Food Quality and Preference 22 (2011) 433–442

Contents lists available at ScienceDirect

Food Quality and Preference

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

Decision strategies for the A Not-A, 2AFC and 2AFC-reminder tasks: Empirical tests

Michael J. Hautus [a,*], Daniel Shepherd [b], Mei Peng [a]

[a] Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand
[b] Department of Psychology, Auckland University of Technology, Private Bag 92006, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 29 March 2010
Received in revised form 20 January 2011
Accepted 22 February 2011
Available online 15 March 2011

Keywords:
2AFC
2AFCR
A Not-A
Duo–trio
Signal detection theory
ROC curve
τ Criterion
β Criterion
Cognitive strategy
Decision strategy
Difference test

ABSTRACT

Signal detection theory provides an approach to modelling sensory difference tests that separates estimates of discriminability from the effect of response bias. However, assuming an incorrect decision strategy can also lead to inaccurate estimates of sensitivity, the most common index of which is d'. Using signal detection theory, Hautus, van Hout, and Lee (2009) describe and develop a number of models for the two-alternative forced choice with reminder (2AFCR) task; a task whose trial structure is identical to that of the constant-reference duo–trio task. There are alternative decision strategies (e.g., β or τ), when judging stimuli presented in the 2AFCR, to the "comparison of distances" strategy assumed to be used in the duo–trio task in food science. We investigated the decision strategies adopted by judges in the 2AFCR task using recently developed detection-theoretic models tested on purposefully collected discrimination data. The results indicate that, in the experimental context employed, the optimal β and τ strategies are more likely to be adopted by judges in the 2AFCR task than the "comparison of distances" strategy. Findings have implications for sensory difference testing and the validity of the sensitivity indices they provide.

1. Introduction

It has long been recognised that the sensory evaluation of foods and beverages is, relative to other sensory modalities such as vision and hearing, a difficult process. As judges become rapidly fatigued or satiated during tasting sessions both the quantity of the stimulus and the frequency of sample presentations must be carefully considered. Additionally, there are carry-over effects between samples so that decisions based on previously tasted samples may influence subsequent decisions. Understandably then, food scientists seek methods that provide the most accurate estimates of sensory performance for the least number of stimulus presentations.

Signal detection theory provides models for most stimulus contexts, including test methods requiring judges to make decisions based on a single sample (e.g., the A Not-A task), or tests in which a sequence of samples must be tasted prior to eliciting a judge's decision (e.g., the duo–trio task). The benefits of using models based on signal detection theory are well documented (Hautus & Irwin, 1995; Lee, van Hout, & Hautus, 2007; Lee, van Hout, Hautus, & O'Mahony, 2007) and include, at the theoretical level, a thorough method of modelling the characteristics of the sensory system under investigation, and at the empirical level, the provision of sensory acuity measures uncontaminated by response bias (Green & Swets, 1966).

There are two additional advantages of using signal detection theory. First, irrespective of the task (e.g., A Not-A or two-alternative forced-choice (2AFC)), identical stimuli yield similar values of the detection-theory index of performance, d' (Macmillan & Creelman, 2005). The same cannot be said of other commonly used measures such as the percentage (or proportion) of correct responses or the *R*-index, which can vary across tasks even when identical stimuli are employed. Fortunately this is not the case with d', as Irwin, Stillman, Hautus, and Huddleston (1993) state: "*Theoretically, a given value of d′ has a given meaning, no matter what psychophysical procedure produced it.*" (p. 230). Second, signal detection theory can elucidate the cognitive strategies underlying the decisional processes adopted by a judge. Decision strategies are the rules or 'sensory calculus' applied by judges to determine their responses to stimuli that are too similar to be perfectly discriminated. Knowledge of the decision strategies associated with a particular difference test is vitally important as different strategies can lead to different response patterns, and consequently to different levels of performance. Two generic decision strategies that have been discussed in the food science literature are the τ ("tau") and β ("beta") decision strategies (Lee, van Hout, Hautus, et al., 2007; O'Mahony and Hautus, 2008; O'Mahony and Rousseau, 2003; Rousseau, 2001; Rousseau, Meyer, & O'Mahony, 1998).

* Corresponding author. Tel.: +64 9 3737599x85924; fax: +64 9 3737450.
E-mail address: m.hautus@auckland.ac.nz (M.J. Hautus).

0950-3293/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2011.02.008

In the τ strategy, a judge uses relative information between stimuli, usually the perceptual difference. Thus the decision variable is the difference between the sensory experiences produced by two (or more) stimuli presented during a test. To use this strategy, the judge assesses the differences between the perceptual evidence arising from various stimulus presentations on a test and chooses a response based on these differences. For example, in a same–different task the judge responds "same" only if this perceptual difference between the two stimuli presented on a test is less than some criterion value (the τ criterion), otherwise the judge responds "different."

In the β strategy, a judge evaluates each stimulus in a test independently of the other stimuli, and makes use of more than just the relative difference information that is used in the τ strategy. Because each presentation of the stimulus is independent, the likelihood that the evidence arising from a presentation belongs to a particular stimulus event can be assessed, and the likelihoods associated with each presentation can be combined and compared to a criterion. For example, in the same–different task, a judge who adopts an unbiased criterion will respond "same" if the ratio of the likelihoods for the two stimuli presented on a test exceeds one (Irwin & Hautus, 1997; Johnson, 1980; Noreen, 1981). This strategy makes optimal use of the information available to the judge, and consequently for some difference tests (e.g., same–different and A Not-A with reminder tasks) performance is higher than when the τ strategy is employed. However, for some tasks performance is theoretically equivalent for both decision strategies (e.g., 2AFC and 2AFC with reminder (2AFCR) tasks) (Hautus et al., 2009).

Accurate measures of discriminability are obtained from judges who have approached asymptotic performance; that is, their performance will no longer improve with further stimulus presentations. However, attaining asymptotic performance is difficult in sensory evaluation as judges do not usually undertake the hundreds of stimulus presentations required to obtain a baseline performance.[1] Furthermore, the products themselves may be unfamiliar, further extending the required training period. Consequently, there is a genuine risk that as judges experience novel procedures and unfamiliar products, their decision criteria will vary considerably. To overcome the issues inherent in exposing judges to only a few stimulus presentations, a family of procedures known as reminder tasks can be used. Reminder tasks constitute a class of difference tests in which the first observation interval always contains an invariant reference stimulus (e.g., A in an A Not-A reminder task). The provision of a reminder stimulus lessens the importance of product familiarity and may stabilise the judge's decision criteria (Hautus et al., 2009).

The 2AFC task has a reminder equivalent, 2AFCR, which minimizes the problems associated with identifying and memorizing specific sensory dimensions (e.g., sweetness). Of interest to food scientists is the relationship between the 2AFCR and another well known difference test in the sensory sciences, the constant-reference duo–trio task. The duo–trio task, with the reminder presented first, was originally developed by the Joseph E. Seagram Quality Laboratory in 1941 (Peryam & Swartz, 1950). However, while the trial structures are identical for the 2AFCR and constant-reference duo–trio tasks, a decision strategy has been assumed by food scientists other than the β or τ decision strategies already described. The strategy, called the "comparison of distances" (COD) decision strategy (Ennis, 1993; Lee & O'Mahony, 2004; O'Mahony, 1995; O'Mahony, Masuoka, & Ishii, 1994) proposes that a judge compares the absolute perceptual distances between the reminder stimulus and each of the other two stimuli presented in a test. Such a strategy makes suboptimal use of the available perceptual information, and does not afford a simple method of calculating d'. However, it must be remembered that the constant-reference duo–trio task is usually used in contexts where the judge has impoverished information about the stimuli under test. For example, it is customary for judges to undertake only one, or very few, trials, to have no familiarization with the stimuli beforehand, and to receive no feedback. In this context, the information available to the judge may be sufficient to only allow a decision based on similarity; a COD decision strategy. In previous research, the duo–trio task has been assumed to elicit no bias, and the value of d' has been estimated from the proportion of correct responses using a formula developed by David and Trivedi (1962), or by referencing published tables (e.g., Ennis, 1993). More recent models of 2AFCR, assuming a COD strategy, are able to eliminate this assumption by having the decision criterion as part of the model.

Hautus et al. (2009), working within the framework of signal detection theory and assuming normally distributed perceptual distributions of equal variance, described models for the A Not-A task, the A Not-A with reminder task (A Not-AR), the 2AFC task, and the 2AFCR task. From these models a number of propositions, some restatements, and some novel insights, can be extracted:

(i) Only the β strategy can be reliably employed in the A Not-A task. For this case $d' = z(H) - z(F)$, where $z(\bullet)$ is the inverse normal transform (i.e., z-scores) of the false-alarm rate, F, or the hit rate, H.
(ii) The calculation to obtain d' from F and H is the same for the A Not-A task and the A Not-AR task only if the judge employs a β strategy in the A Not-AR task. If this condition is satisfied then, for both tasks, $d' = z(H) - z(F)$.
(iii) For the A Not-AR task, and only when the τ strategy is used, $d' = \sqrt{2}(z(H) - z(F))$.
(iv) For the 2AFC task, $d' = (z(H) - z(F))/\sqrt{2}$, irrespective of whether the judge adopts a β strategy or a τ strategy.
(v) Based on (iv), the decision strategy used by a judge need not be taken into account when calculating d' from data collected using the 2AFC task.
(vi) For the 2AFCR task, $d' = (z(H) - z(F))/\sqrt{2}$ irrespective of whether the judge adopts a β strategy or a τ strategy.
(vii) From (iv) and (vi) above, d' is the same for the 2AFC task and the 2AFCR task, irrespective of whether the β strategy or τ strategy is adopted by the judge; that is, $d' = (z(H) - z(F))/\sqrt{2}$ for both tasks.

Further, Hautus, Shepherd, and Peng (2011) reported that, for the case of a judge participating in a 2AFCR task and adopting the COD strategy, it is possible to approximate the relationship between d' for the 2AFCR task with the COD strategy and d' as calculated for the A Not-A task. Fig. 1 illustrates the relationships between $z(H) - z(F)$ (i.e., $d'_{\text{A Not-A}}$) and d' for the A Not-A (β strategy), A Not-AR (β and τ strategies), 2AFC (β/τ strategy) and 2AFCR (β/τ strategy) tasks.[2] The curve is for the 2AFCR task with the COD strategy, and is obtained using the approximation (see Hautus, Shepherd, & Peng, in press) based on the model described in Hautus et al. (2009).

The final comment from Hautus et al. (2009) addresses the need to further elucidate the relationship between the instructions gi-

[1] In consumer research asymptotic performance may not be of particular interest. Rather, the average (or some other normative statistic) ability of a group of consumers as they encounter the "product" in everyday life may be a far more relevant and informative measure.

[2] For the 2AFC and 2AFCR tasks we denote the β and τ strategies as β/τ because they are predicted to yield the same result (excluding factors such as memory). For A Not-AR we list the strategies separately (β or τ) because they are predicted to produce different results. For the same reason we separate β/τ and COD for 2AFCR. However, in all cases, these strategies use different decision processes.