processing durations T_1 and T_2 and that there is also a set of other residual subprocesses that we can combine into the single variable T_0. Note that these must be random variables, since RTs are never perfectly constant from trial to trial. Then, the overall RT can be expressed as $\mathrm{RT} = T_0 + T_1 + T_2$. If, in addition, these random variables are probabilistically independent, then the frequency function of their sum is formed by a mathematical operation called the *convolution* of the individual frequency functions. Furthermore, it can be shown that the mean or expectation of RT is just $\mathrm{E(RT)} = \mathrm{E}(T_1) + \mathrm{E}(T_2) + \mathrm{E}(T_0)$.

Suppose, on the other hand, that the two subprocesses of interest are processed in parallel, but again the residual times are in series with them. Then the total time for completing both, plus the residual time, is the maximum of the two separate times plus T_0; that is, $\mathrm{RT} = \max(T_1, T_2) + T_0$, and $\mathrm{E(RT)} = \mathrm{E}(\max(T_1, T_2)) + \mathrm{E}(T_0)$. Thus, we have already formed two possible models for a mental task that involves two separate sub-processes. Of course, there are other issues or questions about system architecture and processing that can arise in model construction (e.g., Townsend and Ashby, 1983), and the theorist may need to form models of much greater complexity (e.g., Schweickert, 1978; Schweickert and Townsend, 1989).

Qualitative Testing

The second, somewhat less common, approach to model testing is to explore and determine qualitative characteristics that are predicted by a model and then to probe whether and to what extent these are exhibited by the data. By 'qualitative' is meant, for example, inequalities, the general form that functions or graphs should take, and other relationships among various aspects of data. Suppose we are interested in testing a model of choice-responding that says that the mean RT increases with the number of choice alternatives n, in proportion to the logarithm of n plus a constant. That is, $\mathrm{E(RT)} = k\log(n) + C$. Now, one might actually fit this simple model to the RT data and assess the fit, or one could start by checking a qualitative prediction of the model: whether the data curve is concave down, like the logarithm function (and many other functions). If the data curve is not concave down, then the logarithm model (as well as many other models) is rejected (disconfirmed). If the curve is concave down, then it may be reasonable to attempt more precise data fits.

A more powerful kind of qualitative distinction can be found in the discussion above of various types of statistical functions. For instance, a model might predict that the cumulative frequency functions are ordered in a certain way. In fact, it has been found that in memory search experiments, the cumulative frequency functions are indeed ordered in

terms of the number of items that must be searched in memory (e.g., Townsend, 1990b, Figure 3). This ordering implies that in a relatively strong statistical sense (stronger than would follow from an ordering based on mean RT alone, for instance), more items do indeed take longer to search through. Interestingly, it turns out that both serial models and many parallel models can make this prediction.

There is one qualitative approach that was developed in the late 1960s and has evolved into a rather impressive arsenal of related techniques intended to reveal the underlying architecture. The approach began as the 'additive factors method' (Sternberg, 1969; Sanders, 1983). The basic idea is to assume that processing is serial and that each sub-process in a series of sub-processes engaged by a certain task may be associated with an experimental factor that can lengthen or shorten the duration taken by that sub-process to finish its part of the task in the series. For instance, an early sensory sub-process might be affected by stimulus intensity, while a late motor sub-process might be affected by response difficulty. (*See* **Information Processing**)

In an additive factors study, the experimenter manipulates two or more factors intended to target certain sub-processes of interest. Because of the assumed seriality the manipulation of the experimental factors should influence RT in an additive fashion. We can then write $\mathrm{RT}(x_1, x_2) = T_1(x_1) + T_2(x_2) + T_0$, and we have the prediction that $\mathrm{E}(\mathrm{RT}) = \mathrm{E}(T_1(x_1)) + \mathrm{E}(T_2(x_2)) + \mathrm{E}(T_0)$. This prediction can readily be tested in the RT data by standard statistical procedures. If additivity is found then the conclusion is that there do indeed exist two psychological sub-processes, that they operate in series, and that they are suitably affected by the experimental factors. On the other hand, if additivity is not found, then one or more of the experimental factors must affect more than 'its own' sub-process (e.g., stimulus intensity might affect both an early and a late sub-process), or that there are not two separate serially-arranged sub-processes.

The additive factors method was extended over the years by a number of authors to include parallel processes as well as much more complex mental networks (e.g., Schweickert and Townsend, 1989). The point to be made here is that the predicted additivity by serial models, and various kinds of non-additivity by other types of architectures, are all of a qualitative nature. It is not necessary to know exactly what the RTs are, only that, for instance, they obey the additivity property. Of course, it is possible, once one has affirmed, say, the additivity, to attempt to fit particular serial models to the numerical data.

It appears that RT will continue to be one of the most important dependent variables used in studying cognitive processes.

References

Luce RD (1986) *Response Times: Their Role in Inferring Elementary Mental Organization*. New York, NY: Oxford University Press.

Sanders AF (1983) Towards a model of stress and human performance. *Acta Psychologica* **53**: 61–97.

Schweickert R (1978) A critical path generalization of the additive factor method: analysis of a stroop task. *Journal of Mathematical Psychology* **18**: 105–139.

Schweickert R and Townsend JT (1989) A trichotomy method: interactions of factors prolonging sequential and concurrent mental processes in stochastic PERT networks. *Journal of Mathematical Psychology* **33**: 328–347.

Sternberg S (1969) Memory scanning: mental processes revealed by reaction time experiments. *American Scientist* **57**: 421–457.

Townsend JT (1990a) The truth and consequences of ordinal differences in statistical distributions: toward a theory of hierarchical inference. *Psychological Bulletin* **108**: 551–567.

Townsend JT (1990b) Serial vs. parallel processing: sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science* **1**: 46–54.

Townsend JT and Ashby FG (1983) *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge, MA: Cambridge University Press.

Wenger MJ and Townsend JT (2000) $H(t)$ and $C(t)$: basic tools for attention and general processing capacity in perception and cognition. *Journal of General Psychology: Visual Attention* **80**: 67–99.

Further Reading

Ashby FG, Tein JY and Balakrishnian JD (1993) Response time distributions in memory scanning. *Journal of Mathematical Psychology* **37**: 526–555.

Myung IJ, Forster MR and Browne MW (eds) *Journal of Mathematical Psychology* **44**(1). [A technical survey of various important approaches to model testing.]

Smith PL (1995) Psychophysically principled models of visual simple reaction time. *Psychological Reviews* **102**: 567–593.

Townsend JT (1990) Serial vs. parallel processing: sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science* **1**: 46–54.

Van Zandt T (2000) How to fit a response time distribution. *Psychonomic Bulletin and Review* **7**: 424–465.

Van Zandt T and Ratcliff R (1995) Statistical mimicking of reaction time distributions: mixtures and parameter variability. *Psychonomic Bulletin and Review* **2**: 20–54.