Signal Detection Theory with Ambivalent or Missing Responses

Samuel R. Mathias, Emma E. E. M. Knowles, and David C. Glahn

Yale University

Author Note

Samuel R. Mathias, Department of Psychiatry, Yale School of Medicine, Yale University, New Haven, Connecticut; Emma E. E. M. Knowles, Department of Psychiatry, Yale School of Medicine, Yale University, New Haven, Connecticut; David C. Glahn, Department of Psychiatry, Yale School of Medicine, Yale University, New Haven, Connecticut.

This work was supported in part by grants from the National Institute of Mental Health.

Correspondence concerning this article should be addressed to Samuel R. Mathias, Suite 3014, 2 Church Street South New Haven, CT 06519.

E-mail: samuel.mathias@yale.edu

Abstract

Signal detection theory (SDT) allows researchers to calculate psychologically meaningful measures from behavioral data, so long as the design of the experiment conforms to one of the traditional SDT paradigms. These paradigms usually require subjects to make an affirmative response (e.g., "yes" or "no") to each trial. Occasionally, researchers may wish to allow subjects to make ambivalent responses (e.g., "I don't know"), or not force them to respond to all trials. Here, we propose modifications to the standard SDT models that allow ambivalence and omissions for yes-no and two-alternative forced-choice experiments. Under these models, the equations for calculating sensitivity (d') are the same as those of regular SDT, but the equations for response bias (c) are different. Furthermore, we derive two new psychologically meaningful measures we call "indices of uncertainty," u and u', that allow researchers to quantify subjects' ambivalence.

Keywords: signal detection theory, methods, modeling, statistics

Signal Detection Theory with Ambivalent or Missing Responses

Introduction

Signal detection theory is a widely used framework that allows researchers to transform the raw data from an experiment—the counts of trials and responses—into psychologically meaningful measures of sensitivity and bias. Methods for calculating these measures are generally straightforward to implement for common experimental designs (see Green & Swets, 1966; Macmillan & Creelman, 2005). The two most common designs are yes-no (YN) and two-alternative forced-choice (2AFC).

In most experiments with SDT-friendly designs, subjects are required make an affirmative response (e.g., “yes” or “no;” “first” or “second”) to each trial. However, researchers occasionally may wish to conduct experiments that do not require affirmative responses. For example, the standard YN and 2AFC designs could be modified to include an additional response that allows subjects to indicate ambivalence towards the other responses (e.g. “I don’t know”). However, including ambivalent responses flouts the requirements of regular SDT models, making it less straightforward to calculate sensitivity and bias.

Consider the recent study by Laskowska, Łęski, and Koziorowski (2015), in which patients with Parkinson’s disease, patients with schizophrenia, and healthy controls completed a test of facial emotion recognition. On each trial, subjects saw a photograph of a face and reported which of six possible emotions were conveyed, and which were not, by choosing one of three possibilities per emotion: “shown,” “not shown,” and “hard to say.” It was possible for a photograph to display more than one emotion. The authors analyzed the data using a standard SDT model by considering each response (six per photograph) as a separate YN trial. The authors found that, on average, the Parkinson’s disease and schizophrenia patients had lower sensitivity to facial emotions than the controls, and that the Parkinson’s disease patients adopted a more liberal response strategy than the controls. However, the authors had to recode “hard to say” responses before they could calculate sensitivity and bias, which vitiates the interpretation of their statistically significant group differences.

Other examples of experiments that do not conform to the requirements of regular SDT models can be found in the neuroimaging literature, where subjects are often required to perform specific tasks while their brain activity is recorded. Such experiments typically require strict control over the time intervals between trials and cannot guarantee that subjects will respond in time to every trial. In the functional magnetic resonance imaging (fMRI) study by Kreitewolf, Friederici, and von Kriegstein (2014), subjects performed two tasks. In one task, subjects heard sequences of words, and indicated whether each word was the same as or different to the word that preceded it. In the other task, subjects heard the same sequences, and indicated whether sequential words had the same or a different intonation (either rising or flat). The researchers were primarily interested in differences in brain activity related to phonological and lexical processing; therefore, to compensate for differences related to task difficulty, subjects' percent correct scores were used as covariates in the fMRI analysis. Missing responses were simply labelled as incorrect. Under these circumstances, SDT measures of sensitivity and bias might have been better covariates than percent correct.

In the present paper, we present a method for dealing with ambivalent or missing responses in YN and 2AFC experiments. This method involves modifying regular SDT models, and re-deriving the equations for estimating sensitivity and bias. We show that, for both YN and 2AFC experiments, the equations for estimating sensitivity are actually the same as those of the regular SDT models, but the equations for estimating bias are different. Furthermore, our modified models also allow a new psychologically meaningful SDT measure to be estimated from the data. We call this measure the "index of uncertainty" because it reflects the degree of uncertainty or ambivalence within the decision process. We performed simulations to test the utility of our models in practice, and examined how previously used recoding strategies (e.g., Kreitewolf et al., 2014;

Laskowska et al., 2015) effect the estimates of SDT measures.

SDT models

Regular YN model

The *equal-variance Gaussian SDT model for YN experiments*, hereby referred to as the regular YN model, was the first SDT model to be described (Peterson, Birdsall, & Fox, 1954; Tanner & Swets, 1954) and forms the foundation of all SDT analysis. It has been outlined in detail by many previous authors (e.g., Green & Swets, 1966; Macmillan & Creelman, 2005). We outline it again briefly here, so that readers can see the difference between this model and our modified version, described later on.

Under the model, each trial of the experiment contains a single stimulus drawn at random from one of two classes, and the subject indicates to which class the stimulus belonged. Let X represent the stimulus class. When $X = 0$, the stimulus is drawn from the first class, and when $X = 1$, the stimulus is drawn from the second class. The model assumes that the subject generates a random variable (or "observation") from the stimulus. These observations are denoted Ψ. If $X = 0$, Ψ has a standard normal probability distribution, $\Psi_{X=0} \sim \mathcal{N}(0, 1)$. If $X = 1$, Ψ has a normal probability distribution with mean d and unit standard deviation, $\Psi_{X=1} \sim \mathcal{N}(d, 1)$. When d is large, the two distributions are distinct, and the subject can easily distinguish between the stimulus classes. If $d = 0$, the classes are indistinguishable. Thus, d is a measure of sensitivity[1]

Let Y represent the subject's response to a trial. When $Y = 0$, the subject responds that the stimulus was in the first class, and when $Y = 1$, the subject responds that the stimulus was in the second class. Under the model, the subject makes a response based on

[1] In the literature, the more commonly used index of sensitivity is d', where the $'$ symbol represents the fact that the distance between the means of the two distributions of observations is "standardized" (i.e., d divided by the pooled standard deviation). However, since we assume here that the standard deviations of both distributions are 1, there is actually no difference between d and d'. The distinction between d and d' becomes important when either standard deviation is not set to 1, such as in unequal-variance SDT models.

whether Ψ is smaller or greater than some fixed threshold (or “criterion”), denoted k. Formally, this decision rule can be written

$$Y = 0 \text{ if } \Psi < k,$$

$$Y = 1 \text{ if } \Psi \geq k.$$

When $k = \frac{d}{2}$, the subject is said to be “unbiased,” since both responses are equally likely. The distance between true k and unbiased k is denoted c. Thus, c is a measure of bias.

The model is summarized in the left panel of Fig. 1. The figure shows two important additional variables. The blue shaded region represents the probability of a “false alarm,” that is, when $Y = 1$ given $X = 0$. We denote this probability f. The green shaded region represents the probability of a “hit,” that is, when $Y = 1$ given $X = 1$, denoted h. These conditional probabilities can be expressed in terms of d and c:

$$\begin{aligned} P\left\{Y = 1 \mid X = 0\right\} = f &= 1 - \Phi\left(k\right) \\ &= 1 - \Phi\left(c + \frac{d}{2}\right), \end{aligned}$$

where $\Phi\left(.\right)$ denotes the standard normal cumulative distribution function; and

$$\begin{aligned} P\left\{Y = 1 \mid X = 1\right\} = h &= 1 - \Phi\left(k - d\right) \\ &= 1 - \Phi\left(c - \frac{d}{2}\right). \end{aligned}$$

Crucially, because f and h can be estimated directly from the data in an experiment, the above equations can be combined and re-arranged to produce estimates of sensitivity and bias. Specifically, the equation for estimating sensitivity is

$$\hat{d} = \Phi^{-1}\left(\hat{h}\right) - \Phi^{-1}\left(\hat{f}\right), \tag{1}$$

where $\Phi^{-1}\left(.\right)$ denotes the probit function, and a caret denotes the *maximum-likelihood estimate* of a variable. The equation for estimating bias is

$$\hat{c} = -\frac{1}{2}\left[\Phi^{-1}\left(\hat{h}\right) + \Phi^{-1}\left(\hat{f}\right)\right]. \tag{2}$$

It is worth noting that the assumptions of the regular YN model can be relaxed in various ways. For example, it is possible for the probability distributions of $\Psi_{X=0}$ and $\Psi_{X=1}$ to be non-Gaussian, or have different variances. Indeed, work on recognition memory has consistently reported phenomena that are inconsistent with the regular YN model but can be explained by assuming unequal variances (Wixted, 2007; Yonelinas & Parks, 2007) or by assuming that $\Psi_{X=0}$ and $\Psi_{X=1}$ are drawn from mixtures of latent distributions (DeCarlo, 2002). Another possibility is to allow k to vary across trials rather than remaining constant over trials (Cabrera, Lu, & Dosher, 2015). However, relaxing any of these assumptions means that the YN model is no longer exactly identified (i.e., the number of independent unknown variables and the number of known variables are not the same), so consequently cannot be solved for d and c. For the purposes of this article, we adhere to these assumptions, but note that future work could extend our models to incorporate unequal variance, non-Gaussian decision variables, and variable criteria.

YN model with ambivalent or missing responses

One can account for ambivalent or missing responses with two modifications to the regular YN model, as shown in the right panel of Fig. 1. The first modification is that the model contains two criteria, denoted a and b. The second modification is to the decision rule. Let Z denote the subject's response. When $Z = 0$, the subject responds that the stimulus is in the first class. When $Z = 1$, the subject responds ambivalently ("I don't know"), or does not respond at all. When $Z = 2$, the subject responds that the stimulus is in the second class. This decision rule can be written as

$$Z = 0 \text{ if } \Psi < a,$$

$$Z = 1 \text{ if } a \leq \Psi < b,$$

$$Z = 2 \text{ if } \Psi \geq b.$$

We can define false-alarm and hit probabilities for this model in an analogous way to the regular model:

$$\begin{aligned} P\left\{Z=2 \mid X=0\right\} &= f = 1-\Phi\left(b\right) \\ P\left\{Z=2 \mid X=1\right\} &= h = 1-\Phi\left(b-d\right). \end{aligned}$$

By combining these equations and solving for d, and replacing f and h with their respective maximum-likelihood estimates from the data, we again arrive at

$$\hat{d} = \Phi^{-1}\left(\hat{h}\right) - \Phi^{-1}\left(\hat{f}\right).$$

In other words, the equation for estimating sensitivity under the YN model with ambivalent or missing responses is the same as under the regular YN model (Eq. 1). However, care must be taken to calculate $\hat{f}$ and $\hat{h}$ appropriately: the numbers of false alarms and hits must be divided by the *total* number of trials (i.e., including trials with ambivalent or missing trials responses), and ambivalent or missing responses must not be recoded.

Under this model, a subject could be said to be unbiased if the criteria a and b were centered on $\frac{d}{2}$, since both affirmative responses would be equally likely in this case. Thus, an appropriate measure of bias, analogous to c from the regular YN model, is the distance between $\frac{d}{2}$ and the midpoint between a and b. We denote this quantity λ. It is related to a, b, and d via the equation

$$\lambda = \frac{1}{2}\left(a+b\right) - \frac{d}{2}.$$

An additional probability to calculate λ from the observations in an experiment. The red shaded region in the left panel of Fig. 1 shows this probability, $P\left\{Z=1 \mid X=0\right\}$, which we denote g and can be expressed as

$$P\left\{Z=1 \mid X=0\right\} = g = \Phi\left(b\right) - \Phi\left(a\right).$$

Like f and h, g can be estimated directly from the data—it is proportion of ambivalent or missing responses when the stimulus was drawn from the first class. Therefore, the maximum-likelihood estimate of λ is given by

$$\hat{\lambda} \;=\; -\frac{1}{2}\left[\Phi^{-1}\left(\hat{f}+\hat{g}\right)+\Phi^{-1}\left(\hat{h}\right)\right]. \tag{3}$$

The model allows a third psychologically meaningful measure to be calculated from the data. The distance between the two criteria, denoted u, represents the degree of uncertainty or ambivalence within the decision process. We therefore refer to it as the "index of uncertainty." This measure can be estimated from the data using the equation

$$\hat{u} \;=\; \Phi^{-1}\left(\hat{f}+\hat{g}\right)-\Phi^{-1}\left(\hat{f}\right). \tag{4}$$

Note that, in the special case where $u = 0$, the modified model reduces to the regular YN model: λ is equivalent to c, and can be calculated using Eq. 2, as usual.

Regular 2AFC model

The regular YN model can be modified to account for data from 2AFC experiments, which allows the calculation of measures of sensitivity and bias (DeCarlo, 2012). On each trial in a 2AFC experiment, the subject is presented with two stimuli per trial. The stimuli are drawn from two different stimulus classes, and presented in random order. Let W denote the presentation order of the stimuli. When $W = 0$, the first stimulus is drawn from the first class, and the second stimulus is drawn from the second class. When $W = 1$, the first stimulus is drawn from the second class, and the second stimulus is drawn from the first class. The subject makes two observations per trial, Ψ_0 and Ψ_1, corresponding to the first and second stimuli, respectively, and then chooses whichever stimulus had the larger corresponding observation. In order to account for bias, a quantity is always added to the first observation prior to the decision being made (see DeCarlo, 2012). Let Y represent the subject's response to a trial. When $Y = 0$, the subject chooses the first stimulus, and when $Y = 1$, the subject chooses the second stimulus. This decision rule can be expressed formally as

$$Y = 0 \text{ if } \Psi_0 + l > \Psi_1,$$

$$Y = 1 \text{ if } \Psi_0 + l \leq \Psi_1,$$

where l is a measure of bias. The model is summarized in the top panels of Fig. 2. The top-left and top-right panels are illustrations of two different trials. In both of them, the subject has just made an observation of the first stimulus. The left panel shows a trial where this stimulus was drawn from the first class (i.e., $W = 0$). The blue shaded region in this figure represents the conditional probability $P\{Y = 1 \mid W = 0\}$, which for convenience we arbitrarily label as a false alarm, f. The panel makes it clear that, unlike under the YN model, the probability of making a false alarm varies from trial to trial depending on the value of Ψ_0. Specifically,

$$P\{Y = 1 \mid W = 0, \Psi_0 = x\} = 1 - \Phi(x + l)$$

where $x \sim \mathcal{N}(d, 1)$. The corresponding equation not conditional on x is

$$P\{Y = 1 \mid W = 0\} = f = \int [1 - \Phi(x + l)]\, \phi(x - d)\, \mathrm{d}x,$$

where $\phi(\cdot)$ is the normal probability density function. Conveniently, equations of this kind can be simplified, leading to

$$f = \Phi\left(\frac{-d - l}{\sqrt{2}}\right).$$

The top-right panel of Fig. 2 shows a trial where the first stimulus was drawn from the second class (i.e., $W = 1$). The green shaded region, $P\{Y = 1 \mid W = 1\}$ or h, is given by the following equations:

$$P\{Y = 1 \mid W = 1, \Psi_0 = y\} = 1 - \Phi(y + l - d),$$

where $y \sim \mathcal{N}(0, 1)$; and

$$\begin{aligned} P\{Y = 1 \mid W = 1\} = h &= \int [1 - \Phi(y + l - d)]\, \phi(y)\, \mathrm{d}y \\ &= \int \Phi(d - y - l)\, \phi(y)\, \mathrm{d}y \\ &= \Phi\left(\frac{d - l}{\sqrt{2}}\right). \end{aligned}$$

The maximum-likelihood estimate of d under the regular 2AFC model is therefore given by

$$\hat{d} = \frac{1}{\sqrt{2}} \left[\Phi^{-1}\left(\hat{h}\right) - \Phi^{-1}\left(\hat{f}\right) \right], \tag{5}$$

and the maximum-likelihood estimate of l is given by[2]

$$\hat{l} = -\frac{1}{\sqrt{2}} \left[\Phi^{-1}\left(\hat{h}\right) + \Phi^{-1}\left(\hat{f}\right) \right]. \tag{6}$$

2AFC model with ambivalent or missing responses

As before, two modifications are required to allow ambivalent or missing responses in 2AFC experiments. The first is to include two additional variables, a and b. The second is to modify the decision rule. Let Z denote the subject's response. When $Z = 0$, the subject chooses the first stimulus. When $Z = 1$, the subject responds ambivalently ("I don't know"), or does not respond at all. When $Z = 2$, the subject chooses the second stimulus. The decision rule is

$$\begin{aligned} Z = 0 \text{ if } &\quad \Psi_0 - a > \Psi_1, \\ Z = 1 \text{ if } &\quad \begin{cases} \Psi_0 - a \leq \Psi_1 \\ \text{and} \\ \Psi_0 + b > \Psi_1 \end{cases} \\ Z = 2 \text{ if } &\quad \Psi_0 + b \leq \Psi_1. \end{aligned}$$

The model is shown in the lower panels of Fig. 2. The bottom-left panel shows a trial where $W = 0$. The blue shaded region f is given by

$$P\left\{Z = 2 \mid W = 0\right\} = f = \Phi\left(\frac{-d - b}{\sqrt{2}}\right).$$

[2] Equation 6 does not appear in previous introductions to SDT, despite entire chapters of authoritative textbooks being devoted to 2AFC (e.g., Green & Swets, 1966; Macmillan & Creelman, 2005). In fact, Macmillan and Creelman suggest (p. 170), "For measuring response bias [in 2AFC], the methods [for YN] are entirely adequate. No $\sqrt{2}$ adjustment is necessary, because (as the reader may not be surprised to learn) bias in one task cannot be predicted from bias in the other."

The bottom-right panel shows a trial where $W = 1$. The green shaded region h is given by

$$P\left\{Z = 2 \mid W = 1\right\} = h = \Phi\left(\frac{d - b}{\sqrt{2}}\right).$$

It should be obvious from the previous sections that the maximum-likelihood estimate of d under the modified 2AFC model is given by Eq. 5, and is the same as under the regular 2AFC model.

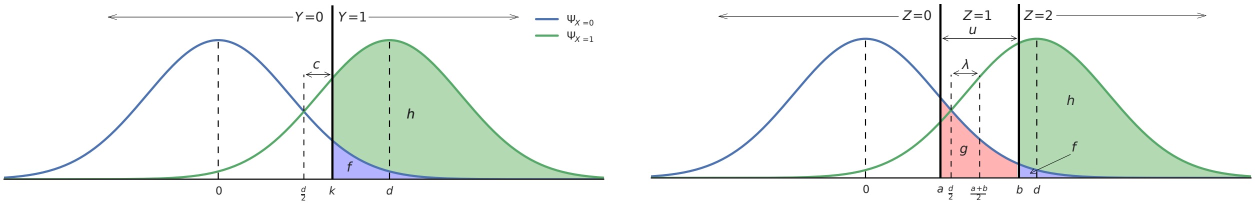

Figure 1. Illustrations of the regular YN model (left panel) and the modified YN model (right panel). In both panels, the two curves represent the probability distributions of observations from the two stimulus classes. Under the regular model, the distance between the means of the distributions, d, is a measure of sensitivity, and the distance between $\frac{d}{2}$ and k, denoted c, is a measure of bias. Both of these measures can be expressed uniquely in terms of the false-alarm and hit probabilities (f and h, respectively). Under the modified model, d is again an appropriate measure of sensitivity and can be estimated in the same way. However, an appropriate measure of bias, λ, is calculated differently, and requires an additional probability, g, to be estimated from the data. The modified model contains an additional measure, u, that represents uncertainty or ambivalence in the decision process.