Decision strategies for the two-alternative forced choice reminder paradigm

M. J. Hautus · D. Shepherd · M. Peng

Published online: 24 December 2010
© Psychonomic Society, Inc. 2010

Abstract The decision strategy adopted by an observer in a psychophysical procedure may be established by comparing estimates of sensitivity obtained by using detection-theoretic models based on each of the available decision strategies. Estimates of sensitivity obtained from 10 observers on auditory-level discrimination tasks using the yes/no, two-alternative forced choice (2AFC), and 2AFC with reminder (2AFCR; sometimes called *duo-trio*) procedures were compared in order to determine the decision strategy used by the judges in the 2AFCR procedure. 2AFCR permits at least three decision strategies: differencing, likelihood ratio, and comparison of distances. The latter strategy has frequently been assumed in the analysis of duo-trio data in the sensory evaluation literature. The present study demonstrates that the comparison-of-distances strategy was not used by 9 of the 10 observers in this auditory experiment. The nature of the instructions given to the observers had no measurable impact on the decision strategy that they adopted.

Keywords Reminder paradigm · Decision strategy · duo-trio · 2AFCR · SDT

M. J. Hautus (✉) · M. Peng
Research Centre for Cognitive Neuroscience,
Department of Psychology, University of Auckland,
Private Bag 92019, Auckland Mail Centre,
Auckland, New Zealand
e-mail: m.hautus@auckland.ac.nz

D. Shepherd
Department of Psychology, AUT University,
Auckland, New Zealand

While it is widely known that signal detection theory (SDT) provides a mechanism to minimize the effect of response bias on estimates of discriminative ability, it has another benefit that is less well known. SDT can accommodate the decision strategy used in the process of judgment, which in itself can have a major impact on performance as measured by other indices such as proportion correct. A decision strategy is the rule applied by an observer to the sensory evidence that stems from the stimuli presented in a trial, to make a decision. There are two major decision strategies commonly reported in the literature: the likelihood ratio decision strategy and the difference decision strategy. For the likelihood ratio strategy, the observer sets a criterion at some level of likelihood ratio (of one type of event over another; e.g., same vs. different) and makes a decision based on where the likelihood of the evidence available from the stimuli presented on a trial falls about this criterion. For the difference strategy, the observer sets a criterion difference that is compared to the difference in sensory evidence that arises from two or more stimuli presented on a trial. The likelihood ratio strategy can—most easily in its simplest form, where the likelihood ratio is 1—be considered an “independent” or “absolute” approach to decision making, as the evidence from each stimulus, or stimulus sequence, is compared independently to a specific criterion. In contrast, the difference strategy can be considered a “relative” or “dependent” approach to decision making because the decision is based on the relation between the evidence arising from two or more stimuli (Hautus, Irwin, & Sutherland, 1994). These two types of decision strategy can lead to different responses for the same perceptual evidence, and hence to different levels of performance (Hautus & Irwin, 1995; Hautus & Meng, 2002; Lee, van Hout, & Hautus, 2007; Lee, van Hout,

Hautus, & O'Mahony, 2007; Noreen, 1981; O'Mahony, 1995; Rousseau, 2001). It is fortunate that SDT provides a mechanism to investigate this issue and uncover the decision strategy that is being used.

In psychophysics, there is a class of psychophysical procedures known as the *reminder paradigm* (Macmillan & Creelman, 2005, pp. 180–182). A reminder stimulus, which is always the same, is presented on every trial. An example of this approach is a yes/no procedure modified to provide a “standard” stimulus, S_1 , before each trial. This reminder stimulus may serve as a memory aid, to “remind” the observer of the nature of the standard stimulus. It may also provide the observer with a comparison for the test stimulus, which may or may not be used in the decision process. It has been argued that familiarity with the stimuli becomes less important to the task when a reminder is available (Hautus, van Hout, & Lee, 2009). The observer could simply be asked, “Were the reminder and test stimuli the same or different?” To participate sensibly in this task, it is thus not necessary for the observer to be able to label or describe the stimuli. This is a benefit of the reminder paradigm that has also been attributed to the standard same–different task (Irwin, Stillman, Hautus, & Huddleston, 1993). However, it is important to remember that the reminder yes/no procedure is not equivalent to a same–different task, even though the question asked of the observer can be framed in a same–different format.

The two-alternative forced choice procedure (2AFC) is also amenable to the application of a reminder. In a 2AFC trial, the two stimuli under consideration can be presented to the observer in two possible sequences, $\langle S_1, S_2 \rangle$ and $\langle S_2, S_1 \rangle$, and the observer's task is to indicate whether S_2 was presented first or second. An observer in this task has available two primary decision strategies. If the two observation intervals are denoted A and B , in the difference strategy the observer computes the difference between the perceptual evidence arising from each observation on a trial; that is, $B - A$. This perceptual difference is then compared to a criterion, which for an unbiased observer would be set to 0, given an equal probability of occurrence of the two stimulus sequences. If the difference exceeds the criterion, the observer responds “second”; otherwise, “first.” For the likelihood ratio decision strategy, the observer determines the ratio of the likelihood of the sequence $\langle S_1, S_2 \rangle$ to $\langle S_2, S_1 \rangle$, given the perceptual evidence on a trial. If the likelihood ratio exceeds a criterion (1 for an unbiased observer and equal probabilities of both stimulus sequences), the observer responds “second”; otherwise, “first.”

If a reminder stimulus (S_1 for consistency) is presented before the 2AFC stimuli—creating the 2AFC-with-reminder task (2AFCR)—the question asked of the observer could change from “Was S_2 presented first or second?” to “Was the reminder stimulus the same as the first or the

second test stimulus?” As with the reminder yes/no procedure, the ability to label or categorize the stimuli is not required. Again, for 2AFCR there are two primary decision strategies. As for the 2AFC task, an observer using the likelihood ratio strategy in 2AFCR would determine the ratio of the likelihood of the sequence $\langle S_1, S_1, S_2 \rangle$ to $\langle S_1, S_2, S_1 \rangle$, given the perceptual evidence on a trial. This simplifies down to the ratio of the likelihoods of the subsequences $\langle S_1, S_2 \rangle$ to $\langle S_2, S_1 \rangle$, which do not involve the reminder stimulus, R , and which yield the same decision rule as for 2AFC. Thus, the reminder can be ignored by the observer with no loss of performance.

In the difference strategy, the observer determines two perceptual differences: $A - R$ and $B - R$. If the difference $(B - R) - (A - R)$ is larger than a criterion difference (0 for an unbiased observer and equal sequence probabilities), then the observer responds “second”; otherwise, “first.” It is clear that $(B - R) - (A - R) = B - A$, and thus the decision strategy leads to the same outcome as the difference strategy in the 2AFC task. This equality also indicates that an observer could potentially perform equally well by ignoring the reminder. However, it does not mean that they necessarily ignore the reminder. What it does mean is that there are two ways an observer can use difference information, and in either case performance is theoretically equal. It may be anticipated that if the observer makes comparisons to the reminder, there will be greater memory degradation effects—due to a greater time delay between R and B than between A and B , plus the interference from A in the comparison of R and B —than if the difference between A and B is used as the basis for the decision.

One way to approach SDT modeling of these reminder procedures is to explore the relationship between the sensitivity measure obtained from the standard SDT yes/no model, $d' = z(H) - z(F)$, and the value of $z(H) - z(F)$ obtained from other procedures, with the goal of estimating the distance between the normalized perceptual distributions of S_1 and S_2 independently of the procedure in which these stimuli are employed. To accomplish this, clear definitions of the false-alarm (F) and hit (H) rates are required for each procedure. For the yes/no procedure, H is the proportion of S_2 trials on which the response is “ S_2 ,” and F is the proportion of S_1 trials on which the response is “ S_2 .” For the 2AFC and 2AFCR procedures, H is the proportion of $\langle S_1, S_2 \rangle$ trials on which the response is “second,” and F is the proportion of $\langle S_2, S_1 \rangle$ trials on which the response is “second.”

If data are collected from a yes/no task, $z(H) - z(F)$ is equal to d' . However, for data collected from other tasks, $z(H) - z(F)$ is not necessarily equal to d' . The theoretical relationship for the 2AFC procedure is $d' = [z(H) - z(F)]/\sqrt{2}$, irrespective of whether the difference or likelihood ratio strategy is assumed (Green & Swets, 1966). More recently,

SDT models for the 2AFCR procedure have been developed (Hautus et al., 2009). These models show that the same relationship holds for the difference and likelihood ratio strategies in the 2AFCR procedure as in the 2AFC procedure; that is, in both cases, for the 2AFCR procedure, $d' = [z(H) - z(F)] / \sqrt{2}$

Thus far, all decision strategies in 2AFC and 2AFCR lead to the same level of performance, and yield the same $\sqrt{2}$ relationship between d' and $z(H) - z(F)$. However, another decision strategy for the 2AFCR procedure has frequently been reported in the literature (David & Trivedi, 1962; Ennis, 1993; Ura, 1960); more recently, it has been named the “comparison-of-distances” decision strategy (Liggett & Delwiche, 2005; O’Mahony, 1995; O’Mahony, Masuoka, & Ishii, 1994; Rousseau, 2001). This decision strategy usually appears in conjunction with a procedure with the same trial structure as the 2AFCR procedure; the duo–trio (constant reference) procedure (e.g., Lawless & Heymann, 1998). Interestingly, there is another version of the duo–trio procedure that has a variable reference (e.g., Lawless & Heymann, 1998); that is, the first stimulus presented on a trial is not always the same, but is randomly chosen on each trial. The trial structure of this procedure is equivalent to that for the ABX procedure commonly employed in psychophysics (see, e.g., Macmillan & Creelman, 2005, chap. 9).

The comparison-of-distances strategy is a restricted form of differencing strategy in which the sign of the difference is ignored; thus, the differences compared by the observer are $|A - R|$ and $|B - R|$. Consequently, the stimulus that provides evidence farthest from that arising from the reminder is selected as the one that is different. This decision-strategy model is traditionally explained in the context of a unidimensional decision space (e.g., O’Mahony et al., 1994, Fig. 3). Figure 1, with a similar unidimensional approach, illustrates the different outcomes for the 2AFCR task. The distributions of sensory evidence have means at 0 and d' and unit variance. A reminder stimulus is drawn from the S_1 distribution and is depicted by a thick vertical line. A test stimulus is drawn from the S_2 distribution (dashed line). In the three examples depicted, these two stimulus values have been held constant. The third stimulus value, drawn from S_1 (solid thin line), varies across the examples, and in every case represents the correct response. For the top panel, an observer using the difference strategy would select the stimulus drawn from S_2 as being S_1 —an incorrect answer. The comparison-of-distances strategy would lead to the correct answer, since the stimulus drawn from S_1 , while greater in perceptual strength than both other stimuli, is closest to the reminder. In the middle panel, the difference strategy leads to the correct response, and in the lower panel both decision strategies lead to the same incorrect

Fig. 1 Outcomes in the 2AFCR task for two decision strategies. The distributions represent the perceptual evidence from S_1 and S_2 . Two stimuli are drawn from S_1 —the reminder (thick vertical line) and a test stimulus (solid thin vertical line)—and one from S_2 (dashed vertical line). Only the test stimulus drawn from S_1 varies in these examples. Labels on top of stimulus lines indicate the stimulus that would be chosen as S_1 (or as matching the reminder) for either the differencing (D) or comparison-of-distances (COD) strategy

response. Such figures can be useful to illustrate the outcomes for specific examples of perceptual evidence; however, it is not obvious how the likelihood ratio decision strategy could be represented, even though it is known that the outcomes are the same as for the difference strategy. Also, there is no clear way to illustrate biased responding, because these depictions do not represent the decision criterion, an integral component of any model based on SDT.

Hautus et al. (2009) presented the 2AFCR model in a two-dimensional representation of the decision space that explicitly included the decision criterion. They represented the difference between the first test stimulus and the reminder ($A - R$) on one dimension, and the difference between the second test stimulus and the reminder ($B - R$) on a second dimension. The distributions of sensory evidence for the two possible stimulus sequences $\langle S_1, S_1 \rangle$ and $\langle S_1, S_2, S_1 \rangle$ are bivariate normal with loci at $(0, d)$ and $(d, 0)$, respectively. Since the reminder stimulus is represented on each dimension, the distributions have a covariance of 1. Figure 2 illustrates the resulting decision space and includes the criterial boundaries for the comparison-of-distances strategy for the case of unbiased performance (see Fig. 7 of Hautus et al., 2009, for examples of the decision space for biased observers).

To understand the depicted boundaries, first consider the standard difference strategy. We assume unbiased responding in the description that follows, since this reflects the

Fig. 2 Decision space for the 2AFCR procedure when the comparison-of-distances strategy is used. Evidence falling within the shaded areas leads to the response “S₂ second”

decision boundaries depicted in Fig. 2. For this case, the observer would respond “B” or “second” if $B - R > A - R$, or alternatively, if $(B - R) - (A - R) > 0$. The decision boundary would therefore be given by the locus of points on the decision space where the difference between the values on each dimension is 0; that is, a line of unit slope and intercept of 0. That line is present in Fig. 2, and any evidence above that line would lead to a response of “S₂ second.” That would include the shaded area above the line, as well as the unshaded area. The entire area below the line would lead to the response “S₂ first.”

For the comparison-of-distances strategy, the unbiased decision rule of the standard difference strategy, to respond “second” if $(B - R) - (A - R) > 0$, becomes respond “second” if $|(B - R)| - |(A - R)| > 0$; that is, the direction of the differences is discarded. This results in a reversal of the responses over the portion of the decision space below a line through the origin with a slope of -1 , leading to the shaded area given in Fig. 2.

For this task, the hit rate has been defined as the probability of responding “S₂ second” when S₂ was presented second. For the comparison-of-distances strategy, this probability is the integral of the $\langle S_1, S_1, S_2 \rangle$ distribution over the shaded area. The false-alarm rate is then the probability of responding “S₂ second” when S₁ was presented second. This probability is given by the integral of the $\langle S_1, S_2, S_1 \rangle$ distribution over the same shaded region. These integrals do not have closed-form solutions, but the probabilities can be calculated using standard numerical methods.

Figure 3 shows the relationship between $z(H) - z(F)$ and d' for the various procedures and decision strategies discussed so far. These functions are independent of response bias; thus, they apply equally to F, H pairs for which $H = 1 - F$ (unbiased responding) and F, H pairs that do not adhere to this relationship (biased responding). The points depicted for the 2AFCR procedure, using the comparison-of-distances strategy, were calculated from the model just described. It is clear that under the assumption of the comparison-of-distances strategy, the relationship is monotonic but not linear. For researchers who want a relatively quick way to calculate values on the curve in Fig. 3, the following interpolating polynomial (see, e.g., Press, Flannery, Teukolsky, & Vetterling, 1992, chap. 3) provides a good approximation to the relationship between $z(H) - z(F)$ and d' :

$$d' = ax^{0.5} + bx + cx^{1.5} + dx^2 + ex^{2.5} + fx^3 + gx^{3.5} + hx^4 + ix^{4.5}, \quad (1)$$

where $x = z(H) - z(F)$, and a through i are the constants 1.480696, -0.092543 , 0.628102 , -1.127742 , 1.624117 , -1.357223 , 0.658614 , -0.165329 , and 0.016113 , respectively. The approximation is illustrated by the dashed curve in Fig. 3. The error of approximation for d' is better than .001 over the domain $0 \leq z(H) - z(F) \leq 4$ and better than .0001 over $0 \leq z(H) - z(F) \leq 3$.

Given the three decision strategies available for the 2AFCR procedure, it is necessary to investigate which is used by observers as their decision rule. While two of the strategies lead to the same level of performance (likelihood

Fig. 3 Relationship between $z(H) - z(F)$, obtained in the yes/no, 2AFC, and 2AFCR procedures, and d' . The horizontal line at $d' = 2$ is a reference line (see text). The relationships hold irrespective of the presence of response bias. Exact values for the 2AFCR comparison-of-distances strategy are fitted well by the approximating curve. (D = differencing, LR = likelihood ratio, COD = comparison of distances)

ratio and difference; but see Hautus et al., 2009, for reasons why performance under the difference strategy might still prove inferior), the use of the comparison-of-distances strategy yields much poorer performance. This can be seen from the horizontal line in Fig. 3 at $d' = 2$. The intersection of each of the functions with this line gives the observed difference, $z(H) - z(F)$, for each procedure. Thus, 2AFCR (comparison of distances) yields the smallest observed difference, followed by yes/no, and then 2AFC and 2AFCR (difference or likelihood ratio). Valid estimates of sensitivity cannot be made without knowledge of the decision strategy used. The remainder of this article presents a determination of the decision strategy used by observers discriminating between tones of different amplitudes using the 2AFCR procedure. To accomplish this, estimates of sensitivity, d' , will be estimated for the same stimuli using three different procedures: yes/no, 2AFC, and 2AFCR. The obtained estimates will be compared on the basis that estimates of d' should be similar across the three procedures.

Method

Observers

Ten students (6 females) from the University of Auckland served as observers. Five of the observers had previous experience in psychophysical experiments, and the remainder were inexperienced. The study was approved by the University of Auckland Human Participants Ethics Committee.

Apparatus and procedure

Tones of 1000 Hz were produced using LabView and generated on a 16-bit converter (National Instruments, PCI 6052E). Attenuators (Tucker Davis Technologies [TDT], Model PA5) were used to control the level of the tones before being fed to a headphone buffer (TDT, Model HB5). The stimuli were presented via a monaural headset (Telephonics, Model TDH 49P). The observer was seated in a sound-attenuating chamber (Amplaid, Model E). The chamber contained a small panel of LEDs, used to provide feedback. Observers listened monaurally with their left ear and responded on a standard keyboard.

The lowest-intensity tone employed in each procedure (S_1) was 70 dB SPL. This level was used for the reminder in the 2AFCR procedure. Tones of 71.0, 71.5, or 72.0 dB SPL were used for S_2 . Each observer underwent testing at two levels of S_2 (vs. S_1). The levels of S_2 employed for each observer were determined during the first session, with the difference between the two levels of S_2 being fixed at 0.5 dB SPL.

Each trial began with a warning interval, signaled by an LED. The latency between the warning light and the first observation interval was 750 ms. All interstimulus intervals were 500 ms, and the intertrial interval was also 500 ms. There were 110 trials in each block, the first 10 of which were discarded as practice.

Three procedures were employed. In the yes/no procedure, the observer rated whether the tone presented was S_1 or S_2 , which occurred with equal probability. In the 2AFC procedure, the observer rated which of the two tones presented was S_2 . For the 2AFCR procedure, the observer was presented with three tones; the first was always 70 dB SPL (S_1). In both the 2AFC and 2AFCR procedures, each stimulus sequence occurred with equal probability. For the 2AFCR procedure, half of the observers (Observers 1–5) were instructed to rate which of the two test stimuli was the most intense, thereby suggesting that the reminder could be ignored and encouraging the use of the likelihood ratio strategy. The remainder of the observers were instructed to rate which of the two test stimuli was different in intensity from the reminder, suggesting comparison between the stimuli and encouraging a difference or comparison-of-distances strategy. Observers were also informed that the reminder stimulus was always S_1 . All responses were made on a 6-point confidence-rating scale worded appropriately for each procedure.

Seven sessions were conducted; the first was for practice and was also used to determine the stimulus levels for S_2 to be used for subsequent sessions. For this session, six blocks of trials were run using the following sequence of procedures—2AFC, yes/no, 2AFCR, 2AFC, 2AFC, 2AFC—to familiarize the observers with each procedure and also to give sufficient evidence from the 2AFC procedure to select stimulus levels that would avoid floor or ceiling effects. Each of the remaining six sessions consisted of one block of trials on each procedure at each level difference (six blocks). Therefore, each observer undertook 600 experimental trials (excluding practice) for each combination of procedure and level. The order of the procedures was randomly arranged for each observer, as were the level differences.

Data analysis

Rating data were collected for each procedure. The equal-variance normal-normal SDT model was fitted to the rating data, using maximum likelihood estimation (e.g., Dorfman & Alf, 1969), yielding a receiver operating characteristic (ROC) curve for each set of ratings. Each ROC was defined by six parameters: five bias parameters and a parameter that represents the distance between the means of the perceptual distributions in units of their common standard deviation, $z(H) - z(F)$. For data collected using the yes/no procedure, this last parameter was equal to d' . For data collected from the 2AFC or 2AFCR procedures, this parameter was not equal to

d' . Estimates of d' for the 2AFC and 2AFCR procedures (assuming likelihood ratio and difference strategies) were obtained by transforming the estimate of $z(H) - z(F)$ obtained from the best-fitting ROC curve by a factor of $1/\sqrt{2}$, because for these procedures and decision strategies $d' = [z(H) - z(F)]/\sqrt{2}$. For the 2AFCR procedure, assuming a comparison-of-distances strategy, estimates of d' were obtained by applying Equation 1 to the obtained estimate of $z(H) - z(F)$. In this way, unbiased estimates of the perceptual separation of S_1 and S_2 were obtained for all procedures under the assumption of the various decision strategies.

Results and discussion

Six ROC curves were fitted to the data obtained from each observer (3 procedures $\times$ 2 levels). Empirical and fitted ROC curves obtained for Observers 2 and 7 are shown in Fig. 4. The fitting process led to six estimates of $z(H) - z(F)$

Fig. 4 Empirical and best-fitting ROC curves for 2 observers in the six conditions of the experiment. Observer 7 showed the most atypical results in the experiment. A lack of separation of the ROC curves for the two levels of stimulus difference in the 2AFCR conditions highlights an issue with the performance of this observer. The results for Observer 2 were more typical

for each observer. These were then converted to estimates of d' for the four distinct combinations of procedure and decision strategy [yes/no, 2AFC (differencing and likelihood ratio), 2AFCR (differencing and likelihood ratio), 2AFCR (comparison of distances)]. This was accomplished using the relationships presented in Fig. 2; namely, $d' = z(H) - z(F)$ for yes/no data, $d' = [z(H) - z(F)]/\sqrt{2}$ for 2AFC data (assuming difference and likelihood ratio strategies) and 2AFCR data (assuming difference and likelihood ratio strategies), and Equation 1 for 2AFCR data (assuming the comparison-of-distances strategy). Table 1 presents the obtained estimates of d' .

To determine the decision strategy used by each observer in the 2AFCR procedure, a comparison can be made between the two estimates of d' obtained for the 2AFCR procedure and those obtained for the yes/no and 2AFC procedures. This experimental approach has been used effectively with other psychophysical procedures (Creelman & Macmillan, 1979; Hautus & Irwin, 1995; Jesteadt & Bilger, 1974). It was expected—since it is a fundamental premise of SDT—that estimates of d' arising from the correct model would be more similar than those arising from incorrect models. In Table 1, the estimate of d' for the 2AFCR procedure that is most similar to the estimate obtained for the 2AFC procedure is in underlined boldface. For all cases, this is the 2AFCR d' value that is also most similar to the yes/no d' value. For all observers except Observer 7, the comparison-of-distances strategy was not adopted; the evidence instead favors use of the likelihood ratio or difference strategies.

For Observer 7, the evidence is equivocal at the lower level difference but favors the comparison-of-distances strategy for the higher level difference. This could be because this observer did employ the comparison-of-distances strategy. An alternative explanation is simply that this observer did not perform well on the 2AFCR procedure, for some reason finding it difficult. There is some evidence for this in the bottom right panel of Fig. 4. It is clear that stimulus level did not influence sensitivity for this observer in the 2AFCR procedure. Notwithstanding this, these data cannot clearly differentiate the alternatives.

The levels of S_2 chosen for each observer in the first session avoided ceiling effects, but there are some fairly low values of d' evident in Table 1. Given the consistency of the results already described, this spread of values does not detract from the outcome. It is relevant to note, however, that Observer 7 was one of the lowest-performing participants.

Table 1 also presents the means and standard errors of estimates of d' across all observers for each combination of procedure and decision strategy. Given the relatively consistent pattern evident in the use of decision strategies,

Table 1 Estimates of d' for the four distinct combinations of procedure vs. decision strategy at each stimulus level difference and instruction group

	Level 1				Level 2			
	Yes/No	2AFC LR, D	2AFCR LR, D	2AFCR COD	Yes/No	2AFC LR, D	2AFCR LR, D	2AFCR COD
Group 1								
1	1.12	1.47	1.51	2.81	1.61	1.77	1.93	3.45
2	0.52	0.54	0.46	1.29	1.04	0.91	0.73	1.69
3	0.93	1.30	1.35	2.58	1.33	1.57	1.53	2.85
4	0.80	0.69	0.75	1.73	0.87	0.95	0.89	1.93
5	0.88	1.06	1.25	2.44	1.28	1.55	1.45	2.72
<i>M</i>	0.85	1.01	1.06	2.17	1.23	1.35	1.31	2.53
<i>SE</i>	0.09	0.18	0.20	0.28	0.13	0.18	0.22	0.32
Group 2								
6	0.72	1.08	1.16	2.31	1.14	1.52	1.47	2.76
7	0.45	0.46	0.24	0.89	0.83	1.09	0.28	0.98
8	0.30	0.58	0.50	1.35	0.47	0.62	0.66	1.59
9	0.79	0.58	0.83	1.84	0.74	1.22	1.00	2.09
10	1.05	1.13	1.29	2.49	1.34	1.13	1.47	2.76
<i>M</i>	0.66	0.77	0.80	1.78	0.90	1.12	0.98	2.04
<i>SE</i>	0.13	0.14	0.20	0.30	0.15	0.50	0.23	0.34
G. Mean	0.76	0.89	0.93	1.97	1.06	1.23	1.14	2.28
<i>SE</i>	0.08	0.11	0.14	0.21	0.11	0.11	0.16	0.24

Note. Bold underlined values indicate the most likely decision strategy adopted in the 2AFCR procedure. The 2AFCR result for Observer 7, Level 1, is not clear. Means and their standard errors (*SE*) are provided below each group and, at the bottom of the table, the grand means across all observers. (D = differencing, LR = likelihood ratio, COD = comparison of distances).

it can be informative to make some statistical comparisons. The estimates of d' averaged across both stimulus levels for the 2AFC and 2AFCR procedures (difference or likelihood ratio) are not significantly different [$t(9) = 0.376$, n.s.], in contrast to those for 2AFC and 2AFCR (comparison of distances) [$t(9) = 8.39$, $p < .0005$]. The former finding is interesting, since the greatest similarity in estimates of sensitivity is expected between procedures that have the most similar structure, primarily because nonperceptual effects on performance would be the most similar. Clearly, the 2AFCR procedure is more similar in structure to the 2AFC procedure than it is to the yes/no procedure.

The estimates of d' averaged across both stimulus levels for the yes/no procedure are smaller than those obtained for the 2AFC procedure [$t(9) = 3.121$, $p = .012$]. This finding is of no surprise; the $1/\sqrt{2}$ relationship predicted by SDT between $z(H) - z(F)$ for the 2AFC procedure and $z(H) - z(F)$ for the yes/no procedure has often been found to be too small (see Macmillan & Creelman, 2005, pp. 175–178, for an overview). In some cases, it has been argued that this factor should be as high as 2—that is, for the 2AFC procedure $d' = [z(H) - z(F)] / 2$ —although the present data support a value around 1.65 (see Table 1; Level 1: $(0.89 \times \sqrt{2}) / 0.76 = 1.66$; Level 2: $(1.23 \times \sqrt{2}) / 1.06 = 1.64$), which falls between these two values. While our analyses control for response bias, it is possible that order effects could explain some of this

finding (e.g., García-Pérez & Alcalá-Quintana, 2010; Nachmias, 2006; Ulrich & Vorberg, 2009; Yeshurun, Carrasco, & Maloney, 2008).

The effect of instruction on performance in 2AFCR was investigated to determine whether the decision strategy used by an observer could be manipulated by directing them either to focus on the two test stimuli independently (Group 1: Observers 1–5) or to focus on the difference between each test stimulus and the reminder stimulus (Group 2: Observers 6–10). Table 1 presents no evidence of a switch in decision strategies. This was confirmed by mixed-design ANOVA, with stimulus level and instruction group as factors and using the difference between d' for 2AFC and 2AFCR as the dependent variable, so that any effect of the sensitivity of each observer was controlled. No effects or interactions were significant. This provides evidence that there was no shift toward the use of the comparison-of-distances strategy when instructions were stimulus-relative because, had there been such a shift, the d' values for 2AFCR (difference or likelihood ratio) would have been more dissimilar from those for 2AFC for Group 2 than for Group 1. This does not provide direct evidence that the use of the difference or likelihood ratio strategy is dependent on instruction type, since performance under these two models was predicted to be equivalent.

While the evidence indicates that the observers used the difference or likelihood ratio strategy in the 2AFCR task,

there is no way to distinguish from these data which strategy they were using. For either of these strategies, observers could, in theory, perform equally well by omitting evidence arising from the reminder from their process of judgment. Comments from several observers suggested that they did make comparisons to the reminder. Indeed, there was no reason why they should not. If instructions suggested that they make such a comparison, it seems reasonable that they would do so. Practice may have led them to take a “shortcut” approach of using the difference between A and B as the decision variable. An interesting finding from this research is that performance on 2AFCR was the same, irrespective of the instruction to the observers. This suggests that performance on these two decision strategies is empirically equivalent.

Little evidence was found for the adoption of the comparison-of-distances strategy in the 2AFCR procedure by observers in this experiment. This indicates that the observers were able to make better use of the information available in the perceptual evidence than is assumed by this decision strategy. This does not mean that observers would never use this strategy. Consider an experiment in which the observer is unfamiliar with the two stimuli of interest. The observer is then provided with a 2AFCR trial using these stimuli. Perhaps the only realistic option for the observer would be to choose as S_2 the stimulus most different from the reminder—that is, to use a comparison-of-distances strategy. While this is an unlikely experiment from the perspective of a psychophysicist, experiments of this kind are conducted frequently in the area of sensory evaluation. The goal here is to establish the sensitivity of the observer to stimuli (usually food products or beverages) under the normal conditions of consumption (see Hautus, O'Mahony, & Lee, 2008, for a description of this usage). It is in this context that the duo-trio procedure is frequently employed, and further research is required to determine whether or not observers use the comparison-of-distances strategy in this context.

Summary and conclusion

The 2AFCR procedure has three associated SDT models in the literature, based on the difference, likelihood ratio, and comparison-of-distances strategies. Two of these models (difference and likelihood ratio) predict the same performance given the same stimuli and observer, even though the observer uses the perceptual evidence to derive a response differently for the two strategies. The comparison-of-distances strategy is important because it is frequently assumed by researchers to be used by observers in the duo-trio procedure. This decision strategy predicts a different outcome from the other two strategies, allowing a determi-

nation to be made of the decision strategy adopted by observers in the 2AFCR procedure. The present study indicates that most of the observers adopted either the difference or the likelihood ratio strategy and that there was no effect of instruction. There was equivocal evidence that 1 observer (of 10) used the comparison-of-distances strategy. These results suggest that researchers who make the assumption that observers use the comparison-of-distances strategy in the duo-trio procedure need to consider the alternative: that observers may instead adopt a difference or likelihood ratio strategy. Drawing the wrong conclusion on this matter can make a considerable difference to the estimated sensitivity of the observers.

Acknowledgements This work was supported by a grant from the University of Auckland Research Committee. R. Philips and V. Lodhia assisted with collection of the data.

References

- Creelman, C. D., & Macmillan, N. A. (1979). Auditory phase and frequency discrimination—Comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 146–156.
- David, H. A., & Trivedi, M. C. (1962). *Pair, triangle and duo-trio tests* (Tech. Rep. 55). Blacksburg, VA: Virginia Polytechnic Institute, Department of Statistics.
- Dorfman, D. D., & Alf, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: Rating-method data. *Journal of Mathematical Psychology*, 6, 487–496.
- Ennis, D. M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies*, 8, 353–370.
- García-Pérez, M. A., & Alcalá-Quintana, R. (2010). The difference model with guessing explains interval bias in two-alternative forced-choice detection procedures. *Journal of Sensory Studies*, 25, 876–898.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hautus, M. J., & Irwin, R. J. (1995). Two models for estimating the discriminability of foods and beverages. *Journal of Sensory Studies*, 10, 203–215.
- Hautus, M. J., Irwin, R., & Sutherland, S. (1994). Relativity of judgements about sound amplitude and the asymmetry of the same-different ROC. *The Quarterly Journal of Experimental Psychology*, 47A, 1035–1045.
- Hautus, M. J., & Meng, X. (2002). Decision strategies in the ABX (matching-to-sample) psychophysical task. *Perception & Psychophysics*, 64, 89–106.
- Hautus, M. J., O'Mahony, M., & Lee, H. S. (2008). Decision strategies determined from the shape of the same-different ROC curve: What are the effects of incorrect assumptions? *Journal of Sensory Studies*, 23, 743–764.
- Hautus, M. J., van Hout, D., & Lee, H. S. (2009). Variants of A Not-A and 2AFC tests: Signal detection theory models. *Food Quality and Preference*, 20, 222–229.
- Irwin, R. J., Stillman, J. A., Hautus, M. J., & Huddleston, L. M. (1993). The measurement of taste discrimination with the same-different task: A detection-theory analysis. *Journal of Sensory Studies*, 8, 229–239.

- Jesteadt, W., & Bilger, R. C. (1974). Intensity and frequency discrimination in 1-interval and 2-interval paradigms. *The Journal of the Acoustical Society of America*, 55, 1266–1271.
- Lawless, H. T., & Heymann, H. (1998). *Sensory evaluation of food: Principles and practices*. New York: Chapman and Hall.
- 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.
- Lee, H. S., van Hout, D., Hautus, M. J., & 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.
- Liggett, R. E., & Delwiche, J. F. (2005). The beta-binomial model: Variability in overdispersion across methods and over time. *Journal of Sensory Studies*, 20, 48–61.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Research*, 46, 2456–2464.
- Noreen, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Proceedings of the symposium in applied mathematics of the American mathematical society and the society for industrial and applied mathematics* (Vol. 13). Providence, RI: American Mathematical Society.
- O'Mahony, M. (1995). Who told you the triangle test was simple? *Food Quality and Preference*, 6, 227–238.
- O'Mahony, M., Masuoka, S., & Ishii, R. (1994). A theoretical note on difference tests: Models, paradoxes and cognitive strategies. *Journal of Sensory Studies*, 9, 247–272.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). New York: Cambridge University Press.
- Rousseau, B. (2001). The β -strategy: An alternative and powerful cognitive strategy when performing sensory discrimination tests. *Journal of Sensory Studies*, 16, 301–318.
- Ulrich, R., & Vorberg, D. (2009). Estimating the difference limen in 2AFC tasks: Pitfalls and improved estimators. *Attention, Perception, & Psychophysics*, 71, 1219–1227.
- Ura, S. (1960). Pair, triangle and duo-trio test. *Reports of Statistical Application Research, Japanese Union of Scientists and Engineers*, 7, 107–119.
- Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research*, 48, 1837–1851.