Journal of Mathematical Psychology 57 (2013) 43–51

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

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

Signal detection models for the same–different task

Lawrence T. DeCarlo

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

HIGHLIGHTS

- Signal detection models for the same–different task are presented.
- The models apply to the full four by two same–different table.
- The approach allows one to consider models that recognize bias and other effects.
- Two basic three-parameter signal detection models are presented.
- It is shown how to fit the models with standard software for nonlinear mixed models.

ARTICLE INFO

Article history:
Received 29 September 2012
Received in revised form
12 February 2013
Available online 26 March 2013

Keywords:
Signal detection
Same–different task
Covert decisions
Differencing
Nonlinear mixed models

ABSTRACT

Signal detection models for the same–different task are presented. In contrast to the standard approach that only considers the proportion correct, the models apply to the full four by two same–different table. The approach allows one to consider models that recognize bias and other effects. Two basic signal detection models, associated with different decision rules, are presented. A version of the covert decisions rule is introduced that directly allows for same–different bias, in contrast to earlier versions. It is shown how to fit the models with standard software for nonlinear mixed models. The models are applied to data from a recent same–different study.

The *same–different* task requires observers to decide if two presented events are the same, say AA or BB, or different, say AB or BA (Macmillan & Creelman, 2005). An interesting aspect of the task is that observers can perform it even if they do not know how the events differ (i.e., how A differs from B). For example, observers can be asked to decide if two products are the same or different without necessarily knowing *how* the products differ. This aspect makes the procedure attractive for certain applications, such as in speech research or in food science, and so the task has been widely used (e.g., see Macmillan, Kaplan, & Creelman, 1977; Noreen, 1981; O'Mahony & Rousseau, 2002; Stillman & Irwin, 1995).

The standard approach to analyzing data from the same–different task is to collapse AA and BB presentations into a 'same' category, and AB and BA presentations into a 'different' category, and to then obtain 'hits' and 'false alarms' from the resulting two by two table (e.g., Macmillan & Creelman, 2005). Note that, because there are only two observations (i.e., hits and false alarms), models with more than two parameters are not identified, and so the standard signal detection models include only one or two parameters, such as detection and criterion location. However, as noted by Petrov (2009), the same–different task actually provides four observations, that is, the task produces a four by two table, with events of AA, AB, BA, and BB as the four rows and responses of 'different' or 'same' as the two columns. By not collapsing the table (and losing information), one can consider more realistic signal detection models for the same–different task. For example, a model that allows for direct bias towards a response of 'same' or 'different' is introduced here.

Two basic signal detection models for the same–different task are presented. The models follow from different decision rules that have been discussed in the literature, but differ from the standard approach in that they apply to the full four by two table and so they use all of the information available in the data. It is shown how to fit the models with standard software for nonlinear mixed models and with software for Bayesian estimation. The models are applied to data from a recent same–different study.

1. The same–different task

Two basic approaches to the same–different task are discussed in Macmillan and Creelman (2005): the 'independent observation'

E-mail address: decarlo@tc.edu.

0022-2496/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmp.2013.02.002

approach and the 'differencing' approach. Both approaches are examined here, with the difference that additional parameters are introduced, which represent 'bias' and 'asymmetry', as discussed below. In addition, it is shown that the 'differencing' approach also has a representation in the original (non-differenced) decision space.

1.1. Covert decisions

One approach is to simply make separate covert decisions for each event, which has been referred to variously as the 'independent observation' decision rule (Macmillan & Creelman, 2005), the 'β-strategy' (e.g., Rousseau, 2001), or the 'covert categorization' approach (Petrov, 2009). It will be referred to here simply as the *covert decisions* approach. The idea is that the observer makes (covert) decisions as to whether each event is A or B, and then responds 'same' for AA or BB decisions and 'different' for AB or BA decisions.

Decision rule. The observer responds 'different' if the two covert decisions differ, otherwise they respond 'same'. The decision rule is generalized here to allow for response bias, which basically allows the criterion to have different locations across the two decisions. In particular, for the same–different task, 'bias' refers to a tendency to favor one of the responses, such as 'same' or 'different', in the same manner as in the simple detection task, where 'bias' refers to a tendency to favor a response of 'yes' or 'no'. A point that has been somewhat overlooked is that this type of bias occurs for the covert decisions approach if the observer changes the location of their response criterion after the first decision. For example, a bias towards responding 'same' will occur if either (1) the first perception is below the criterion and a higher criterion is then used for the second perception or (2) the first perception is above the criterion and a lower criterion is then used for the second perception. Thus, a decision rule with bias for the same–different task is,

$$Y = 1 \quad \text{if} \begin{cases} \Psi_1 < c \quad \text{and} \quad \Psi_2 > c + b \\ \qquad\qquad \text{or} \\ \Psi_1 > c \quad \text{and} \quad \Psi_2 < c - b, \end{cases} \tag{1}$$

else $Y = 0$, where $Y = 1$ indicates a response of 'different' and $Y = 0$ indicates a response of 'same', Ψ_1 is the perception associated with the event in the first position (or first presented) and Ψ_2 is the perception of the second event. Note that a positive value of b in Eq. (1) indicates a bias towards a response of 'same' whereas a negative value indicates a bias towards a response of 'different'.

Fig. 1 illustrates the decision rule for the situation where B (say a 'noise') is presented in the first position and A (say a 'signal') in the second position, that is, BA trials. Consider the case where the bias b is positive, which is a bias towards a response of 'same'. As shown in the top panel of Fig. 1, if a realization from the first perception Ψ_1 (solid circle) is below the response criterion c, so that the first decision is 'B', then the second criterion is raised by b to $c + b$, which makes a decision of 'B' for the second event more likely, and so the probability of a response of 'same' (BB in this case) is greater. Similarly, as shown in the lower panel of Fig. 1, if a realization from Ψ_1 (solid circle) is above the criterion c, the first decision is 'A' and the second criterion is then located at $c-b$, which makes a decision of 'A' for the second event more likely, and so the probability of a response of 'same' (AA in this case) is again greater. The figure shows an example where the response would have been 'different' in both the top and bottom panels if only one criterion location (c) had been used, but because of the positive bias, the response is 'same' in both cases (because the open circle is below the second criterion in the top panel and above it in the bottom panel). Thus, b reflects bias towards one of the responses in the same–different

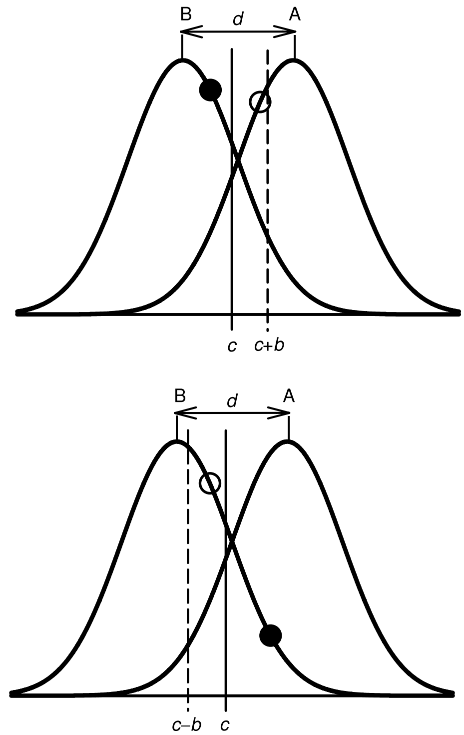

Fig. 1. An illustration of the covert decisions rule with bias for the same–different task.

task in exactly the same manner that the criterion location reflects response bias in the simple detection situation.

Note that families of receiver operating characteristic (ROC) curves can be generated (for different values of d and c) by varying the bias parameter b from minus infinity to plus infinity, which gives hit and false alarm probabilities that vary from zero to one. This is in contrast to other versions of the covert decisions model that have been proposed, as discussed in the ROC section below. Another interesting consequence of the decision rule is that a bias towards a response of 'same' or 'different' means that the *covert decisions are correlated* (the correlation is positive for $b > 0$ and negative for $b < 0$) because the location of the second criterion depends on the location of the first realization.

Structural model. The 'structural model' is the same as that used for m-alternative forced choice (see DeCarlo, 2012), which is

$$\Psi_i = dX_i + \varepsilon_i, \tag{2}$$

for $i = 1, 2$, where ε_i is random variation in the observer's perception Ψ, and X_1 and X_2 are position (temporal or spatial) indicators. For example, $X_1 = 1$ indicates that event A is in the first position and $X_1 = 0$ indicates that event B is in the first position; $X_2 = 1$ indicates that event A is in the second position and $X_2 = 0$ indicates that event B is in the second position.

A response of 'different' occurs if the covert decisions are either 'BA' or 'AB'. It follows from the decision rule of Eq. (1) that the conditional probability of a response of 'different' is

$$\begin{aligned} p(Y = 1|X_1, X_2) &= p(\Psi_1 < c, \Psi_2 > c + b) \\ &\quad + p(\Psi_1 > c, \Psi_2 < c - b). \end{aligned}$$

Substituting the structural model of Eq. (2) and rearranging terms gives

$$\begin{aligned} p(Y = 1|X_1, X_2) &= p(\varepsilon_1 < c - dX_1, \varepsilon_2 > c + b - dX_2) \\ &\quad + p(\varepsilon_1 > c - dX_1, \varepsilon_2 < c - b - dX_2). \end{aligned}$$

If the perceptions are independent (i.e., ε_1 and ε_2), then the above joint probabilities can be written as products and the model can be re-written as

$$\begin{aligned} p(Y=1|X_1,X_2) &= p(\varepsilon_1 < c - dX_1)p(\varepsilon_2 > c + b - dX_2) \\ &\quad + p(\varepsilon_1 > c - dX_1)p(\varepsilon_2 < c - b - dX_2) \\ &= F(c - dX_1)[1 - F(c + b - dX_2)] \\ &\quad + [1 - F(c - dX_1)]F(c - b - dX_2), \end{aligned} \tag{3}$$

where F is a cumulative distribution function (CDF). Eq. (3) gives a general SDT model for the same–different task that follows from the assumption of covert decisions with bias towards a response of 'same' or 'different'. Note that, although the *perceptions* or *observations* (i.e., ε_1 and ε_2) are independent, the *decisions* are *not* independent but rather are correlated, as noted above. The model is also more general than the usual same–different models in that CDFs other than the normal can be used for F in Eq. (3).

The normal theory version of the model follows by using the normal CDF for F in Eq. (3),

$$\begin{aligned} p(Y=1|X_1,X_2) &= \Phi(c - dX_1)[1 - \Phi(c + b - dX_2)] \\ &\quad + [1 - \Phi(c - dX_1)]\Phi(c - b - dX_2). \end{aligned} \tag{4}$$

Eq. (4) is a nonlinear model that is within a family of *nonlinear mixed models* that have recently been discussed for m-alternative forced choice (DeCarlo, 2012), and so it can be fitted in the same manner. For example, it is shown here how to fit the model with the NLMIXED procedure of SAS, as well as with software for Bayesian estimation.

Eq. (4) is a basic signal detection model that follows from the covert decisions approach with response bias, where the bias is towards a response of 'same' or 'different'. Note that a strict assumption is that the bias is symmetrical, in that the criterion shifts up or down by the same amount, b, regardless of whether the realization is above or below the criterion c, as shown in Fig. 1. A generalization is to allow the bias to differ depending on whether the realization is above or below the criterion. More specifically, an *asymmetry parameter a* can be introduced as follows,

$$Y = 1 \quad \text{if} \begin{cases} \Psi_1 < c \quad \text{and} \quad \Psi_2 > c + b + a \\ \qquad\qquad \text{or} \\ \Psi_1 > c \quad \text{and} \quad \Psi_2 < c - b. \end{cases} \tag{5}$$

The parameter a allows for an asymmetry in the 'same' or 'different' bias (b). For example, a positive value of a indicates that the observer has a greater bias towards a response of 'same' when the first realization is below the criterion, as compared to when the realization is above the criterion. In terms of Fig. 1, a positive value of a shifts the bias line shown in the top panel ($c + b$) to the right, whereas a negative value shifts it to the left. Thus, bias towards a response of 'same' can be smaller (negative a) or larger (positive a) in one direction as compared to the other.

Using the structural model of Eq. (2) and the decision rule of Eq. (5), the resulting normal theory version of the model is,

$$\begin{aligned} p(Y=1|X_1,X_2) &= \Phi(c - dX_1)[1 - \Phi(c + b + a - dX_2)] \\ &\quad + [1 - \Phi(c - dX_1)]\Phi(c - b - dX_2). \end{aligned} \tag{6}$$

Eq. (6) includes four parameters and is exactly identified (i.e., it has the same number of parameters as observations); it will be considered along with Eq. (4) and will be referred to as the *asymmetric covert decisions* model.

1.2. Asymmetric yardstick

Another approach to the same–different task is referred to as the 'differencing approach' in Macmillan and Creelman (2005, p. 221); also see Sorkin (1962) and Noreen (1981). The view is that the decisions are based on differenced random variables, such as $\Psi_2 - \Psi_1$. As shown here, however, one does not need to assume that the decision is based upon differenced random variables, because the approach also has a representation in the original decision space. In particular, the observer can be viewed as using a 'yardstick' to help make same–different judgments, without differencing. This has been referred to as the 'τ-criterion' approach (e.g., Lee, van Hout, Hautus, & O'Mahony, 2007, Rousseau, 2001) and will be referred to here simply as the 'yardstick' decision rule, in line with earlier discussions. In the present development, an asymmetry in the yardstick is also allowed for.

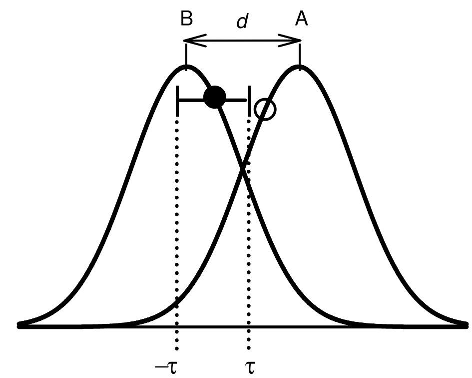

Fig. 2. An illustration of the yardstick decision rule for the same–different task.

Yardstick decision rule. It is assumed that the observer places a "yardstick" around the realization for the first event and makes a decision of 'different' as follows,

$$Y = 1 \quad \text{if} \begin{cases} \Psi_2 > \Psi_1 + \tau + a \\ \qquad \text{or} \\ \Psi_2 < \Psi_1 - \tau, \end{cases} \tag{7}$$

else $Y = 0$. The decision rule is illustrated in Fig. 2 (without the a parameter for visual clarity). In this case, a 'yardstick' (shown by the brackets) is placed around the first realization (solid circle). The parameter τ represents the size of the yardstick (with 2τ being the total length). The decision rule is that if the second realization (open circle) falls within the yardstick brackets, then the decision is 'same', else it is 'different'. Note that any bias towards a response of 'same' or 'different' in this case is simply a part of τ. That is, a larger value of τ means that the probability of a response of 'same' is higher whereas a smaller value indicates that the probability of a response of 'different' is larger (note that, to set a 'no bias' point, one could determine an optimal value of τ and use that as a reference).

The asymmetry parameter a in the decision rule of Eq. (7) allows the size of the yardstick to differ across less-than and greater-than comparisons. With respect to Fig. 2, a positive value of a shifts the rightmost bracket τ to the right, whereas a negative value shifts it to the left. Note that a complication of allowing for asymmetry is that there must be a restriction on the parameter a for negative values, and in particular, if $a < 0$ then $|a| \leq 2\tau$. The reason for the constraint can be seen in Fig. 2—if a is negative and is larger in absolute magnitude than 2τ, then the right-most bracket ($\tau + a$) will be to the left of the left-most bracket ($-\tau$), and the yardstick will have a length that is less than zero, which is clearly not permissible.

It follows from the decision rule of Eq. (7) that the probability of a response of 'different' is

$$p(Y=1|X_1,X_2) = p(\Psi_2 > \Psi_1 + \tau + a) + p(\Psi_2 < \Psi_1 - \tau).$$

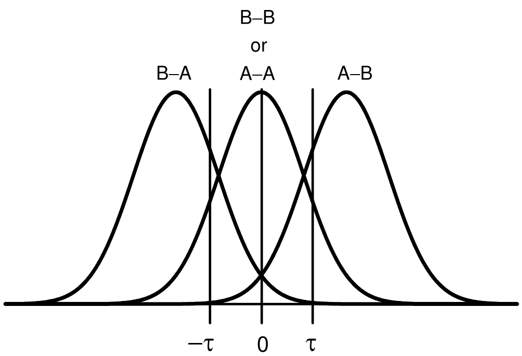

Fig. 3. An illustration of the differencing decision rule for the same–different task. The distributions have standard deviations of $\sqrt{2}$ (instead of 1) in this case.

Substituting the structural model of Eq. (2) and rearranging terms gives,

$$\begin{aligned} p(Y=1|Z) &= p(\varepsilon_2 > \tau + a + dZ + \varepsilon_1) \\ &\quad + p(\varepsilon_2 < -\tau + dZ + \varepsilon_1) \\ &= 1 - F(\tau + a + dZ + \varepsilon_1) + F(-\tau + dZ + \varepsilon_1), \end{aligned}$$

where $Z = X_1 - X_2$. The above can be fit by conditioning on a realization of ε_1, say e_1, and integrating (for details, see DeCarlo, 2012), which gives

$$\begin{aligned} p\,(Y=1|Z) = \int_{-\infty}^{\infty} &[1 - F(\tau + a + dZ + e_1) \\ &+ F(-\tau + dZ + e_1)]f(e_1)de_1. \end{aligned} \tag{8}$$

Eq. (8) will be referred to as the *asymmetric yardstick* signal detection model.

The normal theory version of Eq. (8) can be fit by using maximum likelihood estimation with Gaussian quadrature or by using Bayesian estimation, as discussed in DeCarlo (2012); both approaches are illustrated here. The model can also be written in a simpler form by using a version that follows from a differencing approach, as shown next. Note, however, that the simpler form also follows directly from Eq. (8), because of a relation of the integral to another form, as shown in DeCarlo (2012, p. 198).

Differencing decision rule. Another view of the yardstick decision rule of Eq. (7) arises if one re-writes it in terms of differences,

$$Y = 1 \quad \text{if} \begin{cases} \Psi_2 - \Psi_1 > \tau + a \\ \qquad \text{or} \\ \Psi_2 - \Psi_1 < -\tau, \end{cases}$$

else $Y = 0$. The decision rule is illustrated in Fig. 3 (with $a = 0$ for visual clarity). The three distributions in Fig. 3 are differenced distributions associated with B–A, B–B or A–A, and A–B, respectively. If a realization from one of the distributions in Fig. 3 is between $-\tau$ and τ, then the decision is 'same', else it is 'different'. Fig. 3 shows that $-\tau$ and $+\tau$ are simply fixed response criteria in the differenced decision space, rather than being moving yardsticks (with locations that depend on each realization) as in Fig. 2. Thus, τ has different interpretations in different decision spaces. The asymmetry parameter a allows τ on the right to have a different distance from the zero point than $-\tau$ on the left, and so the criteria locations are asymmetrical. For example, a positive value of a indicates a greater tendency to respond 'same' in one direction as compared to the other. It again follows that the restriction $|a| \le 2\tau$ for $a < 0$ is necessary so that the criterion on the right of Fig. 3 (i.e., $\tau + a$) is not below the left-most criterion ($-\tau$). Thus, the constraint on a for the differencing rule is the same as for the yardstick rule, which simply reflects the relation between the decision rules.

It follows from the differencing decision rule given above that the probability of a response of 'different' is

$$p(Y=1|X_1,X_2) = p(\Psi_2 - \Psi_1 > \tau + a) + p(\Psi_2 - \Psi_1 < -\tau).$$

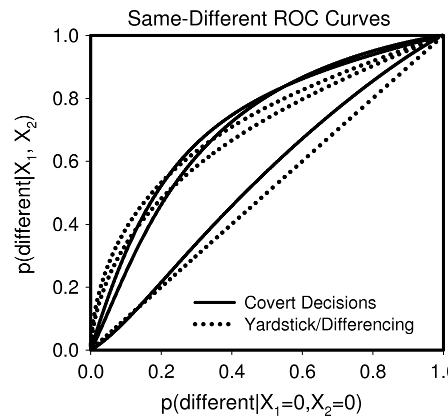

Fig. 4. ROC curves for the asymmetric covert decisions and asymmetric yardstick models for the same–different task.

Substituting the structural model of Eq. (2) and noting that, for the normal theory version of the model, if $\varepsilon_i \sim N(0, 1)$ and ε_1 and ε_2 are independent then $\varepsilon_2 - \varepsilon_1 \sim N(0, \sqrt{2})$, it follows that

$$p\,(Y=1|Z) = 1 - \Phi\left(\frac{\tau + a + dZ}{\sqrt{2}}\right) + \Phi\left(\frac{-\tau + dZ}{\sqrt{2}}\right), \tag{9}$$

where $Z = X_1 - X_2$. The above form of the model is equivalent to the yardstick model of Eq. (8), which follows from a relation discussed in DeCarlo (2012, p. 198). The asymmetric difference model of Eq. (9) is referred to as the DF2 model in Petrov (2009); setting a equal to zero gives the standard differencing model for the same–different task (e.g., Eq. (9.8) in Macmillan & Creelman, 2005; referred to as the DF1 model in Petrov, 2009).

As noted above, Eq. (9) is equivalent to Eq. (8), which can be seen by fitting the two models using the programs provided in the Appendix. Given that the yardstick and differencing approaches lead to equivalent models for the same–different task, they will be considered interchangeably here and will be referred to simply as the yardstick/differencing model. However, this is not meant to imply that the decision approaches are the same in general. That is, although differencing and non-differencing approaches give the same models in several cases (e.g., for the same–different task and for two-alternative forced choice), this is not necessarily the case in general, such as for forced choice with three or more alternatives, as has previously been noted (DeCarlo, 2012). It is also interesting to note that the covert decisions and the yardstick approaches only require qualitative decisions about the perceptions, that is, 'greater than' or 'less than' decisions, whereas the differencing approach involves quantitative information about the perceptions, because of the differencing operation (which implies a unit). In short, covert decisions, yardstick, and differencing are *three* different decision rules for the same–different task.

1.3. Same–different ROC curves

The usual approach to generating receiver operating characteristic (ROC) curves for the same–different task applies to the collapsed two by two table, which gives one ROC curve (for each value of d). However, when the four different types of events (i.e., BB, AA, AB, BA) are recognized, as done here, then there are actually three ROC curves, and not simply one. For example, if the BB row is used as the reference for 'false alarms' (i.e., the observer responds 'different' when BB is presented), then there are three ROC curves (as the false alarm rate varies) associated with each of the other three rows.

Note that, for the simple yes–no detection task, ROC curves can be generated by varying the bias, that is the criterion location, from minus infinity to plus infinity, which gives false alarm probabilities that vary from zero to one and hit probabilities that also vary from zero to one. This also applies to the same–different models presented here, in that varying the bias parameter generates the usual type of ROC curve. For example, for the covert decisions model, a family of ROC curves can be generated by varying b in Eq. (6) (or Eq. (4)) from minus infinity to plus infinity with values of d, c, and a held constant. Similarly, for the yardstick model of Eq. (9), the size of τ reflects bias, and so ROC curves can be generated by varying τ from zero to infinity (for fixed d and a), which gives false alarm and hit probabilities that vary between zero and one.

It is informative to examine the ROC curves that follow from the models of Eqs. (6) and (9) when b or τ is varied. Fig. 4 shows ROC curves for both the asymmetric covert decisions model (Eq. (6), solid lines) and the asymmetric yardstick model (Eq. (9), dotted lines). The plots are for average values of the parameters (d, c, and a) obtained in the analysis presented below (see Tables 3 and 4). Note that, for the covert decisions model, the curve for AA (lowest solid line) is close to the diagonal, whereas for the yardstick model, the curve for AA is exactly on the diagonal (dotted line). This occurs because it follows from the yardstick/differencing model that p('different'|AA) $= p$('different'|BB), as Petrov (2009) noted. Overall, Fig. 4 shows that the ROC curves are fairly similar across the two models, which is relevant to results obtained for the application discussed below.

It is important to recognize that the usual covert decisions model with only a non-moving (fixed) criterion c (i.e., Eq. (4) with $b = 0$, discussed as CC1 in Petrov (2009)) does *not* generate typical ROC curves, in that varying c from minus to plus infinity does not give hit and false alarm probabilities that vary from zero to one (with monotonically decreasing slopes); this is also the case if two fixed criteria are used (e.g., CC2a, CC2s, in Petrov). For example, Fig. 5 shows examples of ROC curves that arise by varying c in the covert decisions model (with $b = 0$), with $d = 2$ in the left panel and $d = 3$ in the right panel (there are actually three curves, but only two are visible because the 'AB' and 'BA' curves are identical). BB presentations are again used as the reference.

Consider what happens when c varies from minus to plus infinity. As shown by Fig. 5, the false alarm probability (the x-axis) at first increases as the criterion moves to the right (because a covert decision of AA becomes less likely, and so the probability of a response of 'different' increases), up to a maximum of 0.5 (which occurs at the midpoint of the B distribution), and then decreases as the criterion goes past the B midpoint (because a covert decision of BB becomes more likely, and so the probability of a response of 'different' decreases). Similarly, the probability of a 'hit', that is a response of 'different' for AB or BA presentations (solid curve), also at first increases, with a maximum at the midpoint of the two distributions, and then decreases, giving the curved ROC curves shown in Fig. 5.

Fig. 5 shows that varying the criterion location c in the covert decisions model with a fixed criterion (e.g., CC1 in Petrov (2009)) does not directly give bias for a response of 'same' or 'different', but rather only directly gives bias for a covert decision of 'A' or 'B'. Using terminology introduced earlier, the ROC curves are not 'proper' in that they do not pass through the points (0, 0) and (1, 1) and they do not have a monotonically decreasing slope (Egan, 1975). This is also the case for covert decisions models with two criteria (e.g., CC2a, CC2s, RC2a, and RC2s in Petrov). For example, ROC curves associated with these models do not pass through the point (1, 1), because, as for the CC1 model, the false alarm rate cannot go above 0.5 (the probability of a 'different' response is a maximum when the 'ambiguous' region is zero, which simply gives the CC1 model). Another problem with the two criteria models is that, in some cases, a same–different decision can be based on only one observation. For example, if the first perception is in the 'ambiguous' region (between the two criteria) then the decision is 'same' irrespective of the second perception, and so a decision is made without observing both events!

In sum, if one wants proper ROC curves, as in Fig. 4, and not improper curves, as in Fig. 5, then one can rule out all of the covert categorization rules discussed in Petrov (2009). In contrast, the bias parameter b in the covert decisions models introduced here directly reflects bias for a response of 'same' or 'different' and gives proper ROC curves. Note that this problem appears to be unique to covert decisions models, and does not arise for the yardstick or differencing models, where τ directly reflects bias for a response of 'same' or 'different' and the resulting ROC curves are proper.

1.4. Fitting the models

The models of Eqs. (4), (6), (8) and (9) are within a family of *nonlinear mixed models* that have recently been discussed for m-alternative forced choice (DeCarlo, 2012), and so they can be fitted in exactly the same manner. The Appendix provides a program that shows how to use the NLMIXED procedure of SAS to fit the models of Eqs. (6), (8) and (9) (other software, such as R, can also be used with similar programs). Programs for Bayesian estimation of Eqs. (6) and (8) (with OpenBUGS; Thomas, O'Hara, Ligges, & Sturtz, 2006) are also provided. Although the Bayesian approach differs from the maximum likelihood approach, parameter estimates obtained for Bayesian estimation with OpenBUGS, as applied to Petrov's (2009) data, were virtually identical to those obtained for maximum likelihood estimation with SAS.

Eqs. (6) and (9) are simply nonlinear models (i.e., they are not mixed because there is no random term). Note that, for Eq. (9), the SAS program given in the Appendix includes a $\sqrt{2}$ term in

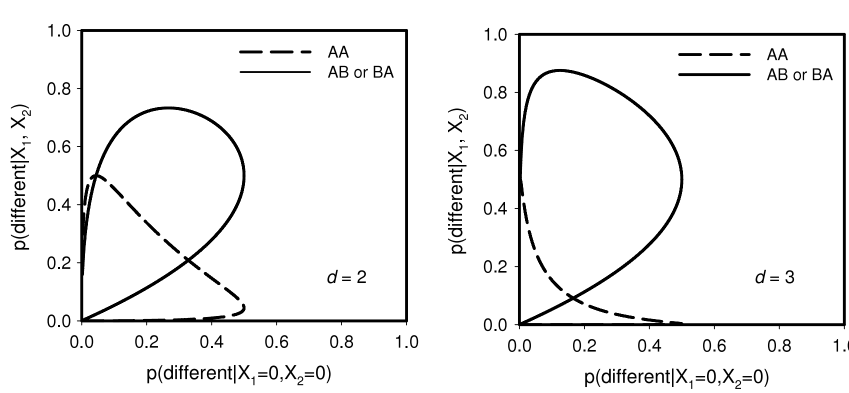

Fig. 5. ROC curves for the covert decisions model with only a 'fixed' criterion c for values of d of 2 and 3.

Table 1
Chi-square and likelihood ratio goodness of fit statistics.

Obs.	Covert decisions				Asymmetric yardstick			
	ChiSq	p	LR	p	ChiSq	p	LR	p
1	3.22	0.07	3.25	0.07	7.27	<0.01	7.33	<0.01
2	1.06	0.30	1.06	0.30	1.78	0.18	1.81	0.18
3	11.94	<0.01	12.00	<0.01	2.18	0.14	2.20	0.14
4	3.01	0.08	3.02	0.08	0.24	0.62	0.24	0.62
5	8.57	<0.01	8.65	<0.01	1.48	0.22	1.49	0.22
6	45.79	<0.01	47.34	<0.01	0.09	0.77	0.09	0.77
7	13.91	<0.01	14.02	<0.01	0.03	0.87	0.03	0.87
8	36.09	<0.01	37.56	<0.01	0.27	0.60	0.27	0.60
9	5.47	0.02	5.49	0.02	0.45	0.50	0.45	0.50
10	1.25	0.26	1.25	0.26	2.40	0.12	2.41	0.12
11	0.17	0.68	0.17	0.68	0.02	0.89	0.02	0.89
12	3.14	0.08	3.15	0.08	0.28	0.59	0.28	0.59
13	3.12	0.08	3.11	0.08	3.35	0.07	3.37	0.07
14	1.60	0.21	1.60	0.21	1.53	0.22	1.53	0.22

Notes: Obs. is the observer's number; the sample size for each observer is 512. The statistics for the differencing model are the same as for the asymmetric yardstick model.

the model specification, so that one obtains direct estimates of d, which eliminates some confusion that has arisen regarding the scale factor $\sqrt{2}$ (such as in forced choice models; see DeCarlo, 2012). Eq. (8) is a nonlinear *mixed* model because it includes a random term in the linear predictor; it can be fitted using Gaussian quadrature (in SAS) or Bayesian estimation (with OpenBUGS). Note that, to fit Eq. (8) with SAS, one must include a 'trial' variable (i.e., the observation number) in the data and the data must be sorted by this variable. Also note that the constraint on the parameter a noted above was not explicitly introduced (although it could be) because it was not necessary for the analysis presented below, in that, for all observers, any negative estimates of a were clearly smaller in absolute value than 2τ. If the constraint on a is not explicitly introduced, then one must check the estimates to ensure that the constraint is satisfied.

2. Application to a same–different experiment

The models are applied to data from a recent same–different experiment (Petrov, 2009).[1] The task was to detect, in two presentations, whether black dots in a display moved in the same or different directions. A detection task was also included, where the task was to determine, in one presentation, whether the dots moved in a clockwise or counterclockwise direction compared to a reference direction. The data were obtained across multiple sessions; only data from the last session (sixth session) are analyzed here. For each observer, there were 512 same–different trials and 512 detection trials. For the same–different data, the signal detection models of Eq. (4) (covert decisions), Eq. (6) (asymmetric covert decisions), and Eq. (9) (asymmetric yardstick/differencing model) are fitted to the individual data of fourteen observers. Parameter estimates and standard errors are obtained, and absolute and relative fit are examined.

2.1. Model fit

Absolute fit. Table 1 shows likelihood ratio (LR) and chi-square goodness of fit statistics for fits of Eqs. (4) and (9) (covert decisions and asymmetric yardstick/differencing). The asymmetric model of Eq. (6) is not included in the table because there are no degrees of freedom available to test absolute fit (i.e., there are four observations and four parameters), however the relative fit of the model can be assessed, as done in the next section. Table 1 shows that the covert decisions model is rejected (at the 0.05 level) in 6 out of 14 cases and the asymmetric yardstick/differencing model is rejected in 1 case.

Relative fit. Table 2 presents relative fit statistics, namely the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), for all three models, that is, Eqs. (4), (6) and (9). Results for the two three-parameter models of Eqs. (4) and (9) are presented first, followed by results for the four-parameter model of Eq. (6). For each statistic, the bold values indicate the smallest value in each row (except for cases with ties; note that, although smaller

[1] I thank Alexander Petrov for making the raw data available.

Table 2
Relative fit statistics for three models.

Observer	AIC			BIC		
	Eq. (4)	Eq. (9)	Eq. (6)	Eq. (4)	Eq. (9)	Eq. (6)
1	653.8	657.9	**652.6**	**666.5**	670.6	669.5
2	**609.7**	610.4	610.4	**622.4**	623.1	627.3
3	598.5	588.7	**588.5**	611.2	**601.4**	605.5
4	548.2	**545.4**	547.2	560.9	**558.2**	564.2
5	671.1	**663.8**	664.3	683.7	**676.5**	681.3
6	628.6	**581.4**	583.3	641.3	**594.1**	600.2
7	670.6	**657.3**	659.2	683.3	**670.0**	676.2
8	669.7	**632.4**	634.1	682.4	**645.1**	651.1
9	680.3	**675.3**	676.9	693.1	**688.0**	693.8
10	**653.2**	654.4	653.9	**665.9**	667.1	670.9
11	626.8	**626.7**	628.7	639.5	**639.4**	645.6
12	551.0	**548.1**	549.8	563.7	**560.8**	566.8
13	638.3	638.6	**637.2**	**651.0**	651.3	654.2
14	647.5	647.5	647.9	660.2	660.2	664.9

Notes: Eq. (4) is the covert decisions model (three parameters); Eq. (9) is the asymmetric yardstick/differencing model (three parameters); Eq. (6) is the asymmetric covert decisions model (four parameters).

Table 3
Parameter estimates: Asymmetric covert decisions model (Eq. (6)).

Observer	d	c	b	a
1	1.63 (0.13)	0.92 (0.14)	−0.23 (0.18)	0.61 (0.34)
2	1.79 (0.13)	0.78 (0.14)	0.61 (0.16)	−0.33 (0.29)
3	1.84 (0.14)	1.21 (0.17)	1.11 (0.18)	−0.97 (0.28)
4	2.30 (0.12)	1.00 (0.13)	−0.00 (0.17)	0.52 (0.30)
5	1.43 (0.14)	0.64 (0.15)	−0.25 (0.16)	0.99 (0.33)
6	1.89 (0.14)	0.63 (0.14)	−0.42 (0.14)	2.00 (0.29)
7	1.54 (0.14)	1.13 (0.15)	0.64 (0.23)	−1.43 (0.38)
8	1.73 (0.14)	0.43 (0.13)	−0.92 (0.15)	2.15 (0.34)
9	1.21 (0.15)	0.36 (0.17)	−0.12 (0.17)	0.84 (0.36)
10	1.14 (0.16)	0.86 (0.19)	0.63 (0.20)	−0.40 (0.35)
11	1.81 (0.12)	0.87 (0.14)	0.11 (0.17)	0.13 (0.31)
12	2.24 (0.14)	1.12 (0.15)	0.26 (0.17)	0.49 (0.28)
13	1.74 (0.13)	0.57 (0.13)	−0.11 (0.16)	0.57 (0.32)
14	1.68 (0.13)	1.05 (0.14)	0.32 (0.19)	−0.42 (0.33)
Average	1.71	0.83	0.12	0.34

Notes: Standard errors are in parentheses.

Table 4
Parameter estimates: Asymmetric yardstick model (Eq. (9)).

Observer	d	τ	a
1	1.88 (0.16)	1.21 (0.13)	0.47 (0.24)
2	2.02 (0.16)	2.04 (0.17)	−0.40 (0.23)
3	2.04 (0.17)	2.33 (0.18)	−0.71 (0.23)
4	2.76 (0.16)	1.74 (0.14)	0.40 (0.24)
5	1.65 (0.17)	1.11 (0.13)	0.78 (0.24)
6	2.19 (0.19)	1.17 (0.12)	1.69 (0.24)
7	1.89 (0.17)	1.72 (0.17)	−0.94 (0.25)
8	2.08 (0.18)	0.75 (0.11)	1.52 (0.24)
9	1.39 (0.17)	1.15 (0.14)	0.58 (0.26)
10	1.24 (0.18)	1.65 (0.16)	−0.17 (0.26)
11	2.15 (0.15)	1.57 (0.15)	0.09 (0.23)
12	2.60 (0.17)	1.84 (0.14)	0.51 (0.23)
13	2.03 (0.16)	1.39 (0.14)	0.37 (0.24)
14	1.99 (0.15)	1.63 (0.16)	−0.28 (0.24)
Average	1.99	1.52	0.28

Notes: Standard errors are in parentheses.

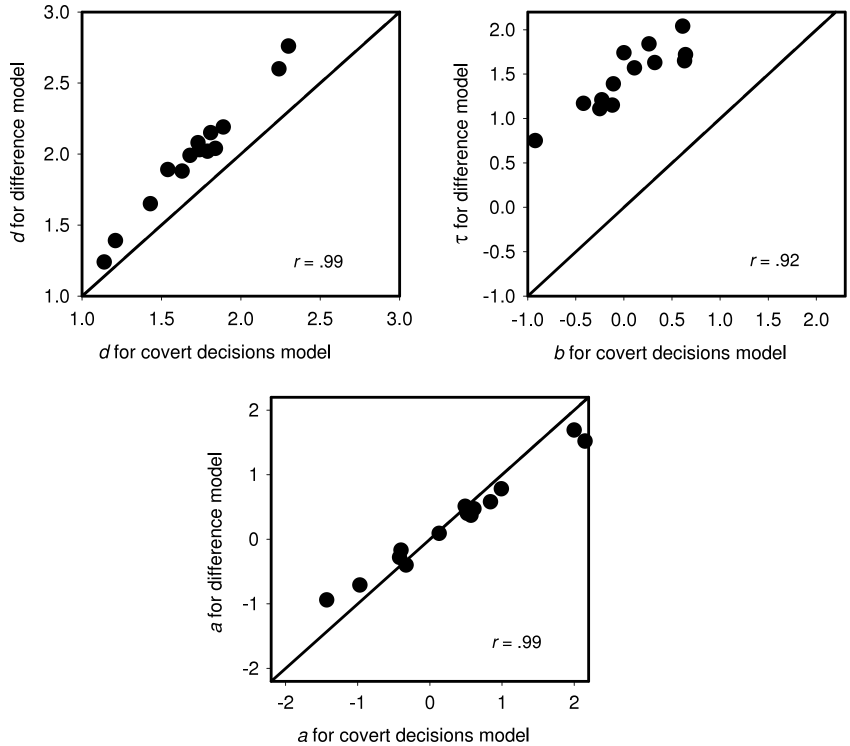

Fig. 6. The top left panel shows estimates of d for the asymmetric yardstick/differencing model plotted against estimates of d for the asymmetric covert decisions model. The top right panel shows estimates of τ for the difference model plotted against estimates of b for the covert decisions model. The bottom panel shows estimates of the asymmetry parameter a for the yardstick/differencing model plotted against the asymmetry parameter a for the covert decisions model.

values indicate better relative fit, a difference of less than, say, 2 can be considered as trivial). For the AIC, the covert decisions models (Eqs. (4) or (6)) are favored in 5 cases and the differencing model is favored in 8 cases. Note, however, that for the 8 cases where the AIC selects (i.e., is smaller) the asymmetric differencing model (Eq. (9)), AIC for Eq. (6) is quite close in magnitude, tending to be within about 2, and so the results do not indicate a clear choice between Eqs. (6) and (9). The right side of Table 1 shows that the BIC is smallest for the differencing model in 9 cases (it is well known that BIC tends to select models with fewer parameters).

ROC curves. As noted earlier (DeCarlo, 2007), ROC curves offer a useful visual supplement to absolute and relative fit statistics. Note that the data analyzed above do not include different conditions where the bias was varied within observers, and so empirical ROC curves are not obtained (i.e., for each observer, there is only one point on each of the three curves). However, it is still informative to plot ROC curves for typical values of the parameters, as done in Fig. 4, to see how the models differ.

Fig. 4 shows ROC curves for the asymmetric covert decisions model (Eq. (6)) and the asymmetric yardstick/differencing model (Eq. (9)). As noted above, the figure shows that the ROC curves are similar, which suggests that it might be difficult in practice to distinguish between the two models. It is informative to note that for Petrov's (2009) data, the x-axis, which gives the false alarm rate, only varied between 0.13 and 0.41 across the 14 observers. If one looks at this region in Fig. 4, it is apparent that there is little difference between the three ROC curves across the two models. This suggests why it is difficult to come to clear-cut conclusions based on fit statistics alone as to which decision rule (if only one) was used by the observers.

2.2. Parameter estimates

Covert decisions model. Table 3 shows parameter estimates and standard errors for fits of Eq. (6) to each observer's data. The average estimate of d is 1.71 and the average estimate of the first criterion location is 0.83. Note that in the simple detection situation, the 'optimal' location of the criterion is at the midpoint of the two distributions (i.e., at $1/2d$) if the observer does not have knowledge about the signal probability (or if they have knowledge about the signal probability but it is 0.50). It is interesting to note that the average estimate of the first criterion location is quite close to this location, that is, one half of the average estimate of d is 0.85 whereas the average estimate of the criterion location is 0.83. With respect to the bias parameter, estimates of b are positive and negative, and are significant in six cases. Estimates of the asymmetry parameter a are also positive and negative, and so the asymmetry is in different directions for different observers, as also found below for Eq. (9).

It is interesting to note that large and significant values of a for Eq. (6) are found in every case where the fit of the covert decisions model of Eq. (4) was rejected (Table 1, Observers 3, 5, 6, 7, 8, and 9). This illustrates why one cannot conclude that rejecting the covert decisions model with bias (Eq. (4)) means that the covert decisions approach was not used—it could simply indicate that the assumption of symmetry in the bias (i.e., $a = 0$) was not valid.

Yardstick/differencing model. Table 4 shows parameter estimates for fits of the differencing model of Eq. (9). The average estimate of d is 1.99, which is larger (by 0.28) than the average d found for the covert decisions model. The average estimate of τ is 1.52. Estimates of the asymmetry parameter a are positive and negative across observers with an average value of 0.28; note that the pattern of positive and negative values of a is exactly the same as that found for a in the asymmetric covert decisions model (Table 3).

Comparisons across models. The top left panel of Fig. 6 presents estimates of d for each observer from the asymmetric yardstick/differencing model (Table 4) plotted against estimates of d from the asymmetric covert decisions model (Table 3). The figure shows that estimates of d from the two models are highly

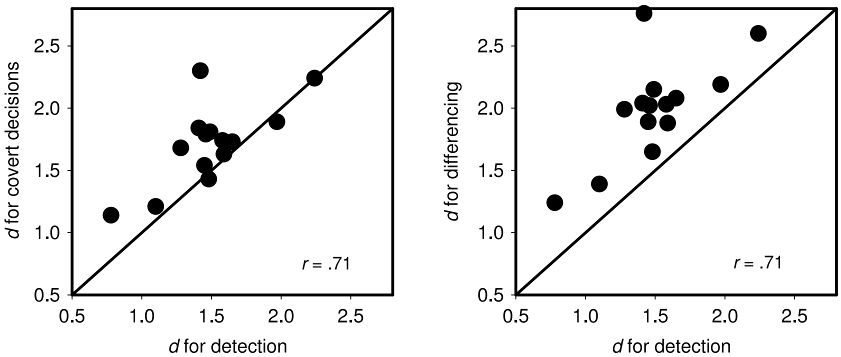

Fig. 7. The left panel shows estimates of *d* for the asymmetric covert decisions model plotted against estimates of *d* from a detection task. The right panel shows estimates of *d* for the asymmetric yardstick/differencing model plotted against estimates of *d* from a detection task.

correlated ($r = 0.99$). Thus, the two models (Eqs. (6) and (9)) give similar estimates of d for the fourteen observers. The figure also shows that estimates of d from the asymmetric yardstick/differencing model are larger than estimates of d from the asymmetric covert decisions model (the points are all above the equality line), which has also been found for fits of the traditional covert decisions model and the differencing model (Irwin, Hautus, & Francis, 2001, Table 1).

The top right panel of Fig. 6 shows that estimates of τ for the asymmetric yardstick/differencing model are related to (but larger than) the estimates of b for the asymmetric covert decisions model, with a correlation of 0.92. The bottom center panel shows that estimates of a for the asymmetric yardstick/differencing model are closely related to those for the asymmetric covert decisions model, with a correlation of 0.99. Thus, like detection, the bias and asymmetry parameters are very similar across the two models.

Relation to detection. The fourteen observers in Petrov's same–different experiment (2009) also participated in a yes–no detection task with the same stimuli. The left panel of Fig. 7 shows a plot of the estimates of d obtained for the asymmetric covert decisions model for the same–different task plotted against estimates of d obtained for the detection task. The figure shows that estimates of d for the same–different task are closely related to estimates of d from detection, with a correlation of 0.71. This is important evidence for the validity of the signal detection approach, in that it shows that the approach gives consistent estimates of d across same–different and detection tasks. The right panel of Fig. 7 shows that estimates of d from the asymmetric yardstick/differencing model are also highly correlated with estimates of d from detection; the figure also shows that the yardstick/differencing estimates of d tend to be larger than in the left panel.

Fig. 7 shows that estimates of d for the same–different task for the covert decisions model are close in magnitude to those found for detection, whereas those for the asymmetric yardstick/differencing model tend to be larger. One could take this as evidence against the yardstick/differencing model, however it is important to recognize that the difference might arise simply because of a violation of an auxiliary assumption. For example, the standard deviations for the differenced distributions are $\sqrt{2}$ if the observations ε_1 and ε_2 are independent; however if they are positively correlated, then the standard deviation is the square root of $2 - 2 \times \text{corr}(\varepsilon_1, \varepsilon_2)$. This means that one can find a positive value for $\text{corr}(\varepsilon_1, \varepsilon_2)$ so that the d are re-scaled to values that are nearly identical to those obtained for the covert decisions model (given that the correlation between the two sets of estimates is 0.99).[2] Thus, the larger estimates of d cannot necessarily be taken as evidence against the yardstick/differencing decision rule, in that it could simply reflect a violation of another assumption.

3. Conclusions

By not collapsing the data, more general models for the same–different task can be considered. Two basic signal detection models that allow for 'bias' and 'asymmetry' are discussed here. The first model is based upon the classic idea of covert decisions. A novel aspect of the model presented here is with respect to how bias is introduced, in that the decision rule allows for direct bias towards a response of 'same' or 'different'. A parameter that allows for asymmetry in the bias also appears to be necessary. The second model is an extension of the standard 'differencing' approach, in that the decision rule allows for asymmetry in the bias, in the same manner as for the asymmetric covert decisions approach.

An important point is that acceptance or rejection of a model does not necessarily mean acceptance or rejection of the associated decision rule. For example, it is shown that the differencing decision rule also has a representation in the original decision space as a 'yardstick' decision rule. Thus, if it is found that the differencing model describes the data, one cannot conclude that the observer is using the differencing decision rule, because the results will be the same if he or she is actually using the yardstick decision rule. As another example, Table 1 suggests rejection of the covert decisions model in six cases, however fits of Eq. (6) showed that, in all of these cases, there was considerable asymmetry in the bias. Thus, it is not necessarily the decision rule (covert decisions) that is being rejected, but possibly rather the assumption of symmetry in the decision rule. It should also be kept in mind that although the yardstick and differencing decision rules lead to equivalent models for the same–different task, this is not necessarily the case in general.

For the data of Petrov (2009), the finding of mixed results for absolute and relative fit statistics, ROC curves that are hard to distinguish for the range of observed false alarms, and parameter estimates that are similar across the models together suggest that one cannot come to any hard and fast conclusions about which decision rule is used (if only one). Further study of the same–different task with the models proposed here is needed.

The models presented here (asymmetric covert decisions and asymmetric yardstick/differencing), although motivated by different decision rules, are apparently closely related. For example, for fits of Petrov's (2009) data, the detection (d), bias (b or τ), and asymmetry parameters (a), were found to be very similar across

[2] This suggests an interesting possible way to estimate the correlation for the yardstick/differencing model, if one has both same–different and detection conditions—one can find the value of the correlation that re-scales d for the same–different condition so that it is the closest in value to d for detection.

the two models. A possible reason for the similar results is suggested by work of Dai, Versfeld, and Green (1996). In particular, they showed that, for likelihood ratio decision rules for the same–different task, having independent observations gives the usual likelihood ratio model (which is closely related to the version of the covert decisions model given here) and having highly correlated observations gives the differencing model. They noted that "Thus, the correlation between observations provides a common thread between the two decision rules" (p. 2). As noted above, the bias parameter in the covert decisions model introduced here allows for correlated responses, whereas correlated responses are also allowed for in the differencing model, in that correlation simply re-scales the detection parameter d. Thus, the fact that both models allow for correlation suggests that this might be the 'common thread' between them.

The models presented here fall within a well-developed statistical framework (nonlinear mixed models) and can easily be fitted with standard software, which should encourage their use. Reporting results for fits of the models, such as absolute and relative fit, parameter estimates, and standard errors, provides a wealth of information about the data and will lead to cumulative knowledge about the same–different task.

Appendix

A SAS NLMIXED program for same–different signal detection models

```
*Note: y is coded as zero/one, z = x1 − x2;
proc nlmixed data = first;
 title 'Asymmetric covert decisions with bias model, Equation 6';
 parms b = 0 a = 0 c = 0 d = 1;
 eta1 = c − d ∗ x1;
 eta2 = c + b + a − d ∗ x2;
 eta3 = c − d ∗ x1;
 eta4 = c − b − d ∗ x2;
 p = probnorm(eta1) ∗ (1 − probnorm(eta2)) + (1 −
probnorm(eta3))  ∗ probnorm(eta4);
 model y ∼ binary(p);
 predict p out = predprob;
 estimate 'rel_c' c − .5 ∗ d;
 ods output parameterestimates = pars FitStatistics = fit;
 ods output additionalestimates = other;
run;
proc nlmixed data = first qpoints = 20;
 title 'Asymmetric yardstick model, Equation 8';
 parms a = 0 t = 1 d = 1;
  eta1 = (t + a + d ∗ z + eps1);
  eta2 = (−t + d ∗ z + eps1);
  p = 1 − probnorm(eta1) + probnorm(eta2);
  model y ∼ binary(p);
  random eps1 ∼ normal(0, 1) subject = trial;
  ods output parameterestimates = subjparm;
run;
 proc nlmixed data = first;
  title 'Differencing model with asymmetry, Equation 9';
  parms a = 0 t = 1 d = 1;
  eta1 = (t + a + d ∗ z)/sqrt(2);
  eta2 = (−t + d ∗ z)/sqrt(2);
  p = 1 − probnorm(eta1) + probnorm(eta2);
  model y ∼ binary(p);
  predict p out = predprob1;
  ods output parameterestimates = subjparm;
run;
```

An OpenBUGS program for same–different signal detection models

```
# Asymmetric covert decisions model
# Note: y is recoded as 1 (for 1) and 2 (for 0)
model SD
{
#priors for parameters d, c, b, a
d ∼ dnorm(0, .1)
c ∼ dnorm(0, .1)
b ∼ dnorm(0, .1)
a ∼ dnorm(0, .1)
for (i in 1 : N) {
  p1[i, 1] <- phi(c − d ∗ x1[i]) ∗ (1 − phi(c + b + a − d ∗ x2[i])) +
(1 − phi(c − d ∗ x1[i])) ∗ phi(c − b − d ∗ x2[i])
  p1[i, 2] <- 1 − p1[i, 1]
  y1[i] ∼ dcat(p1[i, 1 : 2])
  }
}
#data
list(N = 512)
#priors
list(d = 1, c = .5, b = 0, a = 0)
# Asymmetric yardstick model
model SD
{
#priors for parameters d, c, b, a
d ∼ dnorm(0, .1)
t ∼ dnorm(0, .1)
a ∼ dnorm(0, .1)
for (i in 1 : N) {
  eps[i] ∼ dnorm(0, 1.0)
  z[i] <- x1[i] − x2[i]
  p1[i, 1] <- 1 − phi(t + a + d ∗ z[i] + eps[i]) + phi(−t + d ∗ z[i] +
eps[i])
  p1[i, 2] <- 1 − p1[i, 1]
  y1[i] ∼ dcat(p1[i, 1 : 2])
  }
}
#data
list(N = 512)
#priors
list(d = 1, t = 1, a = 0)
```

References

Dai, H., Versfeld, N. J., & Green, D. M. (1996). The optimum decision rules in the same–different paradigm. *Perception & Psychophysics*, *58*, 1–9.

DeCarlo, L. T. (2007). The mirror effect and mixture signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 18–33.

DeCarlo, L. T. (2012). On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*, *56*, 196–207.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. NY: Academic Press.

Irwin, R. J., Hautus, M. J., & Francis, M. A. (2001). Indices of response bias in the same–different experiment. *Perception & Psychophysics*, *63*, 1091–1100.

Lee, H.-S., van Hout, D., Hautus, M., & O'Mahony, M. (2007). Can the same–different test use a β-criterion as well as a τ-criterion? *Food Quality and Preference*, *18*, 605–613.

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

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, *84*, 452–471.

Noreen, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. *SIAM–AMS Proceedings*, *13*, 237–279.

O'Mahony, M. O., & Rousseau, B. (2002). Discrimination testing: a few ideas, old and new. *Food Quality and Preference*, *14*, 157–164.

Petrov, A. A. (2009). Symmetry-based methodology for decision-rule identification in same–different experiments. *Psychonomic Bulletin and Review*, *16*, 1011–1025.

Rousseau, B. (2001). The β-strategy: an alternative and powerful cognitive strategy when performing sensory discrimination tests. *Journal of Sensory Studies*, *16*, 301–318.

Sorkin, R. D. (1962). Extension of the theory of signal detectability to matching procedures in psychoacoustics. *Journal of the Acoustical Society of America*, *34*, 1745–1751.

Stillman, J. A., & Irwin, R. J. (1995). Advantages of the same–different method over the triangular method for the measurement of taste discrimination. *Journal of Sensory Studies*, *10*, 261–272.

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.