Journal of Mathematical Psychology 56 (2012) 196–207

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

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

On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation

Lawrence T. DeCarlo

Department of Human Development, Teachers College, Columbia University, 525 West 120th Street, Box 118, New York, NY 10027, United States

ARTICLE INFO

Article history:
Received 20 November 2011
Received in revised form
30 January 2012
Available online 21 March 2012

Keywords:
Signal detection theory
Bias
Forced choice
Gaussian quadrature
Bayesian estimation
Markov chain Monte Carlo

ABSTRACT

The standard signal detection theory (SDT) approach to m-alternative forced choice uses the proportion correct as the outcome variable and assumes that there is no response bias. The assumption of no bias is not made for theoretical reasons, but rather because it simplifies the model and estimation of its parameters. The SDT model for mAFC with bias is presented, with the cases of two, three, and four alternatives considered in detail. Two approaches to fitting the model are noted: maximum likelihood estimation with Gaussian quadrature and Bayesian estimation with Markov chain Monte Carlo. Both approaches are examined in simulations. SAS and OpenBUGS programs to fit the models are provided, and an application to real-world data is presented.

On each trial of a two-alternative forced choice (2AFC) task, two events (e.g., signal and noise) are presented and the observer's task is to indicate which event was the signal. The approach can readily be extended to M alternatives, resulting in an m-alternative forced choice (mAFC) task. The standard signal detection theory (SDT) approach to mAFC uses the proportion correct as the outcome variable and assumes that there is no response bias (e.g., see Macmillan & Creelman, 2005; Wickens, 2002). The assumption of no bias is not made for theoretical reasons, but rather because it simplifies the model and estimation of its parameters.

Although the importance of bias was recognized and discussed early on by mathematical psychologists (e.g., Luce, 1963), the approach was not developed in any detail for SDT models of mAFC because of the complexity of the resulting models. For example, Luce (1963) noted that "The generalization of the two-alternative signal detectability model to the k-alternative forced-choice design is comparatively complicated if response biases are included and very simple if they are not." (p.137).[1] Similarly, Green and Swets (1988), in a discussion of response bias in forced choice, noted that "Our discussion is limited to the two-alternative forced-choice procedure; the analysis for larger numbers of alternatives is complex and, at this date, has not been accomplished." (p. 409).

The SDT model for mAFC with bias is presented here. In particular, a general *decision rule* and *structural model* for forced choice are presented. These can then be used to derive the mAFC model with any number of alternatives. As examples, SDT models for two, three, and four alternatives are derived. As has long been recognized, the models present a bit of a challenge to fit. Two approaches to fitting the models are discussed: maximum likelihood estimation with Gauss–Hermite quadrature and Bayesian estimation with Markov chain Monte Carlo. Both approaches are shown to be simple to implement in standard software. A SAS program, for the maximum likelihood approach, and an OpenBUGS program, for the Bayesian approach, are provided.

The first section briefly reviews the basic SDT situation where an observer responds "yes" or "no" in response to a presentation of a signal or noise. Next, the extension to two-alternative forced choice, recognizing effects of bias, is shown. The approach can be immediately extended to three or more alternatives. Approaches to estimation via maximum likelihood and Bayesian methods are noted, and simulations that examine parameter recovery are presented. An application to real-world data is also presented.

1. SDT and the yes/no procedure

The simplest detection situation consists of the presentation of a signal or noise, with an observer responding yes (signal present) or no (signal absent). Let the variable X indicate presentation of a signal or noise, with $1 =$ signal and $0 =$ noise, and the variable Y indicate the observer's response, with 1 indicating a response of "yes" (signal perceived to be present) and 0 indicating a response of "no" (signal perceived to be absent). A basic idea

E-mail addresses: decarlo@tc.edu, decarlo@exchange.tc.columbia.edu.

[1] The inclusion of bias has been developed for mAFC in Luce's choice theory (see Luce, 1963), which is related to the present approach.

0022-2496/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2012.02.004

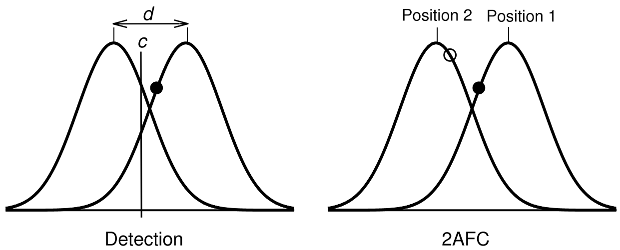

Fig. 1. An illustration of the signal detection approach for detection (left panel) and two-alternative forced choice (right panel). In detection (left panel), the observer reports a signal because their perception (solid circle) in this case is above the criterion; in 2AFC (right panel), the observer chooses Position 1 because their Position 1 perception (solid circle) in this case is above their Position 2 perception (open circle).

in SDT is that the effect on an observer of a presentation of a signal or noise can be represented by a random variable, Ψ, which is a psychological representation of the event. In applications in psychophysics, for example, the psychological representation is usually interpreted as the observer's *perception* of the event, whereas in other applications, Ψ has other interpretations (e.g., as the *familiarity* of a word in recognition memory research).

The observer is viewed in SDT as arriving at an observed response, that is, a decision of yes or no, by using the random variable Ψ along with a *decision rule*,

$$Y = 1 \quad \text{if } \Psi > c,$$
$$Y = 0 \quad \text{if } \Psi \leq c.$$

Thus, the observer responds "yes" if their perception on any given trial is larger than the decision criterion c, otherwise they respond "no". The location of the decision criterion c can be viewed as reflecting the observer's "bias" towards a response of yes or no (see Macmillan & Creelman, 2005), and so c is the same as the bias parameter b discussed below for forced choice. Indeed, an important aspect of SDT is that it recognizes the basic role of response bias in detection, and so it is somewhat odd that bias has tended to be ignored in signal detection approaches to mAFC.

The left panel of Fig. 1 illustrates the decision rule for the yes/no situation. The two distributions represent perceptual distributions associated with signal and noise, with the signal distribution shifted to the right. The distance between the distributions is d and the location of the response criterion is shown by the vertical line marked as c. The solid dot shows the location of a realization from the signal distribution on a trial where the signal was presented. The observer knows that the realization is above the criterion (the vertical line) and so responds "yes", which in this case is a correct response.

The *structural model* (see DeCarlo, 2010) links the psychological variable to the presented event,

$$\Psi = dX + \varepsilon$$

where ε represents random variation in the perception. It follows that the mean of Ψ, that is, the conditional expectation $E(\Psi|X)$, is at zero for a noise presentation ($X = 0$), assuming $E(\varepsilon) = 0$, and is at d for a signal presentation ($X = 1$). If $\varepsilon \sim N(0, 1)$, it follows from the decision rule and structural model that

$$\begin{aligned} p(Y = 1|X) &= p(\Psi > c \mid X) = p(dX + \varepsilon > c) \\ &= p(\varepsilon > c - dX) = 1 - \Phi(c - dX), \\ &= \Phi(-c + dX), \end{aligned}$$

where Φ is the cumulative distribution function (CDF) for the normal distribution. Note that the last step uses the relation $1 - \Phi(a) = \Phi(-a)$, which follows from the symmetry of the normal distribution. The above shows that the basic SDT model with normal underlying distributions is

$$p(Y = 1|X) = \Phi(-c + dX), \tag{1}$$

which is simply a probit model (e.g., see DeCarlo, 2003). The model can also be more generally written for distributions other than the normal by replacing the normal CDF with other CDF's (see below). This is easily implemented by noting that Eq. (1) can be written as a generalized linear model, with the underlying distributions corresponding to the inverse of a link function (see DeCarlo, 1998).

The data for the basic signal detection situation consist of a two by two table, say with X as rows and Y as columns. The essential information is provided by the proportion of "hits" (response of yes when a signal is present) and the proportion of "false alarms" (response of yes when a signal is absent). Thus, there are two observations, the proportion of hits and the proportion of false alarms, and two parameters, c and d, and so the basic SDT model is *exactly identified*. This means that one can directly solve for each parameter. For example, it follows from Eq. (1) that the probability of a hit is

$$p(Y = 1|X = 1) = \Phi(-c + d),$$

and the probability of a false alarm is

$$p(Y = 1|X = 0) = \Phi(-c).$$

It follows that

$$d = \Phi^{-1}[p(Y = 1|X = 1)] - \Phi^{-1}[p(Y = 1|X = 0)],$$

where, for normal distributions, d is the traditional distance measure d' (d is more general notation, for distributions other than the normal, see DeCarlo, 1998). The above shows that one can obtain an estimate of d simply by taking inverse normal transforms, Φ^{-1}, of the proportion of hits and the proportion of false alarms and subtracting. Similarly, solving for the criterion gives

$$c = -\Phi^{-1}[p(Y = 1|X = 0)].$$

Using the above equations along with the observed response proportions (which provide estimates of the probabilities) gives estimates of c and d. Of course, a better approach is to fit Eq. (1), because estimates of the standard errors of the parameter estimates are then also obtained.

2. SDT and 2AFC with bias

In 2AFC, two pieces of information are available to the observer on each trial, namely the perception of the event in the first position and the perception of the event in the second position, where "position" refers to either spatial position (left or right) or temporal position (first or second). A response of $Y = 1$ indicates that the first position is chosen as the signal, and $Y = 2$ indicates that the second position is chosen as the signal. Let the perception associated with the first position be denoted by Ψ_1 and the perception associated with the second position by Ψ_2. The decision rule, assuming no bias, is

$$Y = 1 \quad \text{if } \Psi_1 > \Psi_2,$$
$$Y = 2 \quad \text{if } \Psi_1 \leq \Psi_2.$$

The view in SDT is that the observer has information about their perceptual magnitudes, and so, as shown by the above decision rule, they simply choose as the signal the position with the largest perception. The right panel of Fig. 1 illustrates the decision rule for 2AFC. The figure shows the situation where the signal is presented in the first position. The solid dot shows, for a given trial, the location of a perception associated with Position 1 (signal) whereas the open dot shows a perception associated with Position 2 (noise). The observer knows that the solid observation is above the open observation, just as they know that the observation is above the criterion in the yes/no situation, and so they choose Position 1 as the signal, which in this case is a correct decision.

As has long been recognized (e.g., Luce, 1963; also see Macmillan & Creelman, 2005; Wickens, 2002), the choice might also be affected by response bias. This can be modeled by using one position as the reference and allowing bias to be associated with the other position (i.e., the bias is relative to the reference position). The approach here is to use the last position as the reference and to then use bias parameters for the other positions (one could also use bias parameters for all positions with a sum to zero constraint). Thus, the *decision rule* for 2AFC with bias is

$$\begin{aligned} &Y = 1 \quad \text{if } \Psi_1 + b > \Psi_2 \\ &Y = 2 \quad \text{if } \Psi_1 + b \leq \Psi_2, \end{aligned} \tag{2}$$

where b is a bias parameter, with positive values indicating bias for the first position over the second position, and negative values indicating bias for the second position.

It follows from the decision rule of Eq. (2) that

$$p(Y = 1|S = 1) = p(\Psi_2 < \Psi_1 + b),$$

where $p(Y = 1|S = 1)$ is the probability that the observer chooses Position 1 given that the signal is in Position 1 (a hit). The usual next step (e.g., Macmillan & Creelman, 2005; Wickens, 2002) is to take differences in order to derive the model (see Appendix A). Observers in SDT, however, are assumed to have information about their perceptual magnitudes, and so they know which realization is larger, as illustrated in Fig. 1, without having to take differences. In other words, the view here is that observers can directly evaluate $\Psi_2 < \Psi_1$ without having to subtract Ψ_1 and Ψ_2 (i.e., the process is not one of differencing). This distinction does not make any difference for 2AFC, in that the resulting models are identical, however the two approaches can lead to different models for $m > 2$.

The SDT model can be derived directly from the above as follows. First, note that a *structural model* for 2AFC is

$$\Psi_i = dX_i + \varepsilon_i,$$

for $i = 1, 2$, with $X_1 = 1$ for signal in Position 1 and 0 otherwise, and $X_2 = 1$ for signal in Position 2 and 0 otherwise. As before, the structural model shows how the perception is related to the presentation of a signal or noise (i.e., the conditional mean of the perceptual distribution is shifted by d). The decision rule and structural model together give, for a signal in Position 1,

$$\begin{aligned} p(Y = 1|X_1 = 1, X_2 = 0) &= p(\Psi_2 < \Psi_1 + b) \\ &= p(\varepsilon_2 < b + d + \varepsilon_1). \end{aligned}$$

Note that, for a given realization (value) of ε_1, say e_1, the above conditional probability is simply,

$$\begin{aligned} p(Y = 1|X_1 = 1, X_2 = 0, \varepsilon_1 = e_1) &= p(\varepsilon_2 < b + d + e_1) \\ &= F(b + d + e_1) \end{aligned}$$

where F is a cumulative distribution function (CDF). The response probability, not conditional on e_1, can then be found by integration,

$$p(Y = 1|X_1 = 1, X_2 = 0) = \int_{-\infty}^{\infty} F(b + d + e_1)f(e_1)de_1$$

where f is a probability density function (PDF). Similarly, it can be shown that, for a signal in Position 2,

$$p(Y = 1|X_1 = 0, X_2 = 1) = \int_{-\infty}^{\infty} F(b - d + e_1)f(e_1)de_1.$$

The above two equations together give the signal detection model for 2AFC with bias,

$$p(Y = 1|Z) = \int_{-\infty}^{\infty} F(b + dZ + e_1)f(e_1)de_1,$$

where $Z = X_1 - X_2$, that is, $Z = 1$ indicates that the signal is at Position 1 and $Z = -1$ indicates that the signal is at Position 2.

The above is a general SDT model for 2AFC with bias. The normal theory version of the model follows by using the normal CDF (i.e., Φ) and the normal probability density function (i.e., ϕ) for F and f, respectively,

$$p(Y = 1|Z) = \int_{-\infty}^{\infty} \Phi(b + dZ + e_1)\phi(e_1)de_1. \tag{3}$$

Eq. (3) is the normal theory version of the SDT model for 2AFC with bias. It is important to note that the model was derived simply by assuming that, on any given trial, the observer chooses the alternative associated with the largest perceptual magnitude. The approach can also be immediately generalized to $m > 2$; examples for $m = 3$ and 4 are given below.

The 2AFC model of Eq. (3), and its generalizations to *m*AFC, can be fit directly, as shown below. However, Eq. (3) has generally not been used for 2AFC because a simplification is available (for 2AFC with normal distributions, but not in general for distributions other than the normal or for *m*AFC). In particular, it can be shown that

$$\int_{-\infty}^{\infty} \Phi(b + d + e_1)\phi(e_1)de_1 = \Phi\left(\frac{b + d}{\sqrt{2}}\right),$$

and

$$\int_{-\infty}^{\infty} \Phi(b - d + e_1)\phi(e_1)de_1 = \Phi\left(\frac{b - d}{\sqrt{2}}\right).$$

It follows from these relations that Eq. (3) can be re-written as,

$$p(Y = 1|Z) = \Phi\left(\frac{b + dZ}{\sqrt{2}}\right). \tag{4}$$

An advantage of Eq. (4) is that, as for the yes/no situation, it is a simple probit model, and so it can easily be fit with standard software (e.g., DeCarlo, 2003) using maximum likelihood estimation (see Myung, 2003). Note that, once again, there are two observations (i.e., hits and false alarms) and two parameters, b and d (i.e., the model is exactly identified), and so one can solve directly for the parameters in terms of (inverse normal) transformed hits and false alarms, as done above.

Note that Eq. (4) also follows from a derivation in terms of differences, which is the usual textbook approach in SDT (e.g., Macmillan & Creelman, 2005; Wickens, 2002); this is shown in Appendix A. There is also a large closely related literature on "Thurstonian" modeling, Luce's choice theory, random utility models (e.g., see Bock & Jones, 1968; Böckenholt, 2006; Luce, 1994; Yellott, 1977), and ranking models (see Critchlow, Fligner, & Verducci, 1991). Thurstone (1927), for example, used differenced random variables in his presentation of the law of comparative judgment. Here it is simply noted that the view in SDT is that observers have information about their perceptual magnitudes, and so they simply select, as the signal, the alternative associated with the largest perceptual magnitude, and so the *process* need not involve differencing. The *m*AFC SDT models are developed here directly from this idea – choose the alternative with the largest perceptual magnitude – without the use of differencing.

2.1. Traditional approach: probability of a correct response

As noted above, a common approach to forced choice is to analyze the total proportion correct, and not the proportion of hits and false alarms separately (or, equivalently, the proportion correct for each position). In practical terms, this means that, instead of keeping track of the proportion of times the signal was correctly chosen when it was in Position 1 and the proportion of times the signal was correctly chosen when it was in Position 2, only the total proportion of times the signal was correctly chosen is kept track of, in which case information about possible position bias is lost.

The implications of the above can be clarified by explicitly deriving the model for the probability of a correct response (i.e., the total proportion correct provides an estimate of the probability of a correct response). Let $Z = 1$ indicate that the signal is presented at Position 1 and $Z = -1$ indicate that the signal is at Position 2. The probability of a correct response, p_C, is then

$$\begin{aligned} p_C &= p(Y = 1, Z = 1) + p(Y = 2, Z = -1) \\ &= p(Z = 1)p(Y = 1|Z = 1) \\ &\quad + p(Z = -1)[1 - p(Y = 1|Z = -1)]. \end{aligned}$$

That is, a correct response occurs if the observer chooses Position 1 when the signal is in Position 1 or the observer chooses Position 2 when the signal is in Position 2. Note that the conditional probabilities in the above are given by the SDT model of Eq. (4), and so

$$p_C = p\Phi\left(\frac{b+d}{\sqrt{2}}\right) + (1-p)\Phi\left(\frac{-b+d}{\sqrt{2}}\right),$$

where p is shorthand for $p(Z = 1)$. Note that the model has two parameters, b and d, but only one observation is available (the proportion correct, which provides an estimate of p_C). Thus, there are fewer parameters than observations and the model is under-identified. However, if it is assumed that there is no bias ($b = 0$), then the above reduces to

$$p_C = \Phi\left(\frac{d}{\sqrt{2}}\right),$$

as, for example, given by Eq. (7.6) of Macmillan and Creelman (2005). The above model is exactly identified, and so one can use the observed proportion correct (an estimate of p_C) to obtain an estimate of d. This is basically the motivation for assuming zero bias in 2AFC. Note that, by using the observed proportion correct as the outcome, one has lost information about bias (because information about position is ignored). On the other hand, keeping track of the position chosen allows one to use Eq. (3) or (4) and does not require any assumptions about bias. As shown below, this is also the case for models with $m > 2$.

3. SDT and *m*-alternative forced choice

The generalization of the SDT approach to m-alternatives follows immediately from the above ideas. As before, observers are assumed to have information about their perceptual magnitudes, and so once again the observer simply selects, on any given trial, the alternative associated with the largest perceptual magnitude. Thus, a signal detection approach to mAFC does not require any new assumptions, and the decision rule is exactly the same as for 2AFC, simply extended to more alternatives. The model is shown next for three and four alternatives; the approach is immediately applicable to m-alternatives.

3.1. The decision rule for 3AFC

The SDT approach assumes that the observer chooses, as the signal, the alternative associated with the largest perceptual magnitude, exactly as for 2AFC. Thus, the SDT approach can easily be applied to forced choice with any number of alternatives. For example, the decision rule for 3AFC, without bias, can be compactly written as,

$$Y = i \quad \text{if } \Psi_i > \max(\Psi_j, \Psi_k),$$

for $i = 1, 2,$ or 3 (for Position 1, 2, or 3), where max is the maximum of the set of values. The above decision rule is easily generalized to allow for response bias, as done below. First examined, however, is the traditional approach, where the probability of a correct response is used as the outcome variable and it is assumed that there is no bias.

3.2. Traditional approach: probability of a correct response and 3AFC

For 3AFC, a correct response can be made in one of three ways: when the signal is in the first position and the first position is chosen, the signal is in the second position and the second position is chosen, or the signal is in the third position and the third position is chosen. The traditional approach, however, ignores this and simply considers the total number of correct responses (and so once again information is lost).

In the traditional approach, the signal is chosen (ignoring the actual position chosen) if $\Psi_S > \max(\Psi_{N1}, \Psi_{N2})$, where Ψ_S is the perceptual distribution associated with signal and Ψ_{N1} and Ψ_{N2} are perceptual distributions associated with noise (for two positions). Note that it is assumed that there is no response bias. It follows that the probability of a correct response is

$$p_C = p(Y = S) = p[\max(\Psi_{N1}, \Psi_{N2}) < \Psi_S],$$

where $p(Y = S)$ is the probability that the observer chooses the signal (i.e., a correct response). The above requires finding the maximum of a signal in noise, which is a well-known problem in communication systems, as was recognized early on by Tanner and Swets (1954).

Note that

$$p[\max(\Psi_{N1}, \Psi_{N2}) < \Psi_S] = p(\Psi_{N1} < \Psi_S \cap \Psi_{N2} < \Psi_S).$$

The above events are not independent, given the dependence of both on the random variable Ψ_S, however they are independent for a given realization of $\Psi_S = \psi_S$, and so

$$\begin{aligned} p(Y = S|\Psi_S = \psi_S) &= p(\Psi_{N1} < \psi_S \cap \Psi_{N2} < \psi_S) \\ &= p(\Psi_{N1} < \psi_S)\, p(\Psi_{N2} < \psi_S), \\ &= F_{N1}(\psi_S) \times F_{N2}(\psi_S) = [F_N(\psi_S)]^2, \end{aligned}$$

where $F_{N1}(\psi_S)$ and $F_{N2}(\psi_S)$ are cumulative distribution functions for the two noise distributions (which are both the same, and so the subscripts $N1$ and $N2$ can be replaced by N). Note that the approach can easily be generalized to m-alternatives, which gives

$$p(Y = S|\Psi_S = \psi_S) = [F_N(\psi_S)]^{m-1}.$$

To get the unconditional probability of a correct response, one can integrate over ψ_S,

$$p(Y = S) = \int_{-\infty}^{\infty} [F_N(\psi_S)]^{m-1} f_S(\psi_S - d)\, d\psi_S$$

where f_S is a probability density function for the signal perception, and d is the usual distance measure (not to be confused with $d\psi_S$, which is the derivative with respect to ψ_S). If one assumes normal distributions, so that $\Psi_N \sim N(0, 1)$ and $\Psi_S \sim N(d, 1)$, then the above is

$$p_C = \int_{-\infty}^{\infty} [\Phi(\psi_S)]^{m-1} \phi_S(\psi_S - d)\, d\psi_S, \tag{5}$$

where ϕ_S is the normal probability density function for signal. Eq. (5) can be found in one form or another in many references (e.g., Green & Dai, 1991; Hacker & Ratcliff, 1979; Macmillan & Creelman, 2005; Swets, Tanner, & Birdsall, 1961; Tanner & Swets, 1954; Wickens, 2002). Hacker and Ratcliff (1979) provide tables of the proportion correct and d for the normal model for a number of m-alternative situations (also see Frijters, Kooistra, & Vereijken, 1980). Using the approach shown below, however, one can directly fit Eq. (5) and obtain estimates of d and its standard error.

As for 2AFC, the "standard" model of Eq. (5) is quite wasteful of the data, given that it reduces all of the observations to one—the total proportion correct. For example, for 3AFC, there are six proportions in the three by three table that are free to vary. This is shown in Table 1, which presents data that are analyzed below.

Table 1
Response frequencies for two conditions of a 3AFC experiment (Ennis & O'Mahony, 1995).

Prior stimuli	Sequence	"1"	"2"	"3"	Total
WW	SWW	54	5	1	60
WW	WSW	0	60	0	60
WW	WWS	5	3	52	60
SS	SWW	40	12	8	60
SS	WSW	6	49	4	59
SS	WWS	6	5	49	60

Notes: WW = water, water prior stimuli; SS = salt water, salt water prior stimuli; SWW = salt-water–water sequence, and similarly for the other sequences, responses of "1", "2", and "3" indicate the frequency with which the 1st, 2nd, or 3rd stimulus was chosen as *S*.

Consider the first three by three table at the top of Table 1. The three rows of the table are for when the signal was presented in the first, second, or third position, whereas the three columns are for a choice by the observer of the first, second, or third position. Note that the row totals are fixed by design (i.e., they depend on how many times the signal was presented in each position, which is 60 in this case), and so only two of the three values in each row are free to vary (because the sum must equal 60). Thus, two free values per row by three rows gives a total of $2 \times 3 = 6$ observations that are free to vary, and so there are six pieces of information available. Proportion correct, on the other hand, is simply the sum of the main diagonal (divided by the total), and so six pieces of information are reduced to one. In contrast, the SDT model with bias developed here uses all of the information available in the data.

3.3. SDT and 3AFC with bias

In 3AFC, the observer has, on each trial, three perceptions, each associated with one of the three positions. Response bias can be introduced by including bias parameters (b_i) in the decision rule, exactly as done for 2AFC above. The bias again reflects the added or subtracted value of the position, relative to a reference position. For three positions, there are two bias parameters, denoted here as b_1 and b_2 for Positions 1 and 2, respectively, with the third position serving as the reference. The decision rule for 3AFC with bias is,

$$\begin{aligned} Y &= 1 \quad \text{if } \Psi_1 + b_1 > \max(\Psi_2 + b_2, \Psi_3), \\ Y &= 2 \quad \text{if } \Psi_2 + b_2 > \max(\Psi_1 + b_1, \Psi_3), \\ Y &= 3 \quad \text{if } \Psi_3 > \max(\Psi_1 + b_1, \Psi_2 + b_2). \end{aligned} \tag{6}$$

That is, Position 1 is chosen ($Y = 1$) if the associated perception (Ψ_1) is greater than the perceptions for Positions 2 and 3, including effects of bias (b_1 and b_2), and similarly for the other responses. Note that the decision rule is the same as for 2AFC: choose the alternative with the largest perceptual magnitude, including effects of bias.

The structural model is

$$\begin{aligned} \Psi_1 &= dX_1 + \varepsilon_1, \\ \Psi_2 &= dX_2 + \varepsilon_2, \\ \Psi_3 &= d(1 - X_1 - X_2) + \varepsilon_3, \end{aligned} \tag{7}$$

where $X_1 = 1$ for signal in Position 1 and zero otherwise, and $X_2 = 1$ for signal in Position 2 and zero otherwise. The decision rule and structural model together give the SDT model for 3AFC, in the same manner as for 2AFC above. The result is six equations that correspond to the six cells of the three by three table that are free to vary, as discussed above.

More specifically, it follows from the decision rule that the first position is chosen if

$$\begin{aligned} p(Y = 1|X_1, X_2) &= p[\max(\Psi_2 + b_2, \Psi_3) < \Psi_1 + b_1] \\ &= p[(\Psi_2 + b_2 < \Psi_1 + b_1) \cap (\Psi_3 < \Psi_1 + b_1)]. \end{aligned}$$

One can develop the model by conditioning on Ψ (as done above for the traditional approach), however it is more informative to use the structural model and condition on ε, as done for 2AFC above. In particular, substituting the structural model given above and rearranging terms gives

$$\begin{aligned} p(Y = 1|X_1, X_2) = p[&(\varepsilon_2 < b_1 - b_2 + dX_1 - dX_2 + \varepsilon_1) \\ &\cap (\varepsilon_3 < b_1 + dX_1 - d(1 - X_1 - X_2) + \varepsilon_1)]. \end{aligned}$$

The above events are independent conditional on a given value of $\varepsilon_1 = e_1$,

$$\begin{aligned} &p(Y = 1|X_1, X_2, \varepsilon_1 = e_1) \\ &\quad = p(\varepsilon_2 < b_1 - b_2 + dX_1 - dX_2 + e_1)p(\varepsilon_3 < b_1 \\ &\qquad + dX_1 - d(1 - X_1 - X_2) + e_1) \\ &\quad = F(b_1 - b_2 + dX_1 - dX_2 + e_1) \\ &\qquad \times F[b_1 + dX_1 - d(1 - X_1 - X_2) + e_1]. \end{aligned}$$

The unconditional probability can then be found, as above, by integrating,

$$\begin{aligned} p(Y = 1|X_1, X_2) = \int_{-\infty}^{\infty} &F(b_1 - b_2 + dX_1 - dX_2 + e_1) \\ &\times F[b_1 + dX_1 - d(1 - X_1 - X_2) + e_1]f(e_1)de_1. \end{aligned}$$

The above gives three equations, for the probabilities of choosing Position 1 when the signal is in Position 1, 2, or 3 (indicated by the values of X_1 and X_2). A second set of three equations can be derived in a similar manner for the choice of Position 2, which gives

$$\begin{aligned} p(Y = 2|X_1, X_2) = \int_{-\infty}^{\infty} &F(-b_1 + b_2 - dX_1 + dX_2 + e_2) \\ &\times F[b_2 + dX_2 - d(1 - X_1 - X_2) + e_2]f(e_2)de_2. \end{aligned}$$

The model (and programs, see Appendix B) can be simplified by letting $Z = X_1 - X_2$ (as for 2AFC), $Z_1 = 1 - 2X_1 - X_2$, and $Z_2 = 1 - X_1 - 2X_2$, which gives

$$\begin{aligned} p(Y = 1|\mathbf{Z}) &= \int_{-\infty}^{\infty} F(b_1 - b_2 + dZ + e_1)F(b_1 - dZ_1 + e_1) \\ &\qquad \times f(e_1)de_1 \\ p(Y = 2|\mathbf{Z}) &= \int_{-\infty}^{\infty} F(-b_1 + b_2 - dZ + e_2)F(b_2 - dZ_2 + e_2) \\ &\qquad \times f(e_2)de_2. \end{aligned} \tag{8}$$

Eq. (8) is a general SDT model for 3AFC with bias. Using Φ for F and ϕ for f gives the normal theory version of the model. The approach can also be immediately applied to *m*AFC with $m > 3$; as an example, the model for 4AFC is presented below. Eq. (3) for 2AFC and Eq. (8) for 3AFC show that including bias in the model for *m*AFC results in the inclusion of bias parameters in the psychometric function.

Note that the two components of Eq. (8) are conditioned on different random variables, ε_1 and ε_2, and so the model in essence consists of two binomial models (c.f. Begg & Gray, 1984). In practical terms, this means that, in order to fit the model, responses Y of 1, 2, and 3 (indicating the position chosen) must be recoded into two dichotomous responses, say Y_1 and Y_2, where $Y_1 = 1$ indicates that Position 1 was chosen and $Y_1 = 0$ indicates that Position 2 or 3 was chosen; $Y_2 = 1$ indicates that Position 2 was chosen and $Y_2 = 0$ indicates that Position 1 or 3 was chosen. The two sets of binary models, that is for responses of $Y_1 = 1$ (i.e., $Y = 1$ in the first three equations of Eq. (8)) and $Y_2 = 1$ (i.e., $Y = 2$ in the second three equations), can then be simultaneously fit, as shown by the programs provided in Appendix B.

3.4. Fitting the *m*AFC SDT model with bias

As has long been recognized, fitting the SDT version of the *m*AFC model with bias presents a bit of a challenge. In contrast to Eqs. (1) and (4), for example, Eq. (8) has a more complex form and involves an integral with limits from minus to plus infinity. There are, however, several approaches available for fitting the model. One involves the use of maximum likelihood estimation (MLE) and Gaussian quadrature, whereas the second involves the use of a Bayesian approach with Markov chain Monte Carlo. It is beyond the scope of this article to introduce and explain these approaches in any detail, and so it is simply shown how to use the approaches to fit the *m*AFC SDT models presented here, with references to articles that discuss details of the methods, and programs provided in an Appendix.

A first step is to recognize that the *m*AFC SDT model can be viewed as a type of *nonlinear mixed model* (NLMM; Davidian & Giltinan, 1995; Lindstrom & Bates, 1990; Vonesh & Chinchilli, 1996), and so it can be fit using methods developed for NLMMs. Note that this places *m*-alternative forced choice SDT models within a well developed statistical framework. The use of a special case of NLMMs, namely *generalized linear mixed models* (GLMMs; see Breslow & Clayton, 1993; Fahrmeir & Tutz, 2001; McCulloch & Searle, 2001), for SDT has earlier been noted (DeCarlo, 1998).

The model can be expressed as follows. First note that the joint probability of the dichotomous responses Y_1 and Y_2 defined above is given by the product of their separate probabilities, because of independence, which follows from the assumption of independence of ε_1 and ε_2. Let the responses Y_1 and Y_2 be denoted by Y_j for $j = 1$ and 2. The dichotomous response variables are assumed to follow Bernoulli distributions with parameters p_j,

$$Y_j \sim \text{Bernoulli}(p_j). \tag{9}$$

Next, the Bernoulli parameters (p_j) are given by the 3AFC model of Eq. (8), which can be written more compactly as,

$$p_j = \int_{-\infty}^{\infty} \Phi(\eta_{1j})\Phi(\eta_{2j})\phi(e_j)de_j, \tag{10}$$

where η_{1j} and η_{2j} are *linear predictors* for Y_1 and Y_2. Note that there are different linear predictors for Y_1 and Y_2, which is why there are four predictors in the next equation. In particular, as shown by Eq. (8), the linear predictors for the 3AFC SDT model are

$$\begin{aligned}
\eta_{11} &= b_1 - b_2 + dZ + e_1 \\
\eta_{12} &= -b_1 + b_2 - dZ + e_2 \\
\eta_{21} &= b_1 - dZ_1 + e_1 \\
\eta_{22} &= b_2 - dZ_2 + e_2.
\end{aligned} \tag{11}$$

Note that the linear predictors include random effects (e_j), as is the usual case for NLMMs and GLMMs.

The 3AFC SDT model with normal distributions is specified by the above. Eqs. (9) through (11) basically show how to implement the model for both the maximum likelihood and Bayesian approaches discussed below (compare the equations to the syntax given in Appendix B). A final detail has to do with how to deal with the intractable integral in Eq. (10). Two approaches are noted here: maximum likelihood with Gaussian quadrature and a Bayesian approach. The first approach is implemented with a SAS program given below; the second approach is implemented with the Bayesian software OpenBUGS (Thomas, O'Hara, Ligges, & Sturtz, 2006).

3.5. MLE and Gaussian quadrature

Gauss–Hermite quadrature is used to approximate an integral of a function that is multiplied by another function, where the second function has the shape of a normal density, that is,

$$\int_{-\infty}^{\infty} f(e_j)\exp(-e_j^2)de_j,$$

where notation relevant to the models discussed here is used. The basic idea of Gauss–Hermite quadrature is that the above can be approximated by a finite weighted sum,

$$\int_{-\infty}^{\infty} f(e_j)\exp(-e_j^2)de_j \approx \sum_{q=1}^{Q} f(e_{jq})w_q,$$

where e_{jq} are *quadrature nodes*, w_q are *weights*, and Q is the number of nodes. For example, using the above, the probability of Eq. (10) can be approximated as

$$p_j \approx \sum_{q=1}^{Q} \Phi(\eta_{1jq})\Phi(\eta_{2jq})w_q,$$

with

$$\begin{aligned}
\eta_{11q} &= b_1 - b_2 + dZ + e_{1q} \\
\eta_{12q} &= -b_1 + b_2 - dZ + e_{2q} \\
\eta_{21q} &= b_1 - dZ_1 + e_{1q} \\
\eta_{22q} &= b_2 - dZ_2 + e_{2q},
\end{aligned}$$

where the Gaussian nodes (e_{jq}) and weights (w_q) are used in place of e_j and $\phi(e_j)$, respectively. Thus, this approach replaces the intractable integral of Eq. (10) with a numerical approximation, and the resulting approximated likelihood can then be maximized with standard algorithms, giving maximum likelihood estimates of the parameters. This approach is implemented in the NLMIXED procedure of SAS, except that NLMIXED uses *adaptive* Gaussian quadrature, as discussed by Pinheiro and Bates (1995); (also see Liu & Pierce, 1994). Lesaffre and Spiessens (2001) cautioned against using too few quadrature points and noted that the use of adaptive quadrature, as in NLMIXED, solved the problem for an example they presented.

With respect to the SAS program given in Appendix B, the first part of the program simply recodes the variables and gets the data into the form needed for the NLMIXED procedure. Note that the model is fit as a *multivariate* nonlinear mixed model. Also note that indicator variables ($i1$ and $i2$), which indicate whether the response is Y_1 or Y_2, are created. The program includes a "trick", which is that the indicators are used to select the appropriate linear predictors for Y_1 and Y_2. The linear predictors are denoted as eta1 and eta2 in the program, and are transformed by the normal CDF (phi in the program) and multiplied, as in Eq. (10). To approximate Eq. (10), as discussed above, adaptive Gaussian quadrature with 20 nodes is used (using more nodes did not appear to change the results). The program also shows that the responses Y_j are specified as Bernoulli variables (with the 'binary' command) and that the ε_j (denoted as eps1 and eps2 in the program) are specified as being random variables with variances of one and a covariance of zero.

Note that a similar approach has been used for item response theory (IRT) models, which can also be expressed as NLMMs and are closely related to the models presented here. The approach is usually discussed in IRT as *marginal maximum likelihood* (e.g., de Ayala, 2009) because, as shown in Eq. (8), the random effect (ε_j in this case) is integrated out of the conditional response probability (and so the 'marginal' likelihood is maximized). Some useful references that discuss relevant details of NLMIXED (and other programs) in this context are Rijmen, Tuerlinckx, De Boeck, and Kuppens (2003) and Sheu, Chen, Su, and Wang (2005).

3.6. Bayesian estimation and MCMC

For the Bayesian approach, the statistical model is again the same as in Eqs. (9) through (11), however priors are now specified for the model parameters d, b_1, and b_2. The priors are used, along with the likelihood (which follows from the model), to get posteriors,

posterior $\propto$ prior $\times$ likelihood.

Markov chain Monte Carlo (MCMC) methods can be used to sample from the posterior distributions of the parameters given the observed data.

The OpenBUGS program in Appendix B shows that the model is again specified exactly as given in Eqs. (9) through (11), with Y_j having Bernoulli distributions (using the 'dcat' command) and ε_j (eps1 and eps2) having independent $N(0, 1)$ distributions. Note that the OpenBUGS syntax uses the *precision*, which is the inverse of the variance, $1/\sigma^2$, instead of the variance, and so 0.1 in the program indicates a variance of 10. Priors are specified for the parameters as follows,

$$d \sim N(0, \surd 10)$$
$$b_1 \sim N(0, \surd 10)$$
$$b_2 \sim N(0, \surd 10),$$

which can be viewed as being "mildly" informative (in practice, one usually has knowledge about typical estimates obtained in SDT studies). Indeed, an advantage of the Bayesian approach over maximum likelihood is that the Bayesian approach allows one to incorporate information from previous research. For the real-world example analyzed below, the reader can perform a sensitivity analysis and verify that the results are unchanged if non-informative priors (e.g., normal with a variance of 100) are instead used. As noted in the program in Appendix B, one can also use the bounds function of OpenBUGS to restrict d to positive values.

The Bayesian approach has also been used for IRT models, with a useful reference being Patz and Junker (1999), who provide details about Metropolis–Hastings sampling, which is used for the forced choice signal detection models discussed here (i.e., it is used in OpenBUGS); also see Albert (1992). Related work for Bayesian Thurstonian models is discussed by Ansari and Iyengar (2006) and Yao and Böckenholt (1999). Some useful textbooks on Bayesian analysis are Congdon (2005), Gelman, Carlin, Stern, and Rubin (2004), Jackman (2009), and Lynch (2007).

It should be noted that although the maximum likelihood approach with Gaussian quadrature and the Bayesian approach with Markov chain Monte Carlo come from conceptually different backgrounds, they are actually similar (apart from the introduction of priors for the model parameters in the Bayesian approach). For example, with respect to approximating the integral, the nodes and weights in Gaussian quadrature are fixed, whereas they can be viewed in the Monte Carlo approach as being random, as was noted by Pinheiro and Bates (1995).

3.7. Simulations: 3AFC with bias

Simulations were conducted in order to obtain information about parameter recovery for 3AFC with bias using both approaches, MLE and Bayesian. A "small" sample size was used (recovery appeared to be quite good for large sample sizes). For example, it is typical in applied research (e.g., in memory research) to obtain 100–200 observations per observer in a single session; sample sizes of this sort are also found in other applications, such as the food science example discussed below (sample size of 180). A sample size of 150 was used for the simulation, with the signal appearing in each position 50 times. The population parameters used were similar to those found for analyses of some real-world data, with $d = 1.5$, $b_1 = 0.6$ (small bias for the first position over the third) and $b_2 = -0.4$, and so the first position is the most preferred, the third position is the next preferred, and the second position is the least preferred (but the bias is 'small' in both cases). SAS was used to generate 50 datasets.

For the maximum likelihood approach, SAS was run for each dataset with 20 quadrature points, whereas for the Bayesian approach, OpenBUGS was run for each dataset, with 5000 burn-ins and 20,000 iterations (per dataset). The use of 20,000 iterations appeared to be adequate for convergence, in that the Monte Carlo errors, which reflect between-simulation variability (see Flegal, Haran, & Jones, 2008; Geyer, 1992; Koehler, Brown, & Haneuse, 2009) were less than 5% of the posterior standard deviations (the Monte Carlo error and posterior standard deviations are both given in the OpenBUGS output). This has been suggested as a criterion for convergence (e.g., see Spiegelhalter, Thomas, Best, & Lunn, 2003); inspection of multiple chains also suggested convergence.

Table 2 presents the results. For both simulations, the table shows the average parameter estimates over 50 replications, the average standard error (or posterior standard deviation), the percent coverage, and the percent of cases that were significant at the 0.05 level. For MLE, the percent coverage is for 95% confidence intervals, which have an interpretation in terms of containing the population value in repeated sampling. In the Bayesian approach, percent coverage is for 95% credible intervals, which have an interpretation in terms of the probability that the parameter is in the interval. Similarly, the percent significant is the percent of cases where zero was not in the 95% confidence interval, for MLE, or not in the 95% credible interval, for the Bayesian approach; both provide information about the power to detect a nonzero effect of a given size. The top part of Table 2 shows results for MLE with NLMIXED of SAS; the lower part shows results for Bayesian estimation with OpenBUGS. The table shows that the results for the two approaches are virtually identical, and so they will be discussed together.

Table 2 shows that the average estimates of d, b_1, and b_2 are quite close to their population values (with generally less than 10% error), and so the results suggest that parameter recovery for both approaches is quite good, even with a small sample size. The percent coverage obtained for b_1 and b_2 (94%) is close to the nominal value of 95%, whereas that for d is slightly lower (90%), but still quite high. The '% significant' column shows that d was detected as being non-zero in 100% of the cases. Of particular interest are the significance results for the bias parameters, because in practice one wants to know if there is non-zero bias. Table 2 shows that there is adequate power (84% for MLE and 92% for Bayesian) to detect the larger value of bias, $b_1 = 0.6$, however the power for the smaller negative bias, $b_2 = -0.4$, is clearly inadequate at 36%–38%. The results provide useful information about the magnitude of bias that can be detected for 3AFC with a sample size of 150 (in a balanced design).

To summarize, the results suggest that parameter recovery for 3AFC is good for both approaches (MLE and Bayesian), even with a fairly small sample size. Of course, this conclusion is limited to the situation examined here (which is similar to the real-world example analyzed below). Given that it is quite simple to generate data and fit the model, the recommendation here is to conduct at least a small simulation in order to obtain some information about parameter recovery and power in the particular situation being studied.

3.8. A real-world example: 3AFC with bias

An extensive search for 3AFC data found that in virtually all cases where researchers provided their data, only the proportion of correct responses was given (ignoring position), and not the full three by three table. An exception was a study on taste that

Table 2
Results for maximum likelihood and Bayesian estimation for a 3AFC simulation, $N = 150$.

Maximum likelihood with NLMIXED of SAS					
Parameter	Pop. value	Average estimate	Av. SE	% Coverage	% Significant
d	1.50	1.54	0.16	90	100
b_1	0.60	0.63	0.23	94	84
b_2	−0.40	−0.41	0.26	94	38
Bayesian estimation with OpenBUGS					
Parameter	Pop. value	Average estimate	Av. PSD	% Coverage	% Significant
d	1.50	1.56	0.16	90	100
b_1	0.60	0.65	0.24	94	92
b_2	−0.40	−0.40	0.26	94	36

Notes: Results are for 50 datasets. SE is the standard error. PSD is the posterior standard deviation.

Table 3
Parameter estimates for two conditions of a 3AFC experiment (Ennis & O'Mahony, 1995).

	d	b_1	b_2
WW prior stimuli			
Table (H&R, 1979)	2.39	–	–
Eq. (5) (MLE)	2.41 (0.18)	–	–
Eq. (6) (MLE)	2.68 (0.23)	0.35 (0.31)	0.96 (0.36)
Eq. (6) (Bayesian)	2.73 (0.24)	0.36 (0.31)	1.00 (0.37)
SS prior stimuli			
Table (H&R, 1979)	1.52	–	–
Eq. (5) (MLE)	1.52 (0.14)	–	–
Eq. (6) (MLE)	1.47 (0.12)	−0.29 (0.20)	0.04 (0.20)
Eq. (6) (Bayesian)	1.47 (0.12)	−0.28 (0.20)	0.04 (0.21)

Notes: WW is water, water prior tasted stimuli; SS is salt-water, salt-water prior stimuli. H&R refers to the table given in Hacker and Ratcliff (1979), using the proportion correct as the outcome. MLE is maximum likelihood estimation, with standard errors shown in parenthesis. For Bayesian estimation, values are means and standard deviations (in parenthesis) of the posterior distributions.

considered sequence effects and response bias (Ennis & O'Mahony, 1995). For the data reported here, the three alternatives were (a) an alternative that consisted of salt added to water (signal) and (b) two alternatives that consisted of water alone (noise). The observer's task was to detect the salt solution. In addition, detection of salt was examined in two conditions, one where two samples of water were first tasted (before the 3AFC trial) and one where two salt-water samples were first tasted. Table 1 shows, as two three by three tables, the results (response frequencies) separately for the two conditions (labeled as WW or SS prior stimuli). The purpose here is simply to illustrate aspects of the analysis with some real-world data.

Table 3 shows results for a traditional analysis (i.e., without bias) as well as results for the 3AFC with bias model. For the WW condition (top of table), the first line shows an estimate of d obtained using the proportion correct and the tables of Hacker and Ratcliff (1979), which is 2.39. The second line shows an estimate of d obtained by fitting (with MLE) the traditional 3AFC model without bias (i.e., Eq. (5)). The table shows that the estimate of d is virtually the same as that found using Hacker and Ratcliff's table (about 2.4), with the advantage that the model-based approach also provides an estimate of the standard error. The third line of Table 3 shows estimates of d, b_1, and b_2, and their standard errors, obtained by fitting the 3AFC with bias model (Eq. (8)) with MLE. The fourth line shows the results for the Bayesian approach with OpenBUGS. Once again, the results for MLE and Bayesian estimation are virtually identical and so will be discussed together. For the 3AFC model with bias, the estimate of d (about 2.8) is larger than that obtained for the 3AFC model without bias (about 2.4). The results for the bias parameter estimates suggest a significant bias in favor of the second position, with an estimate of b_2 of 1.0 and standard error (or posterior standard deviation) of about 0.36.

The lower part of Table 3 shows results for the SS condition. In this case, all of the approaches give an estimate of d of around 1.5. Note that this is considerably smaller than the estimate of 2.8 found in the WW condition, which suggests that detection of salt-water was lower when two tastings of salt-water were used as prior stimuli. The bias parameter estimates are small, and so there is little or no bias for either Positions 1 or 2 (of course power considerations have to be kept in mind, as shown in the 4AFC simulation presented below; in any case, the parameter estimates suggest that the magnitude of the biases in this example, if any, are small).

Although of course the true values are not known for real-world data, the results are interesting in that they suggest that, when bias is present, as appears to be the case in the WW condition, d is underestimated if the bias is ignored (as suggested by the finding of smaller estimates of d for the model without bias than for the model with bias), whereas when bias is not present, as appears to be the case in the SS condition, d is *not* underestimated (as suggested by the finding that the estimates of d are about the same across models with and without bias). Macmillan and Creelman (2005) and Wickens (2002) have previously noted that ignoring bias in mAFC will lead to underestimation of d, and the results found here are consistent with this.

3.9. SDT and 4AFC with bias

As noted above, models for any number of alternatives follow immediately from the decision rule and structural model given above, with appropriate extensions. For example, for SDT, the decision rule is always the same: choose the alternative with the largest perceptual magnitude. The structural model is also simply extended to include additional covariates, to represent the additional possible positions that the stimulus can be presented in. Given that the 4AFC procedure has been used in a number of recent studies (whereas $m > 4$ is rarely used), the derivation of the model for 4AFC with bias is shown here, with parameter recovery again examined in simulations.

As before, the model can be derived using the decision rule and structural model. Note that the decision rule is a straightforward extension of Eq. (6) (choose the maximum) whereas the structural model is a straightforward extension of Eq. (7), with the use of additional dummy variables (e.g., X_1, X_2, and X_3) to indicate that the signal is in Position 1, 2, or 3, respectively. The SDT model is then derived in exactly the same manner as for 3AFC above. The result is twelve equations for the twelve cells of the four by four table that are free to vary. Presentation of the model is again simplified by recoding, and in particular let $Z = X_1 - X_2$, $Z_a = X_1 - X_3$, $Z_b = X_2 - X_3$, $Z_1 = 1 - 2X_1 - X_2 - X_3$, $Z_2 = 1 - X_1 - 2X_2 - X_3$, and $Z_3 = 1 - X_1 - X_2 - 2X_3$.

The SDT model for 4AFC is

$$p(Y = 1|\mathbf{Z}) = \int_{-\infty}^{\infty} F(b_1 - b_2 + dZ + e_1)F(b_1 - b_3 + dZ_a + e_1)F(b_1 - dZ_1 + e_1)f(e_1)de_1$$

Table 4
Results for maximum likelihood and Bayesian estimation for 4AFC simulations.

MLE: parameter estimates, $N = 160$					
Parameter	Pop. value	Average	Av. SE	% Coverage	% Significant
d	2.00	2.10	0.18	96	100
b_1	−0.60	−0.58	0.30	90	52
b_2	0.50	0.63	0.28	96	60
b_3	−1.00	−0.98	0.32	94	90
Bayesian estimation, $N = 160$					
Parameter	Pop. value	Average	Av. PSD	% Coverage	% Significant
d	2.00	2.12	0.19	92	100
b_1	−0.60	−0.55	0.32	90	46
b_2	0.50	0.68	0.30	90	64
b_3	−1.00	−0.95	0.33	92	86
MLE: parameter estimates, $N = 400$					
Parameter	Pop. value	Average	Av. SE	% Coverage	% Significant
d	2.00	2.04	0.11	96	100
b_1	−0.60	−0.62	0.18	98	94
b_2	0.50	0.50	0.17	96	82
b_3	−1.00	−1.05	0.19	94	100
Bayesian estimation, $N = 400$					
Parameter	Pop. value	Average	Av. PSD	% Coverage	% Significant
d	2.00	2.05	0.11	90	100
b_1	−0.60	−0.62	0.18	98	94
b_2	0.50	0.52	0.17	96	84
b_3	−1.00	−1.04	0.20	92	100

Notes: results are for 50 datasets. Av. SE is the average standard error. Av. PSD is the average posterior standard deviation.

$$p(Y=2|\mathbf{Z}) = \int_{-\infty}^{\infty} F(-b_1 + b_2 - dZ + e_2)F(b_2 - b_3 + dZ_b + e_2)F(b_2 - dZ_2 + e_2)f(e_2)de_2$$

$$p(Y=3|\mathbf{Z}) = \int_{-\infty}^{\infty} F(-b_1 + b_3 - dZ_a + e_3)F(-b_2 + b_3 - dZ_b + e_3)F(b_3 - dZ_3 + e_3)f(e_3)de_3, \quad (12)$$

where b_i is the bias for the ith position, with three bias parameters for the four positions (with the last position again serving as the reference). Eq. (12) gives the general SDT model for 4AFC with bias. The above shows exactly how to modify, for 4AFC, the SAS and OpenBUGS programs given in Appendix B. Note that, in this case, Y is recoded into three dichotomous Y_j to indicate the position chosen.

3.10. Simulations: 4AFC with bias

Simulations were again conducted in order to obtain some information about parameter recovery for 4AFC with bias. A 'small' sample size of 160 was used in one condition, and a 'medium' sample size of 400 was used in the other. In both cases, the signal appeared in each position either 40 times (for $N = 160$) or 100 times (for $N = 400$). The population parameters were $d = 2.0$, $b_1 = -0.6$, $b_2 = 0.5$, and $b_3 = -1.0$. SAS was used to generate 50 datasets; SAS and OpenBUGS were again used to fit the model.

The top part of Table 4 shows results for $N = 160$. Once again, the MLE and Bayesian results are virtually identical and so will be discussed together. Table 4 shows that the average estimates of d are close to the population value, and so estimation of d appears to be good even for a fairly small sample size. With respect to the bias parameters, estimates of b_1 and b_3 are close to their population values, whereas estimates of the smallest bias parameter, $b_2 = 0.5$, are too large (by about 36%). For all of the parameters, the percent coverage (90%–92%) is close to, but slightly lower than, the nominal 95% level. With respect to significance, power is adequate (86%) for $b = -1$, but is low for the other two (smaller) bias parameters. This shows that, for 4AFC with a sample size of 160, estimation and tests are generally good, but can be poor for small values of bias.

The lower half of Table 4 shows results for the 4AFC simulation with $N = 400$. Again, the results for MLE and Bayesian estimation are quite similar. In this case, the average estimates are quite close to their population values in all cases, and so estimation of both detection and bias appears to be quite good for 4AFC with a sample size of 400. Coverage also appears to be adequate, and ranges around the nominal 95% value (from 90% to 98%). The significance results show that power is now adequate in all cases, even for the smallest bias, $b_2 = 0.5$ (power of 84%). This shows that increasing the sample size from 160 to 400 in 4AFC gives much higher power to detect "small" biases (i.e., a magnitude of around 0.5). Again, the simulations are informative about the magnitude of bias that can be detected in practice for a given sample size and design. The take home message to researchers is that if one is concerned about 'small' bias, then an adequate sample size is needed (another option is to use a hierarchical model, as discussed below).

3.11. Bias and other factors

Bias, in the typical discussion of forced choice, usually refers to a non-perceptual factor that affects the observer's choice, such as preferring the right-most alternative in a forced choice task, which is a 'position' bias (see Macmillan & Creelman, 2005; Wickens, 2002). Here it is noted that, in some applications, the 'bias' might reflect the influence of other factors besides position. An interesting example is provided by two (other) conditions in the study of Ennis and O'Mahony (1995), where the task was, in 3AFC, to detect the weakest stimulus instead of the strongest (i.e., there were two S's and one W instead of two W's and one S). A fit of Eq. (8) (keeping in mind a direction reversal) gave bias parameters that indicated a bias in both conditions towards choosing the third stimulus as the weakest. This could reflect a sensory effect, such as adaptation, rather than a position preference. In short, it should be kept in mind that the 'bias' parameters obtained using the models presented here might reflect effects of other factors besides position. These factors should either be controlled for or explicitly brought into the model as covariates, as done here for position. The use of additional covariates in mAFC is an interesting area for future research.

4. Discussion

As noted by Macmillan and Creelman (2005), "As we have seen, bias is customarily ignored in analyzing mAFC data. That it does not therefore go away is shown in some 4AFC experiments of Nisbett and Wilson (1977)"; they then cite an example where a large bias for the right-most item was apparently present. This motivates the use of mAFC SDT models with bias, as presented here.

Up to this point, there has been a practical reason for ignoring bias—the resulting SDT models are relatively complex, as Green and Swets (1988) and Luce (1963) noted many years ago. The models are also certainly not trivial to fit. However, thanks to advances in statistical modeling, there is no longer any reason to ignore the possibility of bias in the signal detection approach to mAFC—one can easily make use of all the information available in the data and fit the mAFC SDT model with bias parameters, as shown here.

The models can also be extended in various ways. For example, if small sample sizes are obtained for each observer, yet one has a large number of observers, then estimation might be improved by specifying the bias and detection parameters as being random across observers, giving a hierarchical mAFC SDT model. Note that the hierarchical approach can easily be implemented with only minor modifications to the OpenBUGS program given in Appendix B (i.e., use parameters instead of values in the priors). This is an advantage of the Bayesian approach over maximum likelihood. Hierarchical Bayesian models have previously been used for basic SDT models, see for example Lee (2008), Morey, Pratte, and Rouder (2008) and Rouder and Lu (2005); also see the February 2011 issue of the Journal of Mathematical Psychology, which is a special issue on hierarchical Bayesian models.

Another interesting extension would be a mixture version of *m*AFC, which would allow for possible mixing that might arise from various factors, such as attention (DeCarlo, 2002). This can easily be accomplished by including a latent dichotomous variable in the structural models presented above (see DeCarlo, 2010). An advantage of the Bayesian approach is that this type of extension is simple to implement in complex models such as that presented here (for another example, see DeCarlo, 2011), although a caution is that one has to take care that the posterior distribution is still actually being sampled from, as noted by Natarajan and McCulloch (1995) and others (Jackman, 2009; Lynch, 2007). Kellen and Klauer (2011) recently considered a mixture version of SDT for 4AFC, but they only considered a model without bias and they used proportion correct as the outcome variable.

Further study of the *m*AFC SDT model and approaches to estimation, as well as real-world applications, are needed.

Appendix A. The differencing approach for 2AFC

A.1. SDT and differencing

As discussed in the text, the *decision rule* for 2AFC with bias is

$$Y = 1 \quad \text{if } \Psi_1 + b > \Psi_2$$
$$Y = 2 \quad \text{if } \Psi_1 + b \le \Psi_2.$$

In the 'differencing' approach, the decision rule is rewritten as

$$Y = 1 \quad \text{if } \Psi_2 - \Psi_1 < b$$
$$Y = 2 \quad \text{if } \Psi_2 - \Psi_1 \ge b.$$

It follows that

$$p(Y = 1|X_1, X_2) = p(\Psi_2 - \Psi_1 < b|X_1, X_2)$$

with $X_1 = 1$ for signal in Position 1 and 0 otherwise, and $X_2 = 1$ for signal in Position 2 and 0 otherwise. Substituting the structural model (given in the text) gives

$$\begin{aligned} p(\Psi_2 - \Psi_1 < b|X_1, X_2) &= p(dX_2 + \varepsilon_2 - (dX_1 + \varepsilon_1) < b) \\ &= p(\varepsilon_2 - \varepsilon_1 < b + dX_1 - dX_2). \\ &= p(\varepsilon_2 - \varepsilon_1 < b + dZ), \end{aligned}$$

where the last line follows with $Z = X_1 - X_2$, as done in the text for 2AFC. For $\varepsilon_i \sim N(0, 1)$ it follows, assuming that ε_1 and ε_2 are independent, that $(\varepsilon_1 - \varepsilon_2) \sim N(0, \sqrt{2})$ and so

$$p(\varepsilon_2 - \varepsilon_1 < b + dZ) = \Phi\left(\frac{b + dZ}{\sqrt{2}}\right).$$

In short, the $\sqrt{2}$ term arises from the differencing of the two random variables ε_1 and ε_2. The above gives Eq. (4) in the text.

A.2. A note on the invariance of *d*

Note that Eq. (1) for yes/no detection and Eq. (3) for 2AFC both give estimates of *d*. A basic question of interest is whether *d* is invariant across the different procedures, which would be important evidence in favor of SDT. There are in fact a number of experiments in psychophysics that have provided evidence that *d* is invariant (e.g., Green & Swets, 1988; Macmillan & Creelman, 2005; Schulman & Mitchell, 1965; Shipley, 1965; Swets, 1959), as well as some studies in food science (e.g., Lee, van Hout, & Hautus, 2007). Macmillan and Creelman, however, also noted some studies that do not support invariance (although information about the standard errors of the parameter estimates is needed).

It should be noted that the invariance of *d* is often discussed in another way, namely as a "$\sqrt{2}d$" prediction (e.g., Wickelgreen, 1968). Here it is noted that the appearance of $\sqrt{2}$ is simply a consequence of using the differencing approach (i.e., it does not occur if Eq. (3) is used). That is, it follows from either Eq. (4) (with zero bias) or from the forced choice model for proportion correct given above that one of the differenced distributions is located at $d/\sqrt{2}$ and the other is located at $-d/\sqrt{2}$, and so the distance between the differenced distributions is $d/\sqrt{2} - (-d/\sqrt{2}) = 2d/\sqrt{2} = \sqrt{2}d$, which is what gives rise to the "$\sqrt{2}d$" prediction. This does not mean that detection is $\sqrt{2}$ times better in forced choice than detection (this misunderstanding has appeared in the literature), but rather it simply means that the forced choice distance estimate must be divided $\sqrt{2}$ to get an estimate of *d*. Note that, if one uses Eq. (3), then a direct (i.e., not divided by $\sqrt{2}$ as in Eq. (4)) estimate of *d* is obtained, and so estimates of *d* obtained for 2AFC with Eq. (3) will be the same (within sampling error) as those obtained for detection with Eq. (1), if invariance indeed holds (and so there is no $\sqrt{2}$ prediction). The reader can verify that estimates of *d* will be the same by fitting Eqs. (1) and (3) to the yes/no and 2AFC data provided by Shipley (1965), which provides a nice example of invariance. In short, the general prediction is correctly stated as *invariance of d* across detection and 2AFC.

Appendix B. A SAS NLMIXED program for 3AFC with bias

```
filename name 'C:\Desktop\3AFC.txt';
*create Z's, recode y's from 1,2 to 1,0, get data in correct form;
DATA first;
 infile name delimiter = '09'x firstobs = 2;
 input x1 x2 ytemp1 ytemp2; trial = _N_;
 y1 = 2 − ytemp1; y2 = 2 − ytemp2;
 z = x1 − x2; z1 = 1 − 2*x1 − x2; z2 = 1 − x1 − 2*x2;
 array resp[2] y1 − y2;
 do i = 1 to 2; y = resp[i];
 if i = 1 then i1 = 1; if i = 1 then i2 = 0;
 if i = 2 then i1 = 0; if i = 2 then i2 = 1;
 output; end;
 keep trial z z1 z2 i1 i2 y;
 run;
 *create linear predictors, specify random eps covariance
  structure;
 PROC NLMIXED data = first qpoints = 20;
 parms b1 − b2 = 0 d = 0;
 eta1 = (b1−b2+d*z+eps1)*i1+(b2−b1−d*z+eps2)*i2;
 eta2 = (b1 − d*z1 + eps1)*i1 + (b2 − d*z2 + eps2)*i2;
 p = probnorm(eta1)*probnorm(eta2);
 model y ∼ binary(p);
 random eps1 eps2 ∼ normal([0,0],[1,0,1]) subject = trial;
 ods output parameter estimates = subjpar;
 run;
```

An OpenBUGS program for 3AFC with bias

```
# 3AFC with bias
model 3AFC
{
#(mildly informative) priors for parameters d, b1, b2
#one can also use the bounds function I(0,) with d
d~ dnorm(0,.1)
b1~ dnorm(0,.1)
b2~ dnorm(0,.1)
for (i in 1:N) {
   eps1[i] ~ dnorm(0,1.0)
   eps2[i] ~ dnorm(0,1.0)
   z[i] <- x1[i] − x2[i]
   z1[i] <- 1 − 2*x1[i] − x2[i]
   z2[i] <- 1 − x1[i] − 2*x2[i]
   p1[i, 1] <-
phi(b1 − b2 + d*z[i] + eps1[i])*phi(b1 − d*z1[i] + eps1[i])
   p1[i, 2] <- 1 − p1[i, 1]
   p2[i, 1] <-
phi(−b1 + b2 − d*z[i] + eps2[i])*phi(b2 − d*z2[i] + eps2[i])
   p2[i, 2] <- 1 − p2[i, 1]
   y1[i] ~ dcat(p1[i,1:2])
   y2[i] ~ dcat(p2[i,1:2])
   }
}
#data
list(N = 150)
#inits
list(d = 1,b1 = 0,b2 = 0)
```

References

Albert, J. H. (1992). Bayesian estimation of normal ogive response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics*, *17*, 251–269.

Ansari, A., & Iyengar, R. (2006). Semiparametric Thurstonian models for recurrent choices: a Bayesian analysis. *Psychometrika*, *71*, 631–657.

Begg, C. B., & Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, *71*, 11–18.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day.

Böckenholt, U. (2006). Thurstonian-based analyses: past, present, and future utilities. *Psychometrika*, *71*, 615–629.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*, 9–25.

Congdon, P. (2005). *Bayesian models for categorical data*. New York: Wiley.

Critchlow, D. E., Fligner, M. A., & Verducci, J. S. (1991). Probability models on rankings. *Journal of Mathematical Psychology*, *35*, 294–318.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. New York: Chapman & Hall.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*, 186–205.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: theoretical developments with applications to recognition memory. *Psychological Review*, *109*, 710–721.

DeCarlo, L. T. (2003). Using the PLUM procedure of SPSS to fit unequal variance and generalized signal detection models. *Behavior Research Methods, Instruments, & Computers*, *35*, 49–56.

DeCarlo, L. T. (2010). On the statistical and theoretical basis of signal detection theory and extensions: unequal variance, random coefficient, and mixture models. *Journal of Mathematical Psychology*, *54*, 304–313.

DeCarlo, L. T. (2011). Signal detection theory with item effects. *Journal of Mathematical Psychology*, *55*, 229–239.

Ennis, D. M., & O'Mahony, M. (1995). Probabilistic models for sequential taste effects in triadic choice. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1088–1097.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York: Springer-Verlag.

Flegal, J. M., Haran, M., & Jones, G. L. (2008). Markov chain Monte Carlo: can we trust the third significant figure? *Statistical Science*, *23*, 250–260.

Frijters, J. E. R., Kooistra, A., & Vereijken, P. F. G. (1980). Tables of d' for the triangular method and the 3-AFC signal detection procedure. *Perception & Psychophysics*, *27*, 176–178.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. (2004). *Bayesian data analysis* (2nd ed.). New York: Chapman & Hall, CRC.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science*, *7*, 473–483.

Green, D. M., & Dai, H. (1991). Probability of being correct with 1 of *M* orthogonal signals. *Perception & Psychophysics*, *49*, 100–101.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Publishing.

Hacker, M. J., & Ratcliff, R. (1979). A revised table of d' for *M*-alternative forced choice. *Perception & Psychophysics*, *26*, 168–170.

Jackman, S. (2009). *Bayesian analysis for the social sciences*. New York: Wiley.

Kellen, D., & Klauer, K. C. (2011). Evaluating models of recognition memory using first- and second-choice responses. *Journal of Mathematical Psychology*, *55*, 251–266.

Koehler, E., Brown, E., & Haneuse, S. J.-P. A. (2009). On the assessment of Monte Carlo error in simulation-based statistical analysis. *The American Statistician*, *63*, 155–162.

Lee, M. D. (2008). Three case studies in Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*, 1–15.

Lee, H.-S., van Hout, D., & Hautus, M. J. (2007). Comparison of performance in the *A*-Not A, 2-AFC, and same-different tests for the flavor discrimination of margarines: the effect of cognitive decision strategies. *Food Quality and Preference*, *18*, 920–928.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics*, *50*, 325–335.

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, *46*, 673–687.

Liu, Q., & Pierce, D. A. (1994). A note on Gauss–Hermite quadrature. *Biometrika*, *81*, 624–629.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*: *Vol. 1* (pp. 103–189). New York: Wiley.

Luce, R. D. (1994). Thurstone and sensory scaling: then and now. *Psychological Review*, *101*, 271–277.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York: Springer.

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

McCulloch, C. E., & Searle, S. (2001). *Generalized, linear, and mixed models*. New York: Wiley.

Morey, R. D., Pratte, M. S., & Rouder, J. N. (2008). Problematic effects of aggregation in zROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology*, *52*, 376–388.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

Natarajan, R., & McCulloch, C. E. (1995). A note on the existence of the posterior distribution for a class of mixed models for binomial responses. *Biometrika*, *82*, 639–643.

Nisbett, R. E., & Wilson, T. C. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*, 146–178.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*, 12–35.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, *8*, 185–2005.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*, 573–604.

Schulman, A. I., & Mitchell, R. R. (1965). Operating characteristics from yes–no and forced-choice procedures. *The Journal of the Acoustical Society of America*, *40*, 473–477.

Sheu, C.-F., Chen, C.-T., Su, Y.-H., & Wang, W.-C. (2005). Using SAS PROC NLMIXED to fit item response theory models. *Behavior Research Methods*, *37*, 202–218.

Shipley, E. F. (1965). Detection and recognition: experiments and choice models. *Journal of Mathematical Psychology*, *2*, 277–311.

Spiegelhalter, D., Thomas, A., Best, N. G., & Lunn, D. (2003). WinBUGS user manual. Retrieved from: http://www.mrc-bsu.cam.ac.uk/bugs/.

Swets, J. A. (1959). Indices of signal detectability obtained with various psychophysical procedures. *Journal of the Acoustical Society of America*, *31*, 511–513.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.
Thomas, A., O'Hara, B., Ligges, U., & Sturtz, S. (2006). Making BUGS open. *R News*, *6*, 12–17. Retrieved from: http://www.rni.helsinki.fi/~boh/publications/Rnews2006-1.pdf.
Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.
Vonesh, E. F., & Chinchilli, V. M. (1996). *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker.
Wickelgreen, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, *5*, 102–122.
Wickens, T. D. (2002). *Elementary signal detection theory*. NY: Oxford.
Yao, G., & Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sample. *British Journal of Mathematical and Statistical Psychology*, *52*, 79–92.
Yellott, J. I. (1977). The relation between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*, 109–144.