Journal of Mathematical Psychology 55 (2011) 290–301

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

# Performance measures for dynamic signal detection

J.D. Balakrishnan [a,*], Justin A. MacDonald [b]

[a] *Statistics Department, California Polytechnic State University, Building 25, Room 107D, San Luis Obispo, CA 93407-0405, United States*

[b] *Department of Engineering Psychology, New Mexico State University, United States*

ARTICLE INFO

*Article history:*
Received 18 June 2010
Received in revised form
27 April 2011
Available online 15 June 2011

*Keywords:*
Statistical decision theory
Signal detection theory
Sequential decision theory
Discrimination
Decision criterion
Sensitivity
Response bias
Speed–accuracy trade-off
Pure insertion
Motor time

ABSTRACT

For more than half a century, experimental studies of various kinds of detection and discrimination behavior have tended to rely on the simple, two-stage statistical decision model known as signal detection theory. An apparent weakness of this classical framework is its assumption that making a decision is equivalent to choosing a decision criterion or boundary to map perceptual or evidence states to a binary classification response. This static representation leads to several fundamental mispredictions about qualitative properties of discrimination, each of which is characteristic of a dynamic detection process. In this article, we show that there is a robust solution to a second class of problems introduced originally by detection theorists, but later mostly abandoned — the problem of estimating the detectability of the signal when the decision process is sequential. In an empirical application, a detectability statistic defined on a crude description of the temporal dynamics of the detection process is shown to be roughly constant under manipulations of both response preference and response speed. The estimated stringency of the stopping condition decreased in conjunction with a decrease in signal strength in time, consistent with the hypothesis that sensory information is retrieved from a decaying memory store. The analysis also makes it possible to estimate the bivariate distribution of the sensory and non-sensory components of the response time.



In their review of the accelerated development of radar detection systems during the Second World War, Lawson and Uhlenbeck (1950) defined a model for the transformation of the effects of electromagnetic radiation on an antenna to an audio or video image that could be interpreted by a human operator. The model assumed that the information that is presented on the display (the output of the receiver) is statistical in nature. That is, due to random fluctuations in, for example, voltage or current, the content of the display is a waveform that contains a signal plus noise. The effect of this stimulus on the perceptual system of the operator was assumed to be the component frequencies in the stimulus that fall within the perceptible range, plus some additional noise caused by random organic fluctuation. The chapter in the text 'initiated by' Siegert considered the decision rules that could be adopted by the operator when there are two possible events to be discriminated, *Signal* or *Noise*.

This formal description of a human–machine decision system was later developed into a model for human perception by Tanner and Swets (1954), and came to be known as the theory of signal detection (or signal detection theory). In this psychophysical model, the activity on the display was the signal and the decision problem was statistical due to noise in the perceptual system. The representation of the effect of a stimulus as a signal in noise was not new to psychologists—what was novel in the theory was the assumption that human observers make sophisticated statistical decisions, estimating the posterior likelihoods of the stimuli to be discriminated and then choosing the response that would minimize a subjective expected loss. This fundamental statement about the objective of the decision making process makes it possible to develop a decision model for virtually any type of discrimination experiment. As the theory evolved, however, it became increasingly focused on the analysis of two-choice detection and discrimination tasks (e.g., *yes–no* detection and two-alternative forced-choice) in which the stimulus is presented for a fixed period and the observer's goal is to maximize the probability of a correct judgment or an expected payoff defined on the four possible outcomes of a trial. The signal processing part (i.e., how to make the observations) of the original detection model was then mostly forgotten, and the theory became what it is mostly known as today, i.e., a purely statistical model, in which two distributions describe the effect of the stimulus presentation on *Signal* and *Noise* trials and a decision rule assigns each possible effect to one of the two possible detection responses (e.g., Macmillan & Creelman, 2005).

Another early form of the signal detection problem, the sequential detection model, assumed that collecting data is costly and

* Corresponding author.
*E-mail address:* jbalakri@calpoly.edu (J.D. Balakrishnan).

0022-2496/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2011.05.001

the operator must decide, after each observation, whether it is best (per the expected loss) to stop immediately and choose a response or to continue collecting more information (e.g., Birdsall & Roberts, 1965; Degroot, 1970; Peterson, Birdsall, & Fox, 1954). Once the observation process is terminated, a decision rule is applied to the outcomes of the observations taken in order to select a response. These types of decision problems do not have simple solutions, and the models whose predictions have been determined do not appear to be good descriptions of speeded classification behavior (see, e.g., Luce, 1986; Townsend & Ashby, 1983, for reviews). As a consequence, this area of statistical decision theory is not often cited in psychology, despite the increasing emphasis on decision time in many current paradigms in perception and memory research.

Recently, we have contended that even in the classical *yes*–*no* detection or discrimination task, human observers will adopt a sequential decision making strategy (Balakrishnan, 1999; Balakrishnan & MacDonald, 2003, 2008; Balakrishnan, MacDonald, & Kohen, 2003). Indirect support for this claim has been available for a long time. For example, in some of the classical psychophysical methods, an inverse relationship was observed between response time and response confidence and between response time and the size of the stimulus difference (e.g., Henmon, 1911, Johnson, 1939; see Link, 1992; Pleskac & Busemeyer, 2010 for reviews). In some of the early *yes*–*no* detection experiments, the preferred (more frequent) response was shown to be both faster and more likely to be correct than the unpreferred (less frequent) detection response (e.g., Carterette, Friedman, & Cosmides, 1965; Emmerich, Gray, Watson, & Tanis, 1972, see Pike, 1973 for a review). Statistical decision theory predicts, fundamentally, that the decision maker's accuracy will increase with the number of observations taken. Although it is not necessary to equate the sample size with the decision time, the interdependence of response accuracy, latency, and frequency in detection and discrimination experiments certainly appearsto favor the sequential detection model over the fixed-sample model.

More direct support for the sequential model comes from the relationship between preference and confidence, about which both models are explicit in their predictions. Let $O$ be the outcome of the observations taken (the sensory effect) and let $S$ be the stimulus type, *Signal* or *Noise*. The decision maker's confidence that a *Signal* was presented is directly related to the posterior likelihood ratio,

$$l = \frac{P(O|S = \text{Signal})}{P(O|S = \text{Noise})}.$$

In the fixed-sample model, the decision maker will respond *yes* (*no*) if $l$ is greater than (less than) a criterion, $\beta$, which is the measure (or definition) of the decision maker's bias. The optimal value of the criterion depends on the ratio of the prior probabilities of the two stimuli, while the distributions of $l$ on *Signal* and *Noise* trials are independent of $\beta$.

Analysis of the likelihood ratios corresponding to observers' different degrees of confidence in the accuracy of their detection responses suggest exactly the opposite relationship: the distributions of $l$ change under different prior probability conditions, but the criterion value $\beta$ is always equal or close to one (Balakrishnan, 1998, 1999; Balakrishnan & MacDonald, 2003, 2008; Van Zandt, 2000). The effect of the prior probabilities on the distribution of $l$ is predicted by the sequential detection model, because the distribution of $l$ will depend on the observer's stopping rule as well as the stimuli to be discriminated, and the stopping rule will depend on the prior probabilities. The bias in the stopping rule is such that always adopting the unbiased terminal decision rule, even when the prior probabilities are unequal, will have almost no consequence for overall accuracy.

With these specific reasons for rejecting the classical signal detection model in mind, we return in this article to Lawson and Uhlenbeck (1950)'s original threshold signals problem and develop a method of estimating the parameters of a dynamic detection process. From the information contained in the response times, response confidence, and the *yes*-or-*no* detection response, we show how it is possible to obtain a rough description (i.e., with a crude response time scale) of the temporal evolution the sensory effect of the stimulus and the logic of the observer's stopping condition. The analysis allows us to define a statistic that remains roughly constant under manipulations of both response preference and speed pressure, and also to estimate the bivariate distribution of sensory and non-sensory components of the observable response time.

We begin by assuming that after a transformation of the observer's confidence rating, $C$, to a likelihood ratio, $P(C = c|S = \text{Signal})/P(C = c|S = \text{Noise})$, we obtain an estimate of the terminal state of a random walk, where the state of the walk at each time, $k$ is the likelihood ratio defined by the sensory effect of the stimulus up to $k$. Most of the first part of the article (Sections 2 and 3) is devoted to this assumption and its justification. From this terminal state and the response time, the problem is to estimate the parameters of the walk, and from this temporal description, a measure of signal detectability.

## 1. Basic principles of detection theory

In the static, or fixed-sample, form of the threshold signals problem, a device attached to a sensor (e.g., a receiver) records the sensor's output during a specific, well-defined interval, during which one of two types of stimuli is present, *Signal* or *Noise*. Due to some transient effects in the environment or in the sensor itself, the sensory effect, $\psi$, will rarely or never uniquely identify the stimulus. However, it will almost always alter the relative likelihoods of the two stimuli, increasing one at the expense of the other. The sensory effect is passed on to a decision maker, who assigns (or 'maps') it to one of two responses, *yes* or *no*. The observer's detection response, *yes* or *no*, is the terminal decision—i.e., the decision that completes the trial.

This overt response is understood to be a gamble, with the chance of winning determined by the sensory effect of the stimulus, $\psi$, the prior probabilities of the stimuli, $p_{\text{Signal}}$ and $p_{\text{Noise}}$, and the response that is selected, $R$. By choosing one of the two responses, *yes* or *no*, the observer chooses between one of two gambles, with the chance of winning being

$$\varphi_{\text{Signal},\psi} = P(S = \text{Signal}|\psi)$$

if the observer responds *yes*, and

$$\varphi_{\text{Noise},\psi} = P(S = \text{Noise}|\psi) = 1 - \varphi_{\text{Signal},\psi}$$

if the observer responds *no*. Notice that $\varphi_{\text{Signal},\psi}$ and $\varphi_{\text{Noise},\psi}$ are random variables whose values are probabilities.[1] For our purposes, there is no reason to consider objectives other than maximizing the probability of a correct judgment on a given trial. We will therefore refer to $\varphi_{\text{Signal},\psi}$ and $\varphi_{\text{Noise},\psi}$ as the *risks* that the observer is facing during the experiment.

In order to have the best chances of winning the bet, the observer must choose the safer (less risky) gamble, which is the larger of $\varphi_{\text{Signal},\psi}$ and $\varphi_{\text{Noise},\psi}$. The decision rule is *suboptimal* (for accuracy) if, with non-zero probability, the observer will choose the response that does not correspond to the maximum of $\varphi_{\text{Signal},\psi}$ and $\varphi_{\text{Noise},\psi}$, that is, if

$$P(R = \text{yes}|\varphi_{\text{Signal},\psi} < \varphi_{\text{Noise},\psi}) > 0,$$

[1] Here and elsewhere, the variables being defined are random variables if any random variable in the definition is free (i.e., is not bound by a quantifier or assigned a specific value).

or

$$P(R = \text{no}|\varphi_{\text{Noise},\psi} < \varphi_{\text{Signal},\psi}) > 0.$$

In such a case, the probability of a correct response will be less than it would have been if the optimal decision rule was adopted.

When the prior probabilities of the stimuli are equal, choosing the maximum of $\varphi_{\text{Signal},\psi}$ and $\varphi_{\text{Noise},\psi}$ is equivalent to responding *yes* whenever the conditional probability of the effect $\psi$ on *Signal* trials is greater than the conditional probability of this effect on *Noise* trials,

$$P(\psi|S = \text{Signal}) > P(\psi|S = \text{Noise}),$$

and responding *no* whenever this inequality is reversed,

$$P(\psi|S = \text{Noise}) > P(\psi|S = \text{Signal}).$$

This strategy is called the *unbiased decision rule* in detection theory.

The expected percent correct of the optimal decision maker increases as the difference in the prior probabilities of the stimuli increases. To quantify the amount of sensory noise, therefore, we will define the *detectability* of the signal in the fixed-sample detection problem as the probability of a correct response when the decision rule is optimal and the prior probabilities are equal.

## 2. Implications of the order of transformations in the threshold signals model

In order to determine whether the observer's decision rule is suboptimal, the classical signal detection analysis introduces some assumptions about the distributions of the sensory effect $\psi$ on *Signal* and *Noise* trials (e.g., that they are univariate normal) and about the complexity of the decision rule (e.g., that a single, fixed criterion separates the sensory effects into *yes* and *no* responses). One reason for choosing a parametric model was, presumably, the fact that the sensory effect of the stimulus is not an observable quantity. It is worth noting, however, that the same limitation applies to Lawson and Uhlenbeck (1950)'s threshold signals problem, but this is not the reason why they developed parametric models to describe the output of the sensor and receiver. In the engineering model, the effect of the stimulus on the sensor (e.g., some representation of the effect of electromagnetic radiation on an antenna) and the output of the sensor (e.g., voltages or currents) are not the same waveforms. The output of the sensor (or receiver) can be understood as a measure, $\psi_m$, that is taken on the effect of the stimulus on the sensor, $\psi$ (or the effect of the sensor on the receiver). Although the transformations are designed to communicate the most pertinent information in the behavior of the sensor, there is always the possibility – due to noise of various kinds – that $\psi$ will contain at least some information about the stimulus that is not contained in the final signal (e.g., a video or audio display) that is communicated to the decision maker.

Since there is no way to determine $\psi$ from $\psi_m$, it is impossible to determine, from observations of $\psi_m$, the extent of the difference in the amount of information afforded by the two measures. However, it is also important to remember that the representations $\psi$ and $\psi_m$ are not merely two unrelated random processes: $\psi$ is the effect of the stimulus on the sensor, and $\psi_m$ is the result of a conversion applied to the sensor output. The output $\psi_m$ may be a more suitable type of representation for the decision maker to operate on (e.g., a signal in a human observer's perceptible range), but even so, $\psi_m$ is assumed to be further 'downstream' from the stimulus, and hence it would not contain any information about the stimulus that is not also contained in $\psi$—it could only lose information. Under this assumption, the risks faced by a decision maker when both $\psi$ and $\psi_m$ are available pieces of information depends only on $\psi$,

$$P(S = \text{Signal}|\psi = y, \psi_m = x) = P(S = \text{Signal}|\psi = y),$$

and

$$P(S = \text{Noise}|\psi = y, \psi_m = x) = P(S = \text{Noise}|\psi = y),$$

for each $y$ and $x$. In other words, $\psi_m$ and $S$ are conditionally independent given $\psi$, and the optimal decision rule is defined only on $\psi$. Stated in terms of the likelihood ratio,

$$\frac{P(\psi = y, \psi_m = x|S = \text{Signal})}{P(\psi = y, \psi_m = x|S = \text{Noise})} = \frac{P(\psi = y|S = \text{Signal})}{P(\psi = y|S = \text{Noise})},$$

for each $y$ and $x$. Ultimately, this simple consequence of the flow of information in the threshold signals model – which we will refer to henceforth as the *downstream constraint*[2] – makes it possible to draw certain inferences about the information in $\psi$ in a behavioral study from the information contained in an experimentally observable measure, $\psi_m$.

## 3. Sufficient conditions for suboptimal and biased decision rules

Many behavioral measures would be, like $\psi_m$ in the engineering problem, downstream from the sensory effect of the stimulus. Any information about the stimulus that is contained in the observable measure $\psi_m$ must therefore also be contained in $\psi$. The problem is therefore to understand how the risks associated with the unobservable values of $\psi$ are related to the risks associated with the observable values of $\psi_m$. It is obviously impossible to fully resolve this black box problem by any purely statistical analysis of $\psi_m$. However, certain properties of $\psi_m$ will place certain constraints on $\psi$. First, note that by the *law of alternatives*,

$$P(S = \text{Signal}|X = x) = \sum_y P(S = \text{Signal}|X = x, Y = y)P(Y = y|X = x), \quad (1)$$

where the sum is taken over all possible values of $Y$. Because

$$\sum_y P(Y = y|X = x) = 1,$$

the amount of risk associated with the *yes* response when the decision maker knows that $X = x$ but does not know the value of $Y$ is a weighted average over the risks associated with the *yes* response when $X = x$ and $Y = y$, with the weights assigned to these risks being $P(Y = y|X = x)$. This probability law will be crucial to our analysis of, and our solution to, the sequential detection problem.

The terms being averaged over in (1) – that is, $P(S = \text{Signal}|X = x, Y = y)$ for different values of $y$ – need not have any interesting behavioral interpretation in order for the law of alternatives to apply. However, when $X$ is the observable measure taken on the sensor, $\psi_m$, and $Y$ is the true output of the sensor, $\psi$, then the downstream constraint applies, and the risk associated with the observable value $\psi_m$, $P(S = \text{Signal}|\psi_m = x)$, is entirely due to the effect $\psi$,

$$\begin{aligned} P(S = \text{Signal}|\psi_m = x) &= \sum_y P(S = \text{Signal}|\psi = y, \psi_m = x)P(\psi = y|\psi_m = x) \\ &= \sum_y P(S = \text{Signal}|\psi = y)P(\psi = y|\psi_m = x). \end{aligned}$$

This means that when the downstream constraint is satisfied, the conditional probability that the experimenter can estimate, $P(S =$

[2] There may be a better term. We did not find any explicit statement of this assumption in the early detection theory papers.

Signal$|\psi_m = x$), must be a weighted average over the values that $\varphi_{\text{Signal},\psi}$ can assume (i.e., the risks) when $\psi_m = x$, with the weight associated with a given risk being the relative frequency of this risk when $\psi_m = x$.

Because of this special relationship between the risks associated with observable and unobservable measures when the downstream constraint is satisfied, the single most important question about the decision making strategy that classical detection theory poses – is the observer's decision rule biased or suboptimal with respect to $\psi$? – can sometimes be answered unequivocally, despite the fact that $\psi$ is never itself observable. Suppose that on the trials when the dependent measure $\psi_m$ is equal to $x$, the participant will respond *yes* with some non-zero probability. Then the value $P(S = \text{Signal}|R = \text{yes}, \psi_m = x)$ is a weighted average over the values of the risks that are incurred by the observer when $R = \text{yes}$ and $\psi_m = x$. If the observer's decision rule is optimal (for accuracy), then none of these unobservable risk values is less than 1/2, and it is impossible for the weighted average of these values to be less than 1/2.

To establish that an observer makes suboptimal decisions for at least some values of $\psi$, therefore, it is sufficient to show that

$$P(S = \text{Signal}|R = \text{yes}, \psi_m = x) < \frac{1}{2}, \tag{2a}$$

or

$$P(S = \text{Noise}|R = \text{no}, \psi_m = x) < \frac{1}{2}, \tag{2b}$$

for some value $x$.

The derivation of the corresponding sufficient condition for bias in the observer's decision rule is less intuitive, but will turn out to be more important in our analysis of sequential detection processes below. First, define the random variable whose value is the likelihood ratio,

$$l_\psi = \frac{P(\psi|S = \text{Signal})}{P(\psi|S = \text{Noise})},$$

and let

$$l_\psi(y) = \frac{P(\psi = y|S = \text{Signal})}{P(\psi = y|S = \text{Noise})},$$

be the value of the likelihood ratio when $\psi = y$.

Now define the likelihood ratio for a given value of the joint dependent measure, $\{\psi_m, R\}$,

$$\begin{aligned} &B_{\psi_m,R}(x, w) \\ &= \frac{P(\psi_m = x, R = w|S = \text{Signal})}{P(\psi_m = x, R = w|S = \text{Noise})} \\ &= \frac{\sum P(\psi_m = x, R = w|\psi = y, S = \text{Signal})P(\psi = y|S = \text{Signal})}{\sum P(\psi_m = x, R = w|\psi = y, S = \text{Noise})P(\psi = y|S = \text{Noise})}, \end{aligned}$$

where $w$ is *yes* or *no*, and $P(\psi_m = x, R = w) > 0$. Applying Bayes' rule and the downstream constraint,

$$\begin{aligned} &P(\psi_m = x, R = w|\psi = y, S = X) \\ &= \frac{P(S = X|\psi = y, \psi_m = x, R = w)P(\psi_m = x, R = w|\psi = y)}{P(S = X|\psi = y)}, \\ &= P(\psi_m = x, R = w|\psi = y), \end{aligned}$$

where here and elsewhere, $X$ is the stimulus type, *Signal* or *Noise*. Therefore, $B_{\psi_m,R}(x, w)$ can be expressed as

$$\begin{aligned} &B_{\psi_m,R}(x, w) \\ &= \frac{\sum_y P(\psi_m = x, R = w|\psi = y)P(\psi = y|S = \text{Signal})}{\sum_y P(\psi_m = x, R = w|\psi = y)P(\psi = y|S = \text{Noise})}. \end{aligned} \tag{3}$$

The likelihood ratio, $B_{\psi_m,R}(x, w)$, is therefore a weighted combination of the numerators of the unobservable likelihood ratios, $l_\psi(y)$, that have non-zero probability when the event observed is $\{\psi_m = x, R = w\}$ divided by a weighted combination of the denominators of these same unobservable ratios. The weighting is such that the contribution of a given likelihood ratio, $l_\psi(y) = v$, to the average depends on the consistency with which the likelihood ratio $v$ is assigned to (co-occurs with) the event $\{\psi_m = x, R = w\}$ and on the relative frequency of $v$, $P(l_\psi = v)$.

If the decision rule is unbiased, the observer's response is always $R = \text{yes}$ when the numerator of the likelihood ratio is greater than the denominator, and $R = \text{no}$ when the numerator is less than the denominator. A sufficient condition for bias in the decision rule is therefore

$$B_{\psi_m,R}(x, \text{yes}) = \frac{P(\psi_m = x, R = \text{yes}|S = \text{Signal})}{P(\psi_m = x, R = \text{yes}|S = \text{Noise})} < 1, \tag{3a}$$

or

$$B_{\psi_m,R}(x, \text{no}) = \frac{P(\psi_m = x, R = \text{no}|S = \text{Signal})}{P(\psi_m = x, R = \text{no}|S = \text{Noise})} > 1, \tag{3b}$$

for some value $x$.

*Comments.*

The conditions for suboptimality and bias of the decision rule are sufficient under only one behavioral assumption, i.e., that $\psi_m$ satisfies the downstream constraint. It is also important to recognize, however, that these sufficient conditions are not also necessary: Even if the measure $\psi_m$ satisfies the downstream constraint and the conditions for bias or suboptimality are not satisfied, this would not establish that the decision rule is unbiased or optimal with respect to $\psi$. The biased or suboptimal components in the weighted averaging processes might simply be averaged out in the sum. Whether or not the bias or suboptimality of the observer's decision rule can be detected when it exists will depend on the covariance of $\psi_m$ and $\psi$.

The empirical problem is therefore to find dependent measures that are more predictive of the stimulus than the *yes–no* detection response alone (and satisfy the downstream constraint). Confidence ratings turn out to be a convenient example of a strongly predictive measure, enough so that both the suboptimality and bias conditions are satisfied. However, they are satisfied in the situations predicted by sequential detection theory rather than by the static signal detection models (Balakrishnan & MacDonald, 2008).

## 4. Estimating the parameters of a sequential detection process

In the sequential detection problem, a cost on the total number of observations taken is added to the cost of the decision outcome. The decision maker continues to take observations until the expected payoff if either the *yes* or the *no* response is immediately rendered is higher than the expected payoff if at least one more observation is obtained. The sequential detection model is therefore a series of distinct detection models, each with a potentially different set of sensory and decision parameters, and each with a third possible decision, 'continue observing'. At each possible stopping time $k$, there is a new sensory observation, $\psi_k$, and a stopping condition, $\theta_k$. The stopping condition is a pair of probabilities or likelihood ratios, $b_{\text{yes}}(k)$ and $b_{\text{no}}(k)$, which represent the minimum strength of evidence in favor of the *Signal* (*Noise*) stimulus at time $k$ that would cause the expected return if the *yes* (*no*) response is selected at time $k$ to be larger than the expected return if at least one more observation is taken before a decision is made.

As a model for human behavior, the optimal stopping condition at time $k$ in the sequential detection problem depends on the observer's subjective cost function as well as the sensory effects of the stimulus. The pair of probabilities that define the optimal stopping condition at $k$ may therefore depend on $k$ and could also vary from trial to trial. Irrespective of how this cost function is defined, however, on each trial there must be a stopping time, $k_T$, and a sensory effect of the stimulus, $\Psi(k_T) = \{\psi_1, \ldots, \psi_{k_T}\}$. A decision rule maps each possible sensory effect to a *yes*-or-*no* detection response, and this rule will be optimal or suboptimal and biased or unbiased with respect to the sensory effect $\Psi(k_T)$ in the exact same sense as the decision rule in the fixed-sample detection model.

Under the experimenter's instruction in the classical *yes–no* detection task (to be as accurate as possible), $k_T$ should be the maximum number of observations that are possible, $k_{\max}$, in the time that is allowed for a response, but according to our thesis, the observer's imposition of a speed incentive causes the observer to sometimes terminate the sensory observation process at some time prior to the maximum.[3] The problem is to estimate, from the behavior in a given *yes–no* detection experiment (e.g., with unequal priors) the detectability of the signal (probability correct) when the prior probabilities are equal and both the decision rule and the stopping rule are optimal (for accuracy).

This is not a simple problem—it is obviously impossible to recover the parameters of each member of a series of detection models from the relative frequencies of the *yes*-or-*no* responses alone. Assuming, however, that in addition to the detection response, the experimenter also records the observer's response time and response confidence, it becomes possible not only to estimate the detectability of the signal in a robust fashion, but also to partition the response time distribution into its sensory and non-sensory components. In fact, in order to estimate the parameters of the sequential detection process, this classical 'mental chronometry' problem (Ashby & Townsend, 1980; Donders, 1969; Dzhafarov, 1992; Kolev, Falkenstein, & Yordanova, 2006) must also be resolved.

Estimating the detectability of a signal from the overt behavior of a sequential decision maker is orders of magnitude more difficult than the analysis of a fixed-sample detection process. To make the development more manageable, our solution is presented in three phases. In the first phase, we assume that the number of observations taken on a given trial ($k_T$) is known (i.e., that $RT = k_T$) and that the decision maker's stopping condition at each possible stopping time $k$ is the same on each trial. This makes it easier to highlight the most important step in the analysis, as well as the key assumptions that are involved in the estimation procedure. In the second phase, we drop the assumption that the stopping condition at each possible stopping time $k$ is constant across trials, and in the third and final phase, we add a post-stopping time component to the observable $RT$.

### 4.1. Recovering the terminal state from the observable response

Suppose that the dependent measure $\psi_m$ is a confidence rating which is executed at the same time as the *yes*-or-*no* response (e.g., the responses are given on a bipolar rating scale with an explicit *yes–no* response cutoff in the middle). The downstream constraint should be satisfied for such a measure, and the time to respond is the time needed to select and execute the detection response. The sensory effect $\Psi(k_T)$ in the sequential model can be substituted for the sensory effect $\psi$ in the definitions and tests developed above for the fixed-sample model. That is, the sufficient conditions for suboptimality and bias of the fixed-sample decision process are sufficient conditions for suboptimality and bias of the sequential decision maker's terminal decision rule. However, because $k_T$ is sometimes less than $k_{\max}$, the probability correct when an optimal terminal decision rule is applied to $\Psi(k_T)$ (and the priors are equal) underestimates the true detectability of the signal.

The first step in developing a measure to control for the effect of a suboptimal stopping rule on performance is to consider the random variable that defines the unobservable sensory evidence state,

$$L_{k_T=k}(k) = \frac{P(\psi_1, \ldots, \psi_k | S = \text{Signal}, k_T = k)}{P(\psi_1, \ldots, \psi_k | S = \text{Noise}, k_T = k)},$$

and its relationship to the observable random variable,

$$B_{\psi_m,R} = \frac{P(\psi_m, R | S = \text{Signal})}{P(\psi_m, R | S = \text{Noise})}.$$

From Eq. (3), it follows that the average value of the terminal evidence state, $L_{k_T=k}(k)$, when $B_{\psi_m,R} = w$ and $k_T = k$ is the value $w$, where the averaging is of the type defined in the equation. It is at least possible, therefore, that the observable measure $B_{\psi_m,R}$ could be an effective estimate of $L_{k_T=k}(k)$, where 'effective' means that by assuming that the values of $B_{\psi_m,R}$ and $L_{k_T=k}(k)$ are the same on each trial, we will be able to define a set of measures that identify the basic dynamics of the detection process and distinguish decisional from sensory variables in discrimination experiments.

The information contained in the sensory dynamics of the detection process is determined by the distributions of the likelihood ratio,

$$l_{S=X}(k) = \frac{P(\psi_{S=X,k} | \psi_{S=X,1}, \ldots, \psi_{S=X,k-1}, S = \text{Signal}, k_T \geq k)}{P(\psi_{S=X,k} | \psi_{S=X,1}, \ldots, \psi_{S=X,k-1}, S = \text{Noise}, k_T \geq k)},$$

for each stimulus type $X$ and each time $k$. Assuming that $B_{\psi_m,R} = L_{k_T=k}(k)$, therefore, the problem is to recover these likelihood ratio distributions from the distributions of $L_{k_T=k}(k)$ at each time $k$. Although the distribution of $L_{k_T=k}(k)$ for a given stopping time $k$ should depend on the sensory effect at each time $k \leq k_T$, the value of $L_{k_T=k}(k)$ is only observed (by way of $B_{\psi_m,R}$) on a given trial because the sensory effect sequence, $\Psi(k)$, satisfied the observer's stopping condition at some unknown time $k_T \leq RT$. The difference between the observable $RT$ and the unobservable stopping time $k_T$,

$$D = RT - k_T,$$

is the time it takes to select and execute the response $\{\psi_m, R\}$. For now, we will ignore this non-sensory part of the total response time and assume that $P(D = 0) = 1$.

### 4.2. The prior and posterior state distributions

If the decision maker chooses to take another observation at time $k$, the strength of the sensory evidence will change from its value at $k - 1$,

$$L_{S=X,k_T \geq k}(k-1) = \frac{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k-1} | S = \text{Signal}, k_T \geq k)}{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k-1} | S = \text{Noise}, k_T \geq k)}$$

to its new value at $k$,

$$L_{S=X,k_T \geq k}(k) = \frac{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k} | S = \text{Signal}, k_T \geq k)}{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k} | S = \text{Noise}, k_T \geq k)}.$$

The difference between these two evidence states is multiplication by the likelihood ratio of the sensory effect at $k$,

$$L_{S=X,k_T \geq k}(k) = \frac{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k-1} | S = \text{Signal}, k_T \geq k)}{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k-1} | S = \text{Noise}, k_T \geq k)}$$

[3] Of course, if the conditional probability of the Signal at $k < k_{\max}$ reaches the value 1 or 0, the observer can comply with the accuracy instructions without waiting until $k_{\max}$ to respond. However, this possibility alone would not explain the speed–accuracy trade-off or the effect of response preference on response confidence.

$$\times \frac{P(\psi_{S=X,k}|\psi_{S=X,1},\ldots,\psi_{S=X,k-1}, S=\text{Signal}, k_T \geq k)}{P(\psi_{S=X,k}|\psi_{S=X,1},\ldots,\psi_{S=X,k-1}, S=\text{Noise}, k_T \geq k)}$$
$$= L_{S=X,k_T\geq k}(k-1)$$
$$\times \frac{P(\psi_{S=X,k}|\psi_{S=X,1},\ldots,\psi_{S=X,k-1}, S=\text{Signal}, k_T \geq k)}{P(\psi_{S=X,k}|\psi_{S=X,1},\ldots,\psi_{S=X,k-1}, S=\text{Noise}, k_T \geq k)}.$$

The most important step in the analysis, therefore, is to develop a method to estimate, from the distributions of the random variables $L_{S=X,k_T=k}(k)$ at each stopping time $k$, the distributions of $L_{S=x,k_T\geq k}(k-1)$ and $L_{S=X,k_T\geq k}(k)$ for each $X$ and $k$. Since $L_{S=x,k_T\geq k}(k-1)$ is the evidence state before the observation at time $k$ and $L_{S=X,k_T\geq k}(k)$ is the evidence state after this new observation, we will refer to these two distributions as the *prior* and *posterior* state distributions, respectively, for the observation time $k$. At this point it is also convenient to convert the prior and posterior states to their log transformations, $\log L_{S=X,k_T\geq k}(k-1)$ and $\log L_{S=X,k_T\geq k}(k)$, so that in our examples and figures, two evidence states $z$ and $-z$ denote equal evidence strength in opposing directions.

### 4.3. Estimating the posterior state distribution

Suppose that the observation process terminates when the log likelihood ratio,

$$\log L(k) = \log\left[\frac{P(\psi_1,\ldots,\psi_k|S=\text{Signal})}{P(\psi_1,\ldots,\psi_k|S=\text{Noise})}\right],$$

first exceeds one of two boundaries, $b_{\text{yes}}(k)$ or $b_{\text{no}}(k)$, for the first time, where $b_{\text{yes}}(k) > b_{\text{no}}(k)$ for each $k$. Suppose further that these boundary values may depend on $k$ but are constant across trials. Under these conditions, the set of observable terminal evidence states that share the same stopping time $k$ are either greater than $b_{\text{yes}}(k)$ or less than $b_{\text{no}}(k)$, due to the stopping condition. Thus, they are random samples from the left and right tails of a single distribution, that is, the distribution of $\log L_{S=X,k_T\geq k}(k)$. The problem is therefore to estimate the 'complete' posterior state distribution, $P\left(\log L_{S=X,k_T\geq k}(k)=x\right)$, from observations taken from its left and right tails.

This key insight is demonstrated in the upper panel of Fig. 1, which shows several possible trajectories of a process that would produce a response at time $k = 7$ or later, when the stopping condition depends on $k$ (the boundaries converge in time) but is constant across trials. When the observer responds *yes* at time $k = 7$, the observed value of $\log L_{S=X,k_T=k}(k)$ is an observation from the right tail of the $\log L_{S=X,k_T\geq 7}(7)$ distribution, and when the observer responds *no*, the observed value $\log L_{X,k_T=k}(k)$ is an observation from the left tail of the $\log L_{S=X,k_T\geq 7}(7)$ distribution. When the observer responds at some time later than $k = 7$, the state of the process at time 7, $\log L_{S=X,k_T\geq 7}(7)$, was some unobserved value in the middle section of the $\log L_{S=X,k_T\geq 7}(7)$ distribution. The size of the middle section is determined by the difference between the two stopping boundaries, $b_{\text{yes}}(k) - b_{\text{no}}(k)$, at time $k = 7$.

Fig. 2 illustrates the interpolation problem that the experimenter faces when the distribution of the terminal evidence state, $\log L_{S=X,k_T=k}(k)$, must be estimated from a finite sample. The distribution shown in the figure is the result of simulating 10 000 *Signal* trials of a relatively simple sequential detection model and recording the evidence state when the process terminated at time $k = 3$. The boundaries were constant in time, $b_{\text{yes}}(k) = -b_{\text{no}}(k) = 0.5$, and the sensory effects distributions at each $k$ were normal with $\mu_{\text{Signal}}(k) - \mu_{\text{Noise}}(k) = 1$ and $\sigma^2_{\text{Signal}}(k) = \sigma^2_{\text{Noise}}(k) = 1$. The total area of the missing piece (denoted by the dashed rectangle) is determined by the frequency with which $k_T > 3$, making it not too difficult to approximate, with reasonable accuracy, the two stopping boundaries and the shape of the distribution between them.

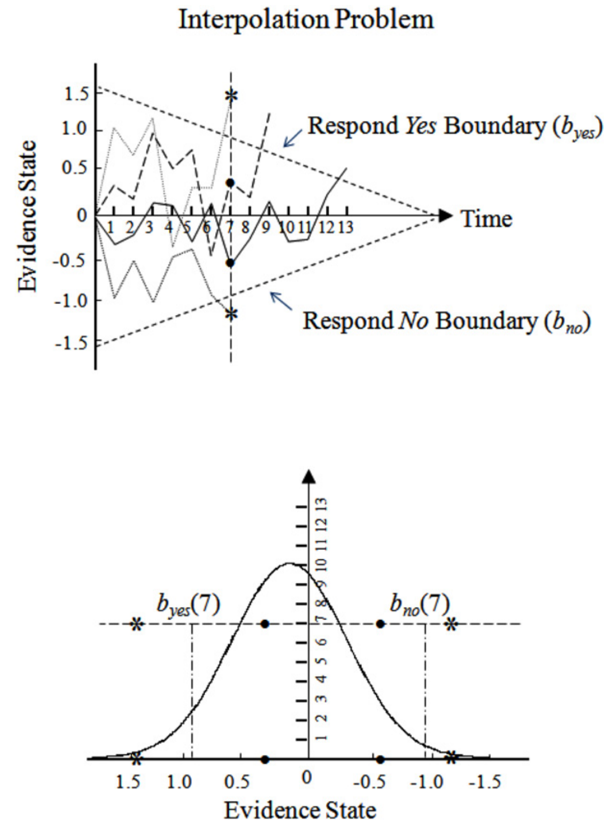


**Fig. 1.** Upper panel: Four possible sample paths of a sequential detection process that terminates at time 7 or greater. Lower panel: the evidence state at time 7 is an observable value in the left or right tail of a distribution when the response occurs at time 7, and an unobservable value in the middle of the distribution when the response time is any value greater than 7.

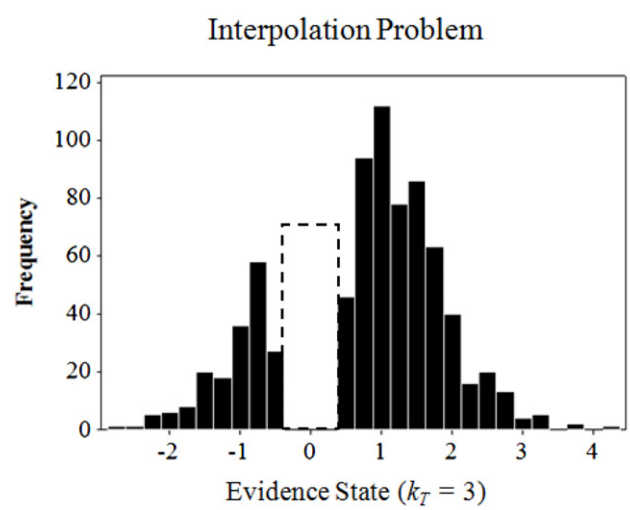


**Fig. 2.** An example of an interpolation problem when the stopping condition is invariant across trials. The middle portion of the underlying evidence state distribution at time $k = 3$ conditioned on $k_T \geq 3$ must be guessed from the shapes of the tails and the size of the rectangle, which is determined by the relative frequency with which the responses are given at some time later than $k_T = 3$.

### 4.4. Determining the prior state distribution

Once an interpolation procedure is developed to find the distribution of $\log L_{S=X,k_T\geq k}(k)$ for each $k$, the prior state distributions can be obtained by a simple transformation, with the interpolated middle portion of the posterior distribution for stopping time $k$

becoming, after multiplication by a constant, the prior state distribution for observation $k+1$. That is, the distribution of the evidence state at $k$ conditioned on $k_T \geq k+1$ – i.e., the prior state distribution for observation time $k+1$ – is the distribution of the posterior state for observation $k$, conditioned on this posterior state being in the interval $[b_{no}(k), b_{yes}(k)]$,

$$P\left(\log L_{S=X,k_T \geq k+1}(k) = z\right) = P\left(\log L_{S=X,k_T \geq k}(k) = z \mid \log L_{S=X,k_T \geq k}(k) \in [b_{no}(k), b_{yes}(k)]\right),$$

for values of $z$ in the interval $[b_{no}(k), b_{yes}(k)]$. Since the probability that $\log L_{S=X,k_T \geq k}(k)$ is in $[b_{no}(k), b_{yes}(k)]$ is the probability that the process does not terminate at $k$, given that it has not terminated yet, $P(k_T > k \mid k_T \geq k, S = X)$, the interpolated middle portion of the posterior state distribution for observation time $k$ is

$$P\left(\log L_{S=X,k_T \geq k+1}(k) = z\right) = \frac{P\left(\log L_{S=X,k_T \geq k}(k) = z\right)}{P(k_T > k \mid k_T \geq k, S = X)},$$

for values of $z$ in $[b_{no}(k), b_{yes}(k)]$.

### 4.5. Distributional assumptions

Assuming that the sensory observations at time $k$ do not depend on the observations prior to $k$, the distribution of the log likelihood ratio, $\log l_{S=X}(k)$, could be recovered from the prior and posterior state distributions by deconvolution (e.g., Sheu & Ratcliff, 1995 and Smith, 1990). Instead of combining an interpolation method with a deconvolution procedure, however, we chose to simply assume that the posterior state distribution is normal at each $k$. This makes it possible to estimate the parameters of the detection process using maximum likelihood.

Under this distributional assumption, the sequential detection model has 6 parameters for each stopping time $k < k_{max}$: the mean and variance of $\log L_{S=X,k_T \geq k}(k)$ for each stimulus type $X$, and the two stopping boundaries $b_{yes}(k)$ and $b_{no}(k)$. At $k_{max}$, the sensory observation process must terminate and the two stopping boundaries must therefore converge on a single 'detection criterion', as in the classical detection model. In order to reduce the size of the problem, we introduced one more assumption about the sensory effects of the stimuli, namely, that they are also univariate normal. This eliminates two of the four sensory parameters at each observation time $k$ (due to implicit assumption of the transformation of the sensory effect to a log likelihood ratio in the definition of the stopping condition, the mean and variance of the distribution of $\psi_k$ on *Noise* trials can be set to 0 and 1, respectively).

### 4.6. Comparing two sequential detection processes

A necessary condition for two experimental conditions to be equivalent with respect to the sensory effects of the stimuli is equivalence with respect to the distributions of the log likelihood ratios, $\log l_{S=X}(k)$, at each observation time $k$. In practice, it is unlikely that these distributions could be estimated with enough accuracy to apply such a strict comparison effectively. Somewhat more feasible would be a weaker condition on the detectability of the signal for each sensory effect $k$. That is, define the probability correct under an optimal decision rule and equal priors,

$$\omega_k = \frac{1}{2}P(\log l_{S=\text{Signal}}(k) > 0) + \frac{1}{2}P(\log l_{S=\text{Noise}}(k) < 0),$$

and compare the values of $\omega_k$ in the two conditions.[4] Since this test is also likely to require a relatively large experiment, the measure that we will adopt here is the detectability of the signal when the priors are equal and the information available is the entire effect sequence $\{\psi_1, \ldots, \psi_{k_{max}}\}$. That is, define

$$\log l_{S=X}(1, \ldots, k_{max}) = \log\left[\frac{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k_{max}} \mid S = \text{Signal})}{P(\psi_{S=X,1}, \ldots, \psi_{S=X,k_{max}} \mid S = \text{Noise})}\right],$$

and then compare

$$\Omega = \frac{1}{2}P(\log l_{S=\text{Signal}}(1, \ldots, k_{max}) > 0) + \frac{1}{2}P(\log l_{S=\text{Noise}}(1, \ldots, k_{max}) < 0),$$

for the two conditions.

### 4.7. Phase two: stochastic stopping rules

The dependence between a human observer's speed and confidence in a perception experiment may be relatively strong, but is never perfect. In many cases, the lowest confidence *yes* and *no* responses will each have non-zero probability at each stopping time. This means that with a sufficiently large sample, the lowest value of $\log B_{\psi_m,R}(x, \text{yes})$ and the highest value of $\log B_{\psi_m,R}(x, \text{no})$ will be observed at each $k$. In order for the likelihood of the data under the model defined above to be non-zero, therefore, the stopping boundaries, $b_{yes}(k)$ and $b_{no}(k)$, at each time $k$ must be the minimum of $\log B_{\psi_m,R=\text{yes}}$ and the maximum of $\log B_{\psi_m,R=\text{no}}$, respectively, which would mean that they are constant in $k$, and close to zero.

At least from the statistical decision theory point of view, the simplest explanation for the strong but imperfect relationship between confidence and response time is that the observers' stopping condition at time $k$ varies across trials, which would happen if the observers' speed incentives can change across trials or if there is some random error in their subjective estimates of the strength of the sensory evidence as it develops during a trial. A similar hypothesis is often put forward by proponents of classical detection theory in order to account for some well-known violations of the static detection model's assumptions, such as the effects of prior outcomes on the supposed placement of the detection criterion (e.g., Benjamin, Diaz, & Wee, 2009, Durlach & Braida, 1969, Ell, Ing, & Maddox, 2009, Healy & Kubovy, 1981, Kubovy, 1977, Mueller & Weidemann, 2008, Nosofsky, 1983, Treisman & Williams, 1984 and Vogels & Orban, 1986).

The effect of adding noise to the stopping boundaries on the interpolation problem is illustrated in Fig. 3, using $\sigma^2_{b_{yes}(k)} = \sigma^2_{b_{no}(k)} = 0.25$ in the model defined above. The noise was normal but truncated so that $b_{yes}(k) \geq 0$ and $b_{no}(k) \leq 0$. Visually, guessing the (average) boundaries and filling in the gap is more difficult, despite the fact that the probability that the state falls between the two boundaries (the probability that $k_T > 3$) can be estimated. However, since we are already assuming that the evidence states, $\log L_{S=\text{Signal},k_T \geq k}(k)$ and $\log L_{S=\text{Noise},k_T \geq k}(k)$, have normal distributions at each stopping time $k$, adding noise to the boundaries does not change the parameter estimation problem in any substantial way. To allow for the possibility of stochastic stopping rules, we assumed that the two boundaries at each $k$ have independent normal distributions, with this noise truncated so that the *yes* response boundary is positive and the *no* response boundary is negative.

### 4.8. Phase three: The non-sensory component of the response time

Once the observation process is terminated, the decision maker must still select and execute a terminal response, which in our paradigm is a simultaneous *yes*-or-*no* judgment and confidence

[4] Note that we are assuming here that $\log l_{S=X}(k) = 0$ has zero probability.

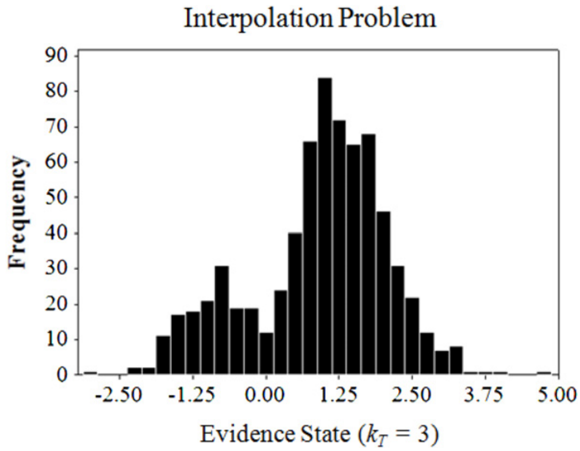


**Fig. 3.** The interpolation problem when noise is added to the two stopping boundary values that define the stopping condition at time $k$. The boundary noise 'blurs' the edges of the gap, making it more difficult to fill in the gap and thereby estimate the distribution of the sensory evidence state.

rating. Unless this extra time component is a constant, it is impossible to determine the length of the delay, $D$, between the stopping time and the response time, $RT$, making it impossible to estimate the stopping time, $k_T$. This would seem to rule out any empirically feasible method of estimating the distributions of $\log L_{S=X,k_T\geq k}(k)$ for any given stopping time $k$. However, this initial impression turns out to be overly pessimistic.

In a discrete time model with a maximum $RT$, the sensory and non-sensory components of the observable decision time have a finite number of possible integer values. Let $k_{\max}$ be, as before, the maximum possible stopping time, and let $d_{\max}$ be the maximum possible delay, $D \leq d_{\max}$. For now, the units of $k_{\max}$ and $d_{\max}$ are determined by the sampling rate of the experimenter's $RT$ timing device, e.g., if the $RT$ is recorded in ms, then $k_{\max}$ and $d_{\max}$ are in ms. The sum, $k_{\max} + d_{\max}$, is the maximum possible $RT$ that the experimenter will allow.

When the stopping time, $k_T$, is equal to a given value $k$ on a given trial, the observed terminal evidence state, $\log L_{S=X,k_T=k}(k)$, will have the value $RT = k+D$ assigned to it, where $D$ is the amount of delay required to select and execute the response on the given trial. The distribution of $\log B_{S=X,R,\psi_m}$ conditioned on a given $RT$ is therefore a mixture of the tails from different distributions, with the relative proportion of a given distribution, $P(\log L_{S=X,k_T=k}(k) = z)$, in the mixture when $RT = i$ being the relative frequency with which the stopping time is $k$ when the $RT$ is $i$, $P(k_T = k \mid RT = i, S = X)$.

The amount of mixing in the distribution of evidence states when $RT = i$ will depend on the number of combinations of stopping and delay times that add to $i$, and will therefore increase as the $RT$ increases. Unless the maximum possible delay is relatively small, the size of the problem will increase very quickly. For example, with $k_{\max} = 5$ and $RT_{\max} = 6$, there are already 13 additional parameters per stimulus. Fortunately, however, the total number of mixture parameters is considerably larger than the number that would need to be estimated, due to three separate factors:

- To the extent that the selection and execution time is under the observer's control, it should be subject to the same speed pressure that is presumed to cause the observer to terminate the sensory observation process prior to $k_{\max}$ on some trials, despite the instruction to maximize accuracy. Therefore, it is reasonable to assume that any variation in the delay would be due to random fluctuations of some kind, as opposed to some deliberate (and hence irrational) strategy. In such a case, any amount of delay that has non-zero probability when the stopping time is $k$ should also have non-zero probability when the stopping time is any other possible value $j$. Under this assumption, the maximum $RT$ must be the maximum stopping time plus the maximum delay, $k_{\max} + d_{\max}$. Several combinations can therefore be eliminated.
- The weights that define the relative proportions in a given mixture (i.e., for a given $RT$) must add to one. Forcing the predicted frequencies in each cell of a column to add to the observed total (the frequency of the given $RT$) eliminates one mixture parameter from each column.
- For each possible stopping time, the delay time $D$ must be one of $d_{\max}$ possible values. The last mixture parameter in each row (i.e., for each stopping time $k$) is therefore completely constrained by the predicted frequency of the stopping time.

**Table 1**
The delay time mixture parameters that can be eliminated due to one of three factors, for the case $k_{\max} = 4$ and $RT_{\max} = 8$. The weight in a cell is $w_{k,i} = P(k_T = k \mid RT = i, S = X)$. Factor 1 ($F1$): each delay has non-zero probability at each stopping time; Factor 2 ($F2$): For a given stopping time $k$, $\sum_i w_{k,i} P(RT = i \mid S = X) = P(k_T = k \mid S = X)$, and $P(k_T = k \mid S = X)$ is determined by the parameters of the sensory observation process; Factor 3 ($F3$): For a given $RT = i$, $\sum_k w_{k,i} = 1$.

| | | $RT$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $k$ | 1 | $w_{1,1}$ $F3$ | $w_{1,2}$ | $w_{1,3}$ | $w_{1,4}$ $F2$ | $w_{1,5}$ $F1$ | $w_{1,6}$ $F1$ | $w_{1,7}$ $F1$ |
| | 2 | | $w_{2,1}$ $F3$ | $w_{2,2}$ | $w_{2,3}$ | $w_{2,4}$ $F2$ | $w_{2,5}$ $F1$ | $w_{2,6}$ $F1$ |
| | 3 | | | $w_{3,1}$ $F3$ | $w_{3,2}$ | $w_{3,3}$ | $w_{3,4}$ $F2$ | $w_{3,5}$ $F1$ |
| | 4 | | | | $w_{4,1}$ $F3$ | $w_{4,2}$ $F3$ | $w_{4,3}$ $F3$ | $w_{4,4}$ $F3$ |

The effect of these three factors on the total number of parameters is illustrated in Table 1, for the case in which $k_{\max} = 4$ and the maximum $RT$ is 8. The notation in the lower right-hand corners of the cells in the table ($F1$, $F2$, $F3$) indicate which parameters can be eliminated and for which reason. Once the impossible cells and the degrees of freedom are taken into account, the size of the problem is reduced substantially. For example, when $k_{\max} = 2$ and $d_{\max} = 2$, there are no mixture parameters at all, and in the Table 1 example, the 22 total possible parameters are reduced to only 6 that would need to be estimated for a given stimulus.

Of course, even after the number of mixture parameters is reduced to a minimum, $k_{\max}$ and $d_{\max}$ cannot be in thousandths or even tenths of a second, and it would be absurd to propose that there might be only about three or four possible stopping times and three or four possible delays in a real experiment. It is not at all absurd, however, to suspect that reducing a relatively accurate measure of time to a relatively crude one merely produces a relatively crude picture of the dynamics of the detection process—and even this crude description turns out to be sufficient to produce a measure that controls for the effects of the stopping condition on an observer's overt performance.

## 5. Empirical application

In order to apply the sequential analysis, the experimenter must solicit a confidence rating and a *yes–no* detection response, and also record the response time. Ideally, the rating and detection responses should be executed simultaneously, so that the downstream constraint is likely to be justified. A numerosity discrimination experiment reported recently by Mueller and Weidemann (2008) satisfied all of these conditions, and also included three different prior probability conditions. Although the authors did not attempt to vary the speed pressure or detectability, there were enough participants so that individual differences can substitute

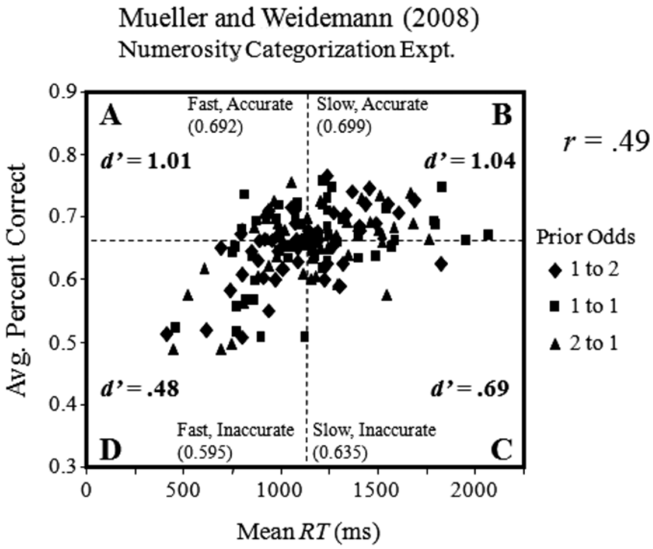


**Fig. 4.** The relationship between average accuracy (the arithmetic average of the hit and correct rejection rates) and mean *RT* in a categorization experiment with fifty participants and three base rate conditions. Accuracy does not depend on the prior probability condition (see legend), but it does depend on speed, with faster participants tending to make more errors. The four quadrants defined by the dashed lines in the figure represent the combination of a median split on accuracy with a median split on speed.

for an explicit experimental manipulation. That is, because both the detectability of the signal and the parameters of the stopping rule will vary among participants, a median split on participants' accuracy crossed with a median split on their speed should simulate a two-by-two factorial combination of a sensory and a speed pressure variable.

The results of this analysis for the three prior probability conditions of Mueller and Weidemann (2008)'s experiment are shown in Fig. 4. According to sequential theory, the manipulation defined by moving diagonally from the lower left (**B**: fast, inaccurate) to the upper right quadrant (**D**: slow, accurate) should be mostly or entirely a speed–accuracy trade-off effect (i.e., decisional), while the manipulation defined by the negative diagonal – slow, inaccurate to fast, accurate performance – should be mostly or entirely a detectability effect. A successful recovery of the participants' ability to discriminate would therefore yield equal values in quadrants **B** and **D** in Fig. 4 and different values in quadrants **A** and **C**.

### 5.1. Data preparation

The participants in the experiment completed 240 trials in each of the three prior probability conditions (1 to 3, 1 to 1 and 3 to 1 *Signal* to *Noise* ratios). The confidence rating scale was bipolar with 4 levels per response and an explicit *yes*-or-*no* cutoff in the middle. Excluding the results from three of the participants who performed at a chance level, the total sample size was 33 440, which was enough to allow us to set the starting point of the final (slowest) *RT* interval to 4 s (roughly the 99th percentile of the dataset) and the total number of response time categories (intervals) to 7. The width of the *RT* intervals was 570 ms, with at least 50 samples in each interval in each of the conditions to be tested. Letting the fastest *RT* interval represent one unit of sensory observation time plus one unit of delay, the slowest *RT* interval is $k_{\max} + d_{\max} = 8$.

Results of the sequential analysis are shown in Fig. 5. In each of the four panels, the estimated detectability of the signal, $\Omega$, is plotted as a function of the maximum stopping time, $k_{\max}$ (i.e., the number of independent detection models in the sequential decision process). The model with the smallest value, $k_{\max} = 1$, is the signal detection theory estimate of sensitivity

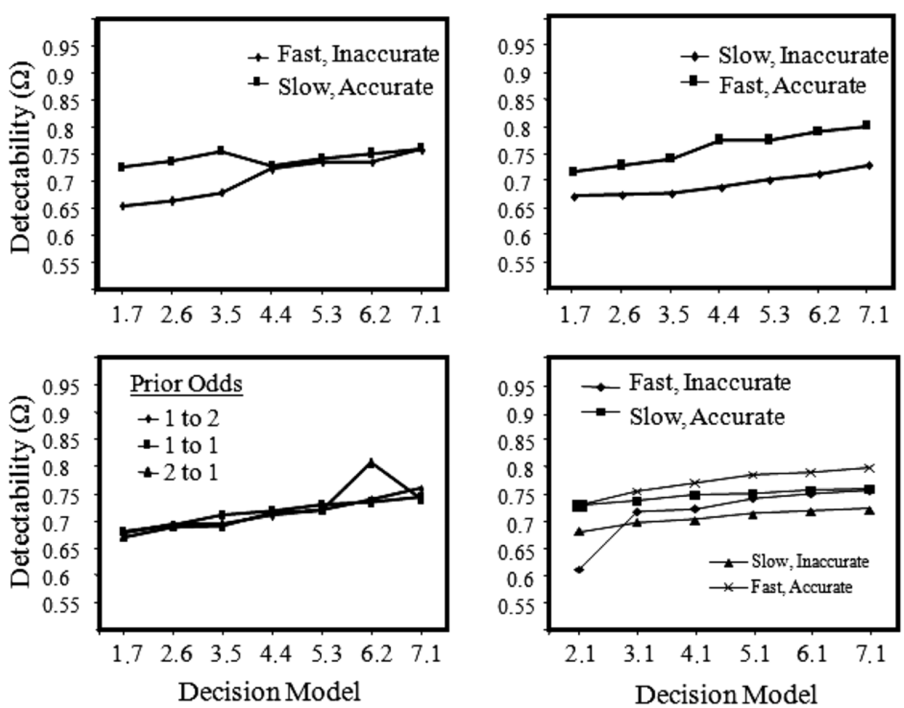


**Fig. 5.** Results of the sequential analysis of the four conditions from Mueller and Weidemann's (2008) experiment that were presented in Fig. 4. In the upper two and lower left panels, $k_{\max} + d_{\max} = 8$ (i.e., the *RT* scores were divided into seven intervals) and the detectability estimate $\Omega$ is compared under the different possible combinations of $k_{\max}$ and $d_{\max}$, as indicated on the abscissa (e.g., "3, 5" is $k_{\max} = 3$ and $d_{\max} = 5$). When $k_{\max} = 1$, the sequential measure $\Omega$ is the static signal detection theory estimate. In the lower right panel, estimates of $\Omega$ for different values of $k_{\max}$ are compared when the delay is ignored (assumed to be constant, $d_{\max} = 1$). In each panel, the estimated detectability of the signal increases with the 'resolution' of the sensory component of the model, $k_{\max}$. The estimates coincide under different priors (lower left panel) and converge in the different speed–accuracy trade-off conditions (upper left and lower right panels). In the 'true' detectability effect comparison (quadrants **A** and **C** in Fig. 4) the difference in estimated detectability is independent of or increasing with $k_{\max}$.

(converted to a percent correct score). The upper two panels of Fig. 5 compare the two speed/accuracy contrasts, one of which should be a signal detectability effect (right panel) and the other which should be due to decision bias (left panel). At $k_{\max} = 1$ (the classical detection model), the two manipulations are indistinguishable, both being apparent sensitivity effects, but with different magnitudes. As $k_{\max}$ increases, however, the size of the difference in estimated detectability is reduced to zero in the speed/accuracy trade-off comparison, while the difference persists, as expected, in the supposed detectability contrast. Thus, the analysis correctly eliminates the illusion of a detectability effect caused by the speed–accuracy trade-off, without confounding the true detectability effect that presumably distinguishes the slow, inaccurate participants from the fast, accurate participants.

The lower left panel of Fig. 5 compares the three prior probability conditions for each value of $k_{\max}$. With the exception of one outlier, there is virtually no difference in estimated detectability in the three different conditions for any value of $k_{\max}$. The results in the lower right panel confirm the 'robustness' of the method that is suggested by the relatively fast convergence of the two estimated functions in the bias condition (upper left panel). That is, partitioning the *RT* data into as few as three speed groups and fixing the delay time to 1 unit (i.e., assuming that the delay contributes virtually nothing to the total variation in *RT*) is enough to effectively remove the false performance advantage, with no loss of power to detect the true change in the detectability contrast.

### 5.2. Non-sensory components of the response time

The detectability measure $\Omega$ would not be expected to be constant under different speed–accuracy trade-off conditions unless its underlying theory adequately describes the dependence of response accuracy on the stopping time of the sensory observation process. Intuitively, therefore, the fact that the estimate of $\Omega$ is most invariant (the measure is most 'successful') under a speed–accuracy trade-off manipulation when the maximum proportion of time in which the sensor is assumed to be "on", $k_{\max}/(k_{\max} + d_{\max})$, is large compared to the maximum proportional amount of the delay, $d_{\max}/(k_{\max}+d_{\max})$, is an indication that the non-sensory component does not contribute much to the total variation in *RT*. The fact that the estimated detectability increases with $k_{\max}$ also suggests that the delay time is relatively small.

In principle, a model selection criterion could be recruited to determine the best choice of $k_{\max}$ and $d_{\max}$ in the model (e.g., McGrory & Titterington, 2007, Myung & Pitt, 1997 and Zhao, Krishnaiah, & Bai, 1986), and in the future it may be worthwhile to explore this possibility. At least in the case of Mueller and Weidemann's (2008) study, however, the question becomes moot, because the conclusions that would be drawn about the non-sensory component did not depend on the choice of $k_{\max}$.

The estimated joint distributions of $k_T$ and $D$ for the three most illustrative cases, $k_{\max} = 3$, 4 and 5, are shown in Fig. 6. Increasing the amount of variability that can be assigned to the non-sensory component (i.e., decreasing $k_{\max}$) has no appreciable effect on the amount that is in fact attributed to this secondary process. In each case, virtually the entire distribution is focused on the smallest possible delay time, $D = 1$, indicating that there is little or no meaningful variation in the non-sensory part of the *RT*.

### 5.3. Dynamics of the detection process

Probably the single most influential source of support for the classical detection model is the supposed invariance of the sensitivity measure $d'$ under different priors and payoffs. Tanner and Swets (1954) considered this to be an important demonstration of the 'internal consistency' of the theory, and its acceptance has caused many investigators to ignore other empirical results, such as the relationship between response time and accuracy, that seem to unequivocally reject the fixed-sample assumption of the model.

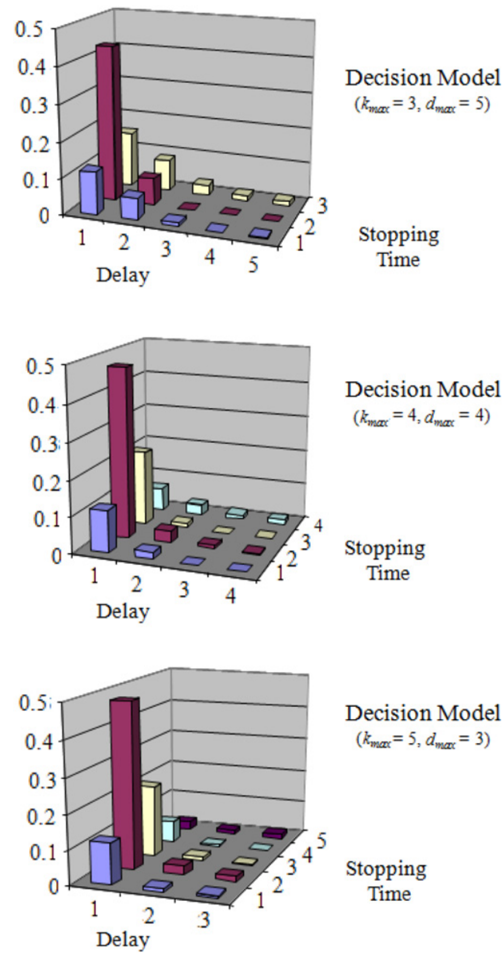


**Fig. 6.** Estimates of the joint distribution of stopping and delay times for three illustrative decision models. In each case, almost all of the variation in the observable *RT* is attributed to variation in the sensory observation process.

One way to reconcile the contradictory findings is to assume that under the conditions that appear to validate the fixed-sample model, the detectability of the signal at a given observation time, $\omega_k$, decreases with $k$. Under these conditions, the stopping boundaries would be expected to converge in time due to the speed pressure, and in such a case the sequential detection model predicts the invariance that is observed in $d'$ (MacDonald & Balakrishnan, submitted for publication).

Estimates of $\omega_k$ and the stopping boundary functions are plotted in the two panels of Fig. 7. Since the conclusions are the same, only the results from the 'longest' sensory observation (i.e., with a constant delay, $d_{\max} = 1$) and the largest sample size (combining across the three prior probability conditions) are shown in the figure. After an initial, sharp increase, the detectability function declines at each point except the final one. The two endpoints each have interpretation issues associated with them, the first time interval (0–570 ms) being, technically, shorter than it is supposed to be (the fastest response is one unit of sensory processing time plus one unit of delay in the model) and the final time interval ($RT \geq 4$ s) being too long. Even ignoring these, the decrease in the function may seem to be modest. It is important to recognize, however, that relatively small changes in $\omega_k$ can have a relatively

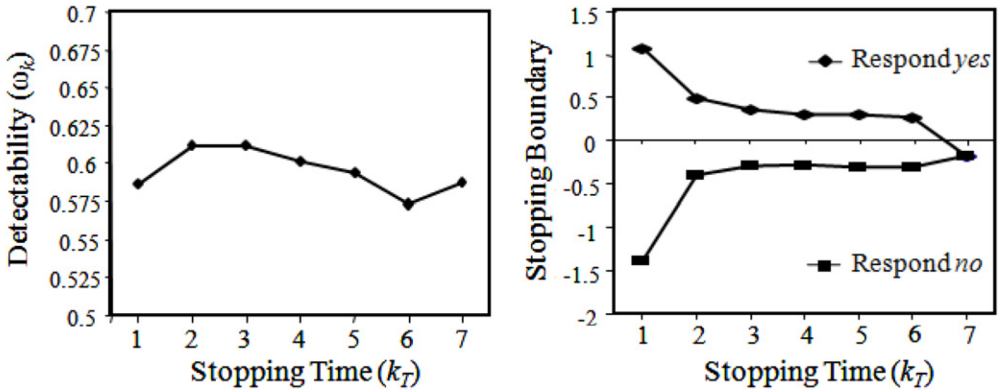


**Fig. 7.** Left panel: Estimated detectability at time $k$, $\omega_k$, when $k_{\max} = 7$ and $d_{\max} = 1$. The first and final points of the estimated function ($k = 1$ and $k = 7$) are not directly comparable to the others due to the sizes of their corresponding *RT* intervals. The function is decreasing in the range that is interpretable. Right panel: The stringency of the estimated stopping condition (distance between the boundaries) relaxes in time, consistent with the loss of signal strength in time.

large impact on the total detectability measure $\Omega$. One way to quantify the effect of the decay in signal strength in time is to compare the total detectability implied by the highest and lowest values of $\omega_k$, respectively. If the maximum value of the estimated $\omega_k$ function, 0.61, was the detectability at each time $k$, then the estimated total detectability would have been $\Omega = 0.78$. In contrast, if the minimum value of estimated $\omega_k$ function, 0.57, was the detectability at each $k$, then $\Omega = 0.69$. The effect of decay in the experiment was therefore a difference on the order of about 9% points in accuracy.

## 6. Discussion

Detection theorists sometimes assert that the static, or fixed-sample, detection model does not make predictions about response times in *yes–no* detection and discrimination tasks. Accepting this statement, there is no inherent contradiction involved in proposing that despite its limitations, the fixed-sample detection model could still be the general framework of a more complete theory, which also predicts response times. It is important to remember, however, that detection theory is, more than anything else, a statistical decision theory. In the fixed-sample detection model, the decision maker stops collecting information because the number of observations is fixed *a priori*. Therefore, the probability that the sensory observation process terminates at time $k$ given that it has not terminated prior to $k$ does not depend on the observations taken up to $k$,

$$P(k_T = k \mid \psi_1 = y_1, \psi_2 = y_2, \ldots, \psi_k = y_k, k_T \geq k) \\ = P(k_T = k \mid k_T \geq k),$$

for each $k$.

In the sequential detection problem, the stopping rule is defined on the sensory effect as it evolves, in such a manner that the cost of the observations can be taken into account along with the cost of errors. The probability of stopping at time $k$ therefore increases as the strength of the evidence collected prior to time $k$ increases. The two stopping rules are clearly mutually exclusive.

Accepting the possibility that experimental devices such as fixing the stimulus presentation time, emphasizing accuracy, and other features of the classical *yes–no* detection task might not be sufficient to justify the fixed-sample detection model, recovering the parameters of the sensory process becomes considerably more difficult. There is an extra dimension, time, and a much wider range of very different, but equally plausible, detection models. Our approach was to adopt the basic structure of the classical threshold signals model, in which there is a flow of information from one noisy device to another, and then introduce additional assumptions that, though implausible as exact statements about the detection process, can be justified with respect to the problem of estimating the detectability of the signal and describing, in qualitative terms, the observer's decision making strategy.

By far the most important assumption of the analysis was that the likelihood ratio transformation of an observer's confidence rating response is a sufficiently accurate measure of the posterior likelihood ratio that is determined by the sensory effect of the stimulus on a given trial. In effect, this presupposes that the observer is able to report the posterior likelihood ratio, which is a highly implausible, if not impossible, proposition. However, under a weak condition (the downstream constraint), the computed likelihood ratio for a given rating response is interpretable as a certain weighted average of the posterior likelihood ratios that co-occur with this rating response. In an important sense, therefore, the likelihood ratio computed from a given rating response is the average amount of sensory information available at the point when the given rating response is selected.

The considerable size and complexity of the space of possible sequential detection models makes it difficult to devise a convincing set of simulations. Most of the weight of our claim about the utility of the detectability measure $\Omega$, therefore, is based on the results of an analysis of empirical data collected by Mueller and Weidemann (2008). The many different concerns that could be raised about how observers make confidence ratings certainly applied to Mueller and Weidemann (2008)'s experiment. For example, the participants were inexperienced with the task and were given no feedback about their consistency in assigning evidence states to rating responses. There were only four levels of confidence on the rating scale, which meant that the parameters of a mixture of normal distributions needed to be estimated from only four points in the left and right tails of the distributions. The experiment was not large enough to perform the analysis for each participant in each condition separately, and the estimates of $\Omega$ were taken therefore over a random combination of different sensory and decisional processes. Despite these many limitations, it was sufficient to divide the responses into only three different levels of speed in order to distinguish the speed–accuracy trade-off manipulation from the sensory effect manipulation. As a basis for categorizing experimental variables, therefore, the measure $\Omega$ appears to be robust to the kinds of violations of its underlying assumptions that would be expected in a typical experiment.

## Acknowledgments

We thank Christoph Weidemann and Jim Townsend for their helpful comments and Ehtibar Dzhafarov for many discussions and detailed comments on previous drafts.

## References


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

Balakrishnan, J. D. (1998). Some more sensitive measures of sensitivity and response bias. *Psychological Methods*, *3*, 68–90.

Balakrishnan, J. D. (1999). Decision processes in discrimination: fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception & Performance*, *25*, 1189–1206.

Balakrishnan, J. D., & MacDonald, J. A. (2003). Decision criteria do not shift: reply to Treisman. *Psychonomic Bulletin & Review*, *9*, 858–865.

Balakrishnan, J. D., & MacDonald, J. A. (2008). Decision criteria do not shift: commentary on Mueller and Weidemann. *Psychological Bulletin & Review*, *15*, 1022–1030.

Balakrishnan, J. D., MacDonald, J. A., & Kohen, H. S. (2003). Is the area measure a historical anomaly? *Canadian Journal of Experimental Psychology*, *57*, 238–256.

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: applications to recognition memory. *Psychological Review*, *116*, 84–115.

Birdsall, T. G., & Roberts, R. A. (1965). Theory of signal detectability: deferred-decision theory. *Journal of the Acoustical Society of America*, *37*, 1064–1074.

Carterette, E. C., Friedman, M. P., & Cosmides, R. (1965). Reaction-time distributions in the detection of weak signals in noise. *Journal of the Acoustical Society of America*, *38*, 531–542.

Degroot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw Hill.

Donders, F. C. (1969). Over de snelheid van psychische processen [On the speed of psychological processes]. (W. Koster, Trans.). In W. Koster (Ed.), *Attention and performance*: *Vol. 2* (pp. 412–431). Amsterdam: North-Holland, (Original work published 1868).

Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *46*, 372–383.

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

Ell, S. W., Ing, A. D., & Maddox, W. T. (2009). Criterial noise effects on rule-based category learning: the impact of delayed feedback. *Attention, Perception, & Psychophysics*, *71*, 1263–1275.

Emmerich, D. S., Gray, J. L., Watson, C. S., & Tanis, D. C. (1972). Response latency, confidence, and ROCs in auditory signal detection. *Perception & Psychophysics*, *11*, 65–72.

Healy, A. F., & Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning & Memory*, *7*, 344–354.

Henmon, V. A. (1911). The relation of time of a judgment to its accuracy. *Psychological Review*, *18*, 186–201.

Johnson, D. M. (1939). Confidence and speed in the two-category judgment. *Archives of Psychology*, *34*, 1–53.

Kolev, V., Falkenstein, M., & Yordanova, J. (2006). Motor-response generation as a source of aging-related behavioural slowing in choice-reaction tasks. *Neurobiology of Aging*, *27*, 1719–1730.

Kubovy, M. (1977). A possible basis for conservatism in signal detection and probabilistic categorization tasks. *Perception & Psychophysics*, *22*, 277–281.

Lawson, J. L., & Uhlenbeck, G. E. (1950). *Threshold signals*. New York: McGraw-Hill.

Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ, England: Lawrence Erlbaum.

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

MacDonald, J.A., & Balakrishnan, J.D. (2011). Why static detection theory should be replaced by dynamic detection theory (submitted for publication).

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide* (2nd ed.) Mahwah, NJ, US: Lawrence Erlbaum.

McGrory, C. A., & Titterington, D. M. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics & Data Analysis*, *51*, 5352–5367.

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: an explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, *15*, 465–494.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: a Bayesian approach. *Psychonomic Bulletin & Review*, *4*, 79–95.

Nosofsky, R. M. (1983). Information integration and the identification of stimulus noise and criterial noise in absolute judgment. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 299–309.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions IRE Profession Group on Information Theory*, *PGIT-4*, 171–212.

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

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, *117*, 864–901.

Sheu, C., & Ratcliff, R. (1995). The application of Fourier deconvolution to reaction time data: a cautionary note. *Psychological Bulletin*, *118*, 285–299.

Smith, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time data. *Psychological Bulletin*, *108*, 533–550.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.

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

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*, 68–111.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 582–600.

Vogels, R., & Orban, G. A. (1986). Decision processes in visual discrimination of line orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 115–132.

Zhao, L. C., Krishnaiah, P. R., & Bai, Z. D. (1986). On detection of the number of signals in presence of white-noise. *Journal of Multivariate Analysis*, *20*, 1–25.