THE EFFECT OF A REMINDER STIMULUS ON THE DECISION STRATEGY ADOPTED IN THE TWO-ALTERNATIVE FORCED-CHOICE PROCEDURE.

Michael J. Hautus, Daniel Shepherd, Mei Peng, Rebecca Philips and Veema Lodhia
Department of Psychology, University of Auckland, Auckland, New Zealand
and Department of Psychology, AUT University, Auckland, New Zealand

Abstract

The decision strategy adopted by an observer in a psychophysical task may be established by comparing estimates of sensitivity obtained by using detection-theoretic models that assume the various available decision strategies. The performance of ten observers on an auditory level discrimination task using the yes-no, two alternative forced choice (2AFC), and 2AFC with reminder (2AFCR; sometimes called duo-trio) procedures was compared to determine the decision strategy used in the 2AFCR procedure. This procedure permits at least three decision strategies: differencing, likelihood ratio, and comparison of distances. The latter strategy has been frequently assumed in the analysis of duo-trio data in the sensory evaluation literature. The current study demonstrates that the comparison of distances strategy was not used by nine of the ten observers.

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 that an observer applies, 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 (LR) decision strategy and the differencing (D) decision strategy. For the LR strategy, the observer sets a criterion at some level of likelihood ratio (same over different events) 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 D strategy, the observer sets a criterion difference which is compared to the difference in sensory evidence that arises from two or more stimuli presented on a trial. The LR strategy can – most easily in its simplest form where the likelihood ratio is one – be considered an ‘independent’ or ‘absolute’ approach to decision making, as the evidence from each stimulus is compared independently to a specific criterion. In contrast, the D 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 stimuli, and hence different levels of performance (Lee, van Hout, & Hautus, 2007; 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 the Yes/No procedure modified to provide a ‘standard’ stimulus, S_1, before each trial. This reminder stimulus provides the observer with a comparison for the test stimulus. Familiarity with the stimulus therefore becomes less important to the task. The observer could simply be asked, “Were the two stimuli the same or different?”

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 are presented in a random order, giving two possible sequences, $<S_1, S_2>$ and $<S_2, S_1>$, and the observer’s task is to indicate whether S_2 was presented first or second. If a reminder is presented before the 2AFC stimuli, then the question asked of the observer could change from “Was S_2 presented first or second?” to “Was the first stimulus you heard presented again second or third?” Consequently the ability to label S_2 may not be required.

One way to approach SDT modeling of these 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; the goal being to estimate the distance between the normalized perceptual distributions of S_1 and S_2 independently of the procedure in which these stimuli are employed. This theoretical relationship for the 2AFC procedure is $d' = \left(z(H) - z(F)\right)/\sqrt{2}$, irrespective of whether the D or LR strategy is assumed (Green & Swets, 1966). More recently, SDT models for the 2AFCR procedure have been developed (Hautus, van Hout, & Lee, 2009). These models show that the same relationship holds for the D and LR strategies in the 2AFCR procedure, as for the 2AFC procedure; that is, in both cases for the 2AFCR procedure, $d' = \left(z(H) - z(F)\right)/\sqrt{2}$.

However, there is another decision strategy for the 2AFCR procedure that has frequently been reported in the literature; the “comparison of distances” (COD) decision strategy (O'Mahony, Masuoka, & Ishii, 1994). This decision strategy usually appears in conjunction with a common pseudonym for the 2AFCR procedure; the duo-trio procedure (fixed reference). The COD strategy is a restricted form of differencing strategy in which the sign of the difference is ignored. Consequently, the stimulus that provides evidence closest to that arising from the reminder is selected as a match. This model is traditionally depicted in a unidimensional form (O'Mahony et al., 1994), however Hautus et al. (2009) presented the model from a multidimensional SDT perspective. The relationship between $z(H) - z(F)$ and d', under the assumption of the COD strategy and unbiased performance, is not linear (see Figure 1). A rudimentary approximation to the points for the 2AFCR procedure depicted in Figure 1 – which were calculated to 1E-7 from the model given by Hautus et al. (2009) – is provided by

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

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, 0.016113, respectively. The error of approximation for d' is better than 1E-3 over the domain 0 <= $z(H) - z(F)$ <= 4 and better than 1E-4 over 0 <= $z(H) - z(F)$ <= 3.

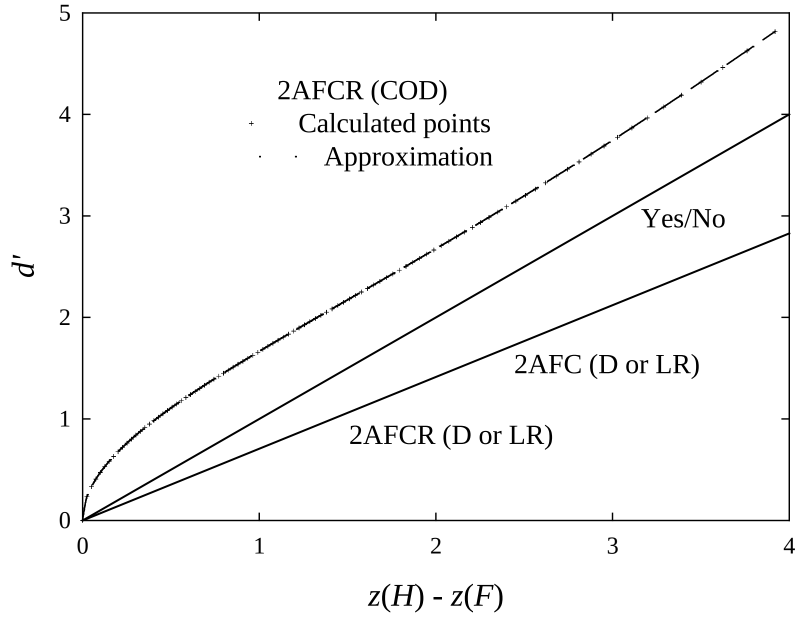

Figure 1. Relationship between $z(H) - z(F)$ obtained in the Yes/No, 2AFC, and 2AFCR procedures, and *d′*. Performance is assumed to be unbiased. Exact values for the 2AFCR COD decision strategy are fitted well by the approximating curve.

Method

Ten students (four males) from the University of Auckland served as observers. Five 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.

Tones of 1000 Hz were produced using LabVIEW and generated on a 16-bit converter (National Instruments, PCI 6052E). Attenuators (Tucker Davis Technologies (TDT), PA5) were used to control the level of the tones before being fed to a headphone buffer (TDT, HB5). The stimuli were presented via a monaural headset (Telephonics, 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 (S_1) was 70 dB SPL. This level was used for the reminder in the 2AFCR procedure, and was the lowest level tone employed in each 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 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 inter-stimulus intervals were 500 ms, and the inter-trial interval was also 500 ms. There were 110 trials in each block, the first 10 of which were discarded as practice. Observers listened to the tones and then rated their confidence on a six-point scale.

Three procedures were employed: In the Yes/No procedure the observer indicated whether the tone presented was S_1 or S_2. In the 2AFC procedure the observer indicated which of the two tones presented was S_2. For the 2AFCR procedure the observer was presented with three tones; the first tone was always 70 dB SPL (S_1). The observer indicated which of the remaining two tones was S_2. All responses were made on a six-point confidence-rating scale worded appropriately for each procedure.

Seven sessions were conducted; the first session was for practice and was also used to determine 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; chosen 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. The remaining six sessions consisted of one block of trials on each procedure at each level difference (six blocks). The order of the procedures was randomly arranged for each observer, as were the level differences.

Results and Discussion

Rating data were collated for each observer in each of the six conditions (3 procedures × 2 levels), yielding 600 trials per condition. The equal-variance normal-normal SDT model was fitted to the data, using Maximum Likelihood Estimation, yielding six Receiver Operating Characteristic (ROC) curves for each observer, and hence six estimates of $z(H)-z(F)$. These were then converted to estimates of d' for the four distinct combinations of procedure and decision strategy. This was accomplished using the relationships presented in Figure 1; namely $d'=z(H)-z(F)$ for Yes/No, $d'=\left(z(H)-z(F)\right)/\sqrt{2}$ for 2AFC (D and LR) and 2AFCR (D and LR), and using Equation 1 for 2AFCR (COD). This approach is possible because these models all produce symmetrical ROC curves. Table 1 presents the obtained estimates of sensitivity.

Table 1. Estimates of d' for the four distinct combinations of procedure vs. decision strategy at each stimulus level difference. Bold values indicate the most likely decision strategy adopted in the 2AFCR procedure. A result for Observer 7, Level 1 is not clear.

	Level 1				Level 2			
	Yes/No	2AFC LR, D	2AFCR LR, D	2AFCR COD	Yes/No	2AFC LR, D	2AFCR LR, D	2AFCR COD
Observer								
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
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
Mean	0.76	0.89	**0.93**	1.97	1.06	1.23	**1.14**	2.28
SE	0.08	0.11	0.14	0.21	0.11	0.11	0.16	0.24

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. It was expected that estimates of d' arising from the correct model will 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 boldface. For all cases, this is the 2AFCR d'-value that is also most similar to Yes/No d'-value. For all observers, except Observer 7, clearly the COD strategy was not adopted. The evidence favors the use of the LR or D strategies.

For Observer 7 the evidence is equivocal at the lower level difference, but favors the COD strategy for the higher level difference. This could be because this observer did employ the COD strategy. An alternative explanation is simply that this observer did not perform well on the 2AFCR procedure; for some reason finding it difficult. These data cannot 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. The mean estimates of d' for the 2AFC and 2AFCR procedures (D or LR) are similar, yet significantly different for the larger stimulus difference (Level 1, $t_{(9)} = 2.03$, n.s.; Level 2, $t_{(9)} = 2.50$, $p = .03$). This is an interesting result as the greatest similarity in estimates of sensitivity are expected between procedures that have the most similar structure, primarily because non perceptual 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. All other paired differences are larger.

The mean estimates of d' for the Yes/No procedure are smaller than those obtained for the 2AFC procedure. This finding is of no surprise. The $\sqrt{2}$ relationship predicted by SDT between $z(H) - z(F)$ for the 2AFC procedure and d' has often been found to be too small (see Macmillan and 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, $d' = (z(H) - z(F))/2$ – although the current data support a value around 1.65 (for both stimulus levels), which falls between these two values.

Summary and Conclusion

The 2AFCR procedure has three associated SDT models in the literature: based on the D, LR, and COD decision strategies. Two of these models (D and LR) 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 COD strategy is important because it is frequently assumed by researchers to be used by observers in the 2AFCR (or duo trio) procedure. This decision strategy predicts a different outcome from the other two strategies, allowing a determination to be made of the decision strategy adopted by observers in the 2AFCR procedure. The current study strongly indicates that most of the observers adopted either the D or the LR strategy. There was equivocal evidence that one (of ten) observer used the COD strategy. These results suggest that researchers who make the assumption that observers use the COD strategy in the 2AFCR procedure need to consider the more likely alternative; that observers typically adopt a D or LR decision strategy. Drawing the wrong conclusion on this matter can make a considerable difference to the estimated sensitivity of observers.

References

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

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 A: Human Experimental Psychology, 47A*(4), 1035-1045.

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*(3), 222-229.

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*(6), 920-928.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). New Jersey: Lawrence Erlbaum.

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.

Rousseau, B. (2001). The β–strategy: An alternative and powerful cognitive strategy when performing sensory discrimination tests. *Journal of Sensory Studies, 16*(3), 301-318.